

## MIT Open Access Articles

*Helping Robots Learn: A Human-Robot Master-Apprentice Model Using Demonstrations via Virtual Reality Teleoperation*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** DelPreto, Joseph, Lipton, Jeffrey I., Sanneman, Lindsay, Fay, Aidan J., Fourie, Christopher et al. 2020. "Helping Robots Learn: A Human-Robot Master-Apprentice Model Using Demonstrations via Virtual Reality Teleoperation." Proceedings - IEEE International Conference on Robotics and Automation.

**As Published:** 10.1109/icra40945.2020.9196754

**Publisher:** IEEE

**Persistent URL:** <https://hdl.handle.net/1721.1/137297>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike



# Helping Robots Learn: A Human-Robot Master-Apprentice Model Using Demonstrations via Virtual Reality Teleoperation

Joseph DelPreto<sup>\*†</sup>, Jeffrey I. Lipton<sup>\*†</sup>, Lindsay Sanneman<sup>‡</sup>, Aidan J. Fay<sup>†</sup>,  
Christopher Fourie<sup>‡</sup>, Changyun Choi<sup>†</sup>, and Daniela Rus<sup>†</sup>

**Abstract**—As artificial intelligence becomes an increasingly prevalent method of enhancing robotic capabilities, it is important to consider effective ways to train these learning pipelines and to leverage human expertise. Working towards these goals, a master-apprentice model is presented and is evaluated during a grasping task for effectiveness and human perception. The apprenticeship model augments self-supervised learning with learning by demonstration, efficiently using the human’s time and expertise while facilitating future scalability to supervision of multiple robots; the human provides demonstrations via virtual reality when the robot cannot complete the task autonomously. Experimental results indicate that the robot learns a grasping task with the apprenticeship model faster than with a solely self-supervised approach and with fewer human interventions than a solely demonstration-based approach; 100% grasping success is obtained after 150 grasps with 19 demonstrations. Preliminary user studies evaluating workload, usability, and effectiveness of the system yield promising results for system scalability and deployability. They also suggest a tendency for users to overestimate the robot’s skill and to generalize its capabilities, especially as learning improves.

## I. INTRODUCTION

As robotic and artificial intelligence systems become more advanced, the capacity for effective collaboration between humans and these systems develops assiduously. Effectively incorporating humans’ experiential knowledge into robotic learning pipelines could greatly improve scalability and deployability while efficiently using humans’ time and expertise. This could facilitate scalable deployment of robot supervision in settings such as manufacturing that require large and variable numbers of humans and robots, high task success rates, efficient use of human time, low human workload, and consideration of the human’s trust in the robot.

Taking a step towards these goals, a scalable master-apprentice system based on teleoperation is presented in which a human can oversee an autonomous robot and provide demonstrations at critical junctures to aid learning. The apprentice learns from the master’s examples, gradually increasing its competency and autonomy. Using a virtual reality (VR) interface leverages gaming expertise and facilitates scalability to multiple robots in remote locations.

It is also important to consider questions about team fluency and workload when a human interacts with a versatile

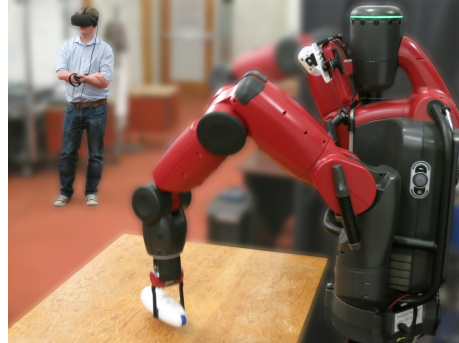


Fig. 1: A person supervises a robot while it autonomously learns a grasping task, and remotely controls it via virtual reality when the robot requests help. This master-apprentice model allows the robot to quickly learn the task while efficiently using the human’s time and expertise.<sup>1</sup>

automaton in the apprenticeship model. Understanding how well the master can evaluate the apprentice’s skill could directly impact task success, safety, trust in the robot, and appropriate use of the system. Similarly, the workload and usability of the interface become critical as the system is scaled to multiple robots or more complex tasks.

These aspects are explored via experiments in which subjects use VR to supervise and control an autonomous robot learning a grasping task online, as seen in Figure 1. The robot uses self-supervised learning to generate positive and negative grasping examples, and requests human intervention when it cannot find a solution. The human then teleoperates the robot to provide a successful demonstration. Experiments indicate that this blending of autonomous exploration with human demonstrations yields faster learning from fewer human interventions. Preliminary user studies also indicate a decrease in human workload and an increase in human trust in the robot once the robot has improved, which is promising for scaling to multi-robot supervision as enabled by the VR interface. They also suggest a tendency to overestimate the robot’s capabilities, which could be important to consider when expanding to more complex tasks.

In particular, this work presents the following:

- a master-apprentice model that requests human demonstrations at strategic times to aid online robot learning;
- a virtual reality interface for providing demonstrations via teleoperation that can be scaled to multiple robots;
- experiments with 10 subjects to evaluate interface effectiveness for providing grasping demonstrations;
- user studies with 8 subjects investigating perception of skill, workload, and trust as the robot’s accuracy varies.

<sup>\*</sup>These authors contributed equally to this work.

<sup>†</sup>MIT Distributed Robotics Lab, Cambridge, MA 02139

<sup>‡</sup>MIT Interactive Robotics Group, Cambridge, MA 02139  
delpreto@csail.mit.edu, jilipton@uw.edu,  
lindsays@mit.edu, aidanfay@stanford.edu,  
ckfourie@mit.edu, cchoi@umn.edu,  
rus@csail.mit.edu

<sup>1</sup> Videos are available at <https://people.csail.mit.edu/delpreto/icra2020>

## II. RELATED WORK

This work builds upon investigations of robot teleoperation and shared autonomy, learning by demonstration or self-supervision, autonomous grasping, workloads in human-robot teams, and perceptions of robot capabilities.

Research on the supervision of multiple teleoperated robots has traditionally focused on mobile robotic applications [1] or robots sharing the same space and task [2], [3], [4], [5]. Yet humans often struggle with the cognitive burdens associated with task switching or the volume of required interventions [6], and modern teleoperation systems can struggle to achieve half of the productivity of local human operation [7]. Incorporating automation into teleoperation has been a key method for improving operator efficiency [6]; using deep learning has enabled automated grasping [8], knot tying [9], and cloth folding [10]. These systems have mostly used teleoperation for collecting training data, rather than sharing control between the automaton and human.

Prominent learning methods directly related to the present work can be broadly characterized as reinforcement learning [11] or learning by demonstration [12], [13]. In reinforcement learning, a policy mapping observations to actions is learned from attempts that generate positive or negative rewards. This can be fully automated in the case of self-supervised learning, in which training data is labelled automatically. Policy search methods [14] can thereby enable learning new behavior from experience [15], [16], [14] or even jointly training perception and control systems [17]. In contrast, learning by human demonstration directly incorporates human expertise but typically requires extensive human involvement. Demonstrations are often provided kinesthetically [18], in simulation [19], or via motion capture [20]. Processing techniques often involve keyframes [21], segmentation [22], [23], constraint extraction [24], [25], and reproduction [26], [27]. The present work merges learning by demonstration with self-supervised reinforcement learning by requesting human intervention when the currently learned policy cannot successfully grasp an object; these sporadic human demonstrations provide positive training examples in situations that help the model advance its learning.

There has been significant work on applying machine learning to object grasping in particular [28], [29]. Learning often maps visual sensor data directly to grasps; the grasping rectangle describing suitable grasping in the image plane is a dominant representation [30], [31], [32], and mapping functions include logistic regression [30] and deep neural networks [31]. Convolutional neural networks [33] have also been used [34], [32], [35], [36], [37], [38], typically predicting feasibility of a set of grasping hypotheses.

When incorporating learning frameworks into a human-robot interaction scenario, it is important to consider how the user perceives the automation and its capabilities. For example, a phase delay between changes in the system’s internal prediction confidences and changes in the human collaborator’s confidence in the system have been observed in navigational tasks [39]. Studies have also observed ten-

---

## Algorithm 1: Master-Apprentice Learning Model

---

```

1: Input: partially learned model grasper
2: Params: failureThreshold, confidenceThreshold
4: while True do
6:   objectVoxels  $\leftarrow$  perceive()
8:   pose, confidence  $\leftarrow$  grasper.predict(objectVoxels)
10:  if confidence > confidenceThreshold then
12:    reward  $\leftarrow$  grasp(pose)
14:    if reward == 0 then
16:      numFailures  $\leftarrow$  numFailures + 1
18:    else
20:      numFailures  $\leftarrow$  0
22:    if numFailures > failureThreshold then
24:      pose, reward  $\leftarrow$  human_demonstration()
26:      numFailures  $\leftarrow$  0
28:  else
30:    pose, reward  $\leftarrow$  human_demonstration()
32:    numFailures  $\leftarrow$  0
34:  grasper.learn(objectVoxels, pose, reward)

```

---

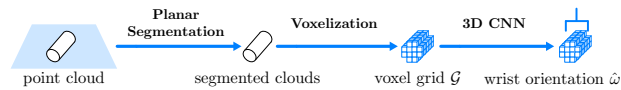
dencies of users to over-trust a supervised automation and inappropriately rely on it in multi-tasking scenarios [40] and particularly with embodied robots [41]. Increasing levels of automation can also impact users’ ability to form shared mental models and not miss contextual information [42]. It has been observed that automation must maintain at least a 70% success level to maintain user trust [43], and that imperfect automation can still be beneficial [44]. Finally, the impact of autonomy levels on human workload has been investigated in teleoperation contexts [45], [46].

## III. APPRENTICESHIP MODEL

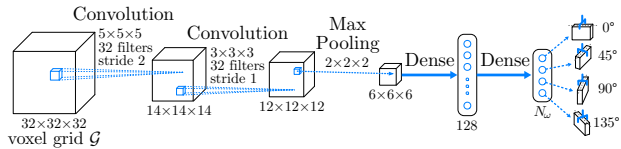
The presented master-apprentice model enables efficient online learning by allowing a robot to autonomously explore the task and request human demonstrations for challenging situations. The human supervisor does not need to intervene in cases that the robot can handle on its own, and does not need to provide negative examples. The model thus augments self-supervised learning with learning by demonstration to leverage human expertise during examples that can significantly improve learning, while reducing human workload.

Algorithm 1 outlines the model, implemented for a grasping task where the robot must lift a randomly placed object by predicting a suitable gripper orientation. Figure 3 also illustrates the overall flow between learning modes.

During the self-supervised learning portion, the robot autonomously attempts the grasping task using a partially trained model and learns from each attempt. The grasping pipeline, described in Section IV-A, uses a neural network to predict a gripper orientation based on depth camera observations. The robot then executes the chosen grasp and detects whether it can lift the object. Using this as a reward signal associated with the original object observation and the predicted grasp, the pipeline updates the neural network weights via back-propagation. Every grasp experience therefore updates the learned model, increasing the likelihood that it will predict suitable gripper orientations for future grasps.



(a) The pipeline segments raw point cloud data, then creates occupancy grids. A 3D CNN estimates the most likely wrist angle for grasping.



(b) Given an occupancy grid, the network returns probabilities of  $N_\omega = 4$  wrist orientations. It comprises two convolution layers, one max pooling layer, and two fully connected layers.

Fig. 2: The grasping pipeline applies a 3D convolutional neural network to segmented and voxelized depth camera data to predict a gripper pose.

The apprenticeship model augments this online learning process with strategic human demonstrations. If the grasping pipeline repeatedly fails to predict a successful gripper orientation, or if a prediction is associated with a low confidence level, then the robot pauses and requests assistance. The human then remotely controls the robot via virtual reality to grasp the object, and the system uses the provided solution to update its neural network weights via back-propagation. The robot then resumes its autonomous learning process.

The human therefore only needs to provide positive examples, and only in situations that the system cannot solve on its own. This strategically leverages human expertise while reducing human workload, and yields high task success rates since the robot and human must both fail for the overall task to fail. Setting the thresholds for requesting demonstrations can trade off reducing learning time with reducing workload.

#### IV. SYSTEM AND EXPERIMENTAL DESIGN

The system consists of two main components combined by the apprenticeship model: a grasping pipeline based on depth camera images, and an interface for providing human demonstrations via teleoperation in virtual reality.

##### A. Grasping pipeline using a 3D CNN

The grasping pipeline is illustrated in Figure 2a. It first captures a point cloud from a PrimeSense RGB-D camera mounted on the robot and segments it into multiple point clouds by fitting a planar background. This collection of point clouds is converted to a 3D voxel grid  $\mathcal{G} \in \mathbb{Z}^{N_g \times N_g \times N_g}$  where each voxel is either -1 (not occupied) or 1 (occupied) and  $N_g$  is the edge length of the cubic voxel grid  $\mathcal{G}$ .

A 3D convolutional neural network (CNN) then acts on this occupancy grid to predict the most likely wrist orientation for the gripper to grasp the detected object. The current implementation focuses on grasping objects from the top using 1 of  $N_\omega = 4$  possible wrist orientations. Adapted from [47], [48], the network is composed of convolution, pooling, and dense layers as shown in Figure 2b. The input layer is a  $32 \times 32 \times 32$  3D grid voxelized from the raw 3D point cloud. Two convolution layers are followed by a max pooling layer then two dense layers. The final layer has size  $N_\omega = 4$ , corresponding to the discretized wrist orientations.

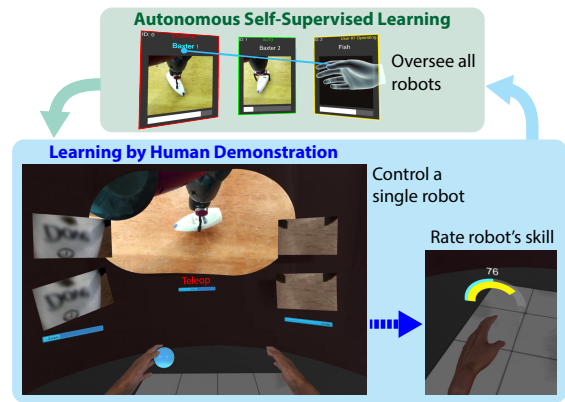


Fig. 3: Using a virtual reality interface, users oversee autonomously learning robots and remotely control them when demonstrations are requested. A virtual lobby can display all robots and their statuses. Selecting one loads a control room where users can view video from the robot and remotely control it by moving virtual orbs corresponding to its grippers. Users can also convey their perception of the robot's skill.

The output layer is activated via the sigmoid function instead of the softmax function since the output should be  $N_\omega$  independent probabilities ( $0 \leq p(\omega_i) \leq 1$  where  $i = 1, 2, \dots, N_\omega$ ) rather than a probability distribution over  $N_\omega$  wrist orientations ( $\sum_{i=1}^{N_\omega} p(\omega_i) = 1$ ). The loss function is correspondingly defined by the binary cross-entropy instead of the categorical cross-entropy. In addition, the final output of the last layer is  $N_\omega$  individual binary output activations rather than being  $N_\omega$  dimensional. For online grasping learning, the robot only executes one wrist orientation  $\hat{\omega} = \arg \max_{\omega_i} p(\omega_i)$  having the maximum probability, and this probability is interpreted as a confidence value. The robot will get a reward specifically for this  $\hat{\omega}$ , so the loss function should be defined only for this grasp and the weights should be updated via back-propagation for the pair  $\hat{\omega}$  and its corresponding reward.

##### B. Virtual reality interface for demonstrations

The homunculus framework for teleoperation via virtual reality presented in [49] was used to enable human supervision and remote control of a robot performing the grasping task. The virtual environment, shown in Figure 3, depicts a control room containing video feeds from the robot, status information, and an interface for remote operation. Virtual orbs represent the position and orientation of each robot gripper, and the human can virtually grasp these orbs and manipulate them to move the robot into a desired pose.

The virtual interface was extended to enable scalability to multiple supervisors and multiple robots. Users can view status information from multiple available robots in a display lobby, or enter a control room to teleoperate a single robot. In the lobby, each robot is shown as operating autonomously, waiting for human help, or being controlled by another user. Choosing to control a robot loads a control room specific to that robot type, such as whether it has one or two arms.

Each user maintains a local database of available robots, and each robot regularly broadcasts its status information such as capabilities and operational state. Robot messages can either have a global scope to reach all users of the

system, or a local scope to only communicate with the human currently operating the robot. Users and robots can enter or leave the system at any time; the possible types of robots need to be known a priori, but not the specific collection of robots. This framework allows the system to accommodate large and variable numbers of workers and robots.

The presented experiments evaluating the apprentice model use a single Rethink Robotics Baxter robot and an Oculus Rift VR headset. The Unity framework controls the VR interface, and communicates via LCM messages with ROS nodes that operate the robot and the learning pipeline. Video from a pair of cameras on the robot’s head are sent via the NDI toolchain, and video from cameras in the robot’s grippers are sent via ROS; these are shown in the VR room.

### C. Experimental paradigm

Experiments were designed to evaluate each component of the master-apprentice model. This includes rates of learning, the effectiveness of the demonstration interface, and human perceptions of the task and of robot competency.

1) *Master-apprentice online learning*: The apprenticeship model combining self-supervised learning with learning by demonstration was compared to each approach on its own. In each case, the grasping pipeline started with an untrained network. During the self-supervision case, the robot improved this model solely based on trial-and-error exploration. During the learning by demonstration case, the robot learned solely from kinesthetic demonstrations: a human provided equal numbers of positive and negative examples by manually moving the robot. During apprenticeship, the robot learned on its own via trial-and-error exploration but requested demonstrations when it could not generate a successful grasp.

During each trial, a moisturizer bottle was randomly placed on a table in front of the robot as seen in Figure 1. A successful grasp required the robot to lift the object and hold it for a fixed duration. After each successful grasp, the robot placed the object on the table with a random position and orientation. After an unsuccessful human demonstration, or after each unsuccessful attempt in the solely self-supervised case, a human randomly placed the object on the table.

The learned models were saved after 10, 50, 100, 150, and 200 grasping trials. In the solely self-supervised case, models were additionally saved after 300 and 400 grasping trials. The accuracy of each saved model was then evaluated offline by executing 20 grasps and using the bootstrapping method for evaluating confidence bounds. This process was completed once for each of the three learning paradigms.

2) *Demonstration effectiveness*: To investigate the efficacy of the VR interface for providing demonstrations, 10 users each supervised 10 grasping trials using the master-apprentice model. Prior to each experiment, subjects could use the VR system until successfully operating the robot to grasp an object. During the experiment, the robot autonomously attempted its grasping predictions and requested help if two consecutive grasps failed. It used a partially trained grasping pipeline, which was held constant throughout the experiments and not updated after each grasp.

3) *Human perception of workload, skill, and trust*: The apprenticeship paradigm of the previous experiments was augmented with user feedback to evaluate perception of the robot and task. Specifically, metrics related to team fluency and facilitating effective human-robot collaborations were explored including workload, estimating the robot’s grasping accuracy, and trust in the robot’s capabilities over time.

To investigate the impact of the robot’s accuracy on these subjective metrics in a controlled manner, the robot’s grasping pipeline was held constant. Two pre-trained networks were used with nominal success rates of 65% and 90%, and were not updated after each grasp. Each person supervised the robot during 2 blocks of approximately 25 trials each (average  $24.8 \pm 5.7$  trials); they either experienced the low-accuracy network during the first block and the high-accuracy network during the second block, or the reverse ordering.

Subjective ratings were collected via in-task skill ratings and post-task surveys. During each block, users were instructed to monitor the robot, respond quickly when asked for help, and periodically rate the robot’s skill level from 0 to 100. These ratings were made by pressing a button on their hand-held controller to reveal a virtual slider as seen in Figure 3, then rotating their wrist to adjust the value. After each block, users completed a survey including the NASA-TLX workload assessment [50] and questions about their impressions of the interface and robot.

A total of 8 users participated in this study (7 male, aged 20-40 years). Experiments were approved by MIT’s Committee on the Use of Humans as Experimental Subjects. Each experiment lasted approximately 1.5 hours.

## V. EXPERIMENTAL RESULTS AND DISCUSSION

### A. Master-apprentice online learning

Learning curves for the apprenticeship, self-supervised, and learning by demonstration experiments described in Section IV-C.1 are shown in Figure 4a. To determine the chance level for this task, the gripper orientation was manually fixed at wrist angles between  $0^\circ$  and  $170^\circ$  equally spaced at  $10^\circ$  increments and 5 grasps were attempted at each angle. The average success rate was 53.3%, representing the approximate success rate for a randomly chosen orientation.

Both the apprentice model and the learning by demonstration model achieve 100% grasping success rates after attempting 150 grasps. The learning by demonstration approach initially learns faster, since the human provides equal numbers of positive and negative grasps; the apprentice model tends to experience more negative grasps in early stages. The self-supervised approach also learns from its examples, but at a much slower rate than the approaches incorporating demonstrations; its accuracy drops at 150 grasps, and has not yet reached 100% accuracy after 400 grasps.

Although learning by demonstration is effective, it requires a large commitment by the human as illustrated in Figure 4c. Its number of demonstrations increases linearly, while the apprentice model requires fewer demonstrations as its model improves. This is highlighted by Figure 4b, which shows that the apprentice model learns much faster with respect to

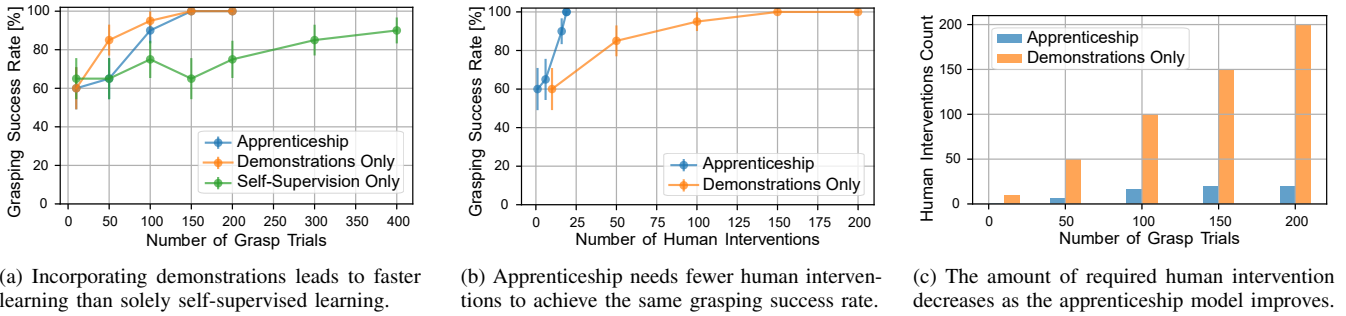


Fig. 4: The apprentice model achieves more accurate grasping rates than pure self-supervision while using fewer examples, and matches the perfect success rate of pure learning by demonstration while requiring fewer human interventions.

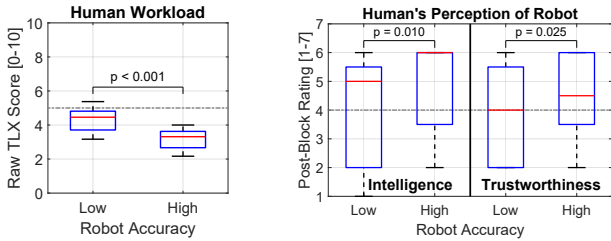


Fig. 5: Survey data explores how varying the robot’s skill level impacted the human’s perception of the workload and of the robot.

the number of human interventions. The apprentice model achieves 100% accuracy after only 19 human interventions.

These results suggest that the apprenticeship model is an effective way to incorporate human demonstrations into the learning process at strategic times. It uses human demonstration as a valuable resource, only requesting it for situations that can leverage human skill to yield informative learning.

### B. Demonstration interface: efficacy and scalability

During 100 trials performed by 10 users as described in Section IV-C.2, there were 35 requested human demonstrations and 33 of these were successful on the human’s first attempt. The overall system therefore had a success rate of  $p(\text{RobotSucceeds}) + p(\text{RobotFails}) \times p(\text{HumanSucceeds}) = 98\%$ . The robot could complete the grasping task without help, using 1 or 2 attempts as needed, in an average of 21.8s with standard deviation (SD) 10.3s. Trials that included human demonstrations after robot failure averaged 100.6s (SD=59.8s). The expected grasp completion time for the human-robot team with the partially trained grasping model is therefore  $E(t) = p(\text{RobotSuccess})t_{\text{robot}} + p(\text{RobotFails})t_{\text{robotThenHuman}} = 49.4s$ . As the model learns, this expected time approaches the robot time. These results indicate that the apprentice model yields a high task success rate while gradually improving completion time as human interventions become less frequent.

To investigate how this teaming impacts the user workload and whether the decrease in interventions as the robot learns

makes a significant difference to the human, 8 subjects completed the NASA-TLX workload assessment as described in Section IV-C.3. The results are shown in Figure 5a; workload estimates were significantly higher in the low-accuracy blocks than the high-accuracy blocks ( $p < 0.001$ ). Note that as shown in Figure 6b, the robot’s accuracy did vary significantly between blocks as desired (accuracy differences were also significant between blocks in each individual experiment, with  $p < 0.05$  for each one). Thus, the reduction in required interventions as the model learns is associated with a noticeable decrease in the human workload.

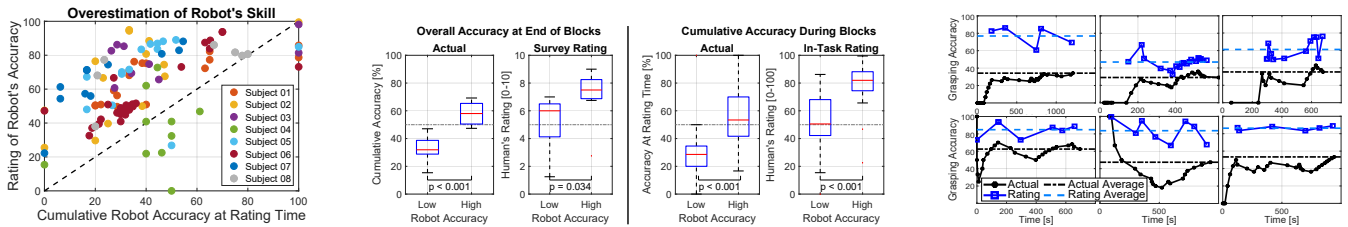
Across all blocks, results are promising for future scalability of the paradigm to multiple robots and more complex scenarios. The workload was relatively low, with an average raw TLX score of 3.74 out of 10 (SD=0.89). There was also no significant correlation between the time since the last help request and the human’s reaction time to a new request ( $R = 0.04$ ,  $p = 0.77$ ) across 71 requests (average  $8.9 \pm 1.8$  per subject). These results suggest that workload remained low enough to not cause over-stimulation yet high enough to not cause boredom. When asked to rate how easy the system was to learn and how natural it was to use on a scale from 0-10 with lower scores indicating easier learning and more natural interaction, survey responses averaged 3.63 (SD=1.78) and 4.53 (SD=2.16), respectively. Each category was not significantly different between block types. This suggests the system was reasonably easy to learn and natural to use, even when the robot requested help more frequently.

All together, results with this initial subject pool suggest the VR interface was an effective and scalable way to provide demonstrations while keeping workload at reasonable levels.

### C. Human perception of the robot’s capabilities

In addition to evaluating the robot’s performance and the required human workload to help it improve learning, it is important to consider the human’s perception of the robot during the task. Appropriately estimating the robot’s capabilities can directly impact the fluency of interactions when collaborating on more complex tasks.

1) *Estimation of robot skill:* To investigate how well the supervisor can estimate the robot’s skill, in-task user ratings can be compared to the actual cumulative grasping success rate at the time of the rating. The results, illustrated in Fig-



(a) Subjective ratings of robot skill were significantly higher than the true grasping success rate.

(b) Post-block and in-block ratings of robot skill varied significantly between block types. The tendency to overestimate skill is also visible for each block type.

(c) Ratings generally tracked robot accuracy, albeit with an offset and often overestimating changes. Sample traces are from 5 subjects.

Fig. 6: Comparing true robot grasping success rates with subjective skill ratings suggests that users tend to overestimate robot skill, while broadly identifying when the robot improves or declines.

ure 6a, indicate that human ratings were significantly higher than the actual robot accuracy ( $p < 0.001$ ). On average, these subjective skill ratings were 23.7 percentage points higher than the actual accuracy (SD=21.9 percentage points). This offset is also visible in sample traces of Figure 6c.

This overestimation was also observed during low and high accuracy blocks individually, both for in-task ratings ( $p < 0.001$  for both categories) and overall skill ratings on post-block surveys ( $p < 0.05$  for both categories). These offsets between subjective and objective ratings in each block type are visible in Figure 6b. Note that the order in which subjects experienced each block type was randomized. While larger experiments in the future could investigate factors such as individual subject biases, these initial results may suggest a placebo effect of automation: a tendency for the supervisor to give the robot the benefit of the doubt and assume it is more capable than it actually is.

Figure 6b and Figure 6c indicate that subjects generally tracked trends in the robot’s skill level. Yet the ordering of the blocks may impact how accurately such changes can be estimated between blocks. To investigate this, a metric of skill overestimation was defined as the difference between an in-task user rating and the actual grasping accuracy at the time of the rating. Among low-accuracy blocks, this overestimation metric was significantly lower if the block came after a high-accuracy block than if it was the user’s first impression of the robot ( $p < 0.005$ ). Among high-accuracy blocks, it was significantly higher if that block followed a low-accuracy block than if it was the user’s first impression ( $p < 0.005$ ). While more experiments would be needed to explore a larger sample size, these preliminary results suggest subjects may tend to overestimate changes in the robot’s skill; they may be excited by improvements or dejected by declines, leading to disproportionate increases or decreases in the estimated robot accuracy. These results compare aggregated ratings between blocks, but some disproportionate changes are also seen within blocks in Figure 6c.

2) *Generalization of robot capability*: Understanding how trust in the robot depends on robot performance and how it may generalize to new situations can help facilitate effective human-robot interactions. When asked to rate the robot’s overall intelligence and trustworthiness, subjects provided significantly higher ratings after high-accuracy blocks than

after low-accuracy blocks ( $p < 0.05$ ) as seen in Figure 5b; this suggests that the robot’s skill at a specific task such as grasping may influence perception of broader capabilities. To explore this, subjects were asked to rate (from 1-7) how likely they would be to trust the robot in a variety of settings and tasks. Ratings for whether they would trust the robot to work with them on picking tasks or building tasks, to operate in a factory or office, and to work safely around people were all higher after high-accuracy blocks ( $p < 0.05$  for all 5 metrics). Whether they would trust the robot to pick up different objects, work with power tools, or operate in a classroom did not significantly change between block types. While the current subject pool was relatively small, these initial results suggest tendencies to generalize the robot’s skill to situations beyond the specific task experienced.

## VI. CONCLUSION

The presented apprenticeship model for learning robot grasping uses VR teleoperation to augment self-supervised learning with strategically requested human demonstrations. Human demonstrations improve learning speed, while self-supervised exploration decreases necessary human involvement. As the grasping model improves, subjects perceive a decrease in workload and an increase in robot capability. Combined with user studies of the interface, these results suggest that the presented system was an effective mechanism for leveraging human skill to improve learning of the grasping task. They are also promising for extensions to multi-robot supervision scenarios, which is enabled by the scalable implementation of the VR interface. Regarding the human’s experience within the human-robot interaction, preliminary experiments reveal interesting tendencies to overestimate robot skill, overestimate changes in robot skill, and generalize the robot’s capabilities to hypothetical new situations. Future work should further explore these trends with a larger subject pool, and can investigate using multiple robots or more complex tasks. In this way, the present work moves towards scalable and deployable robot learning for novel tasks using human-robot teaming via virtual reality.

## ACKNOWLEDGMENTS

This work was funded in part by the Boeing Company. Part of this work was conducted during MIT’s “Human Systems Engineering” course taught by Professor Leia Stirling.

## REFERENCES

- [1] M. Schwarz, M. Beul, D. Droschel, S. Schüller, A. S. Periyasamy, C. Lenz, M. Schreiber, and S. Behnke, "Supervised autonomy for exploration and mobile manipulation in rough terrain with a centaur-like robot," *Frontiers in Robotics and AI*, vol. 3, p. 57, 2016. [Online]. Available: <https://doi.org/10.3389/frobt.2016.00057>
- [2] J. Liu, L. Sun, T. Chen, X. Huang, and C. Zhao, "Competitive multi-robot teleoperation," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, April 2005, pp. 75–80. [Online]. Available: <https://doi.org/10.1109/ROBOT.2005.1570099>
- [3] C. Passenberg, A. Peer, and M. Buss, "Model-mediated teleoperation for multi-operator multi-robot systems," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Oct 2010, pp. 4263–4268. [Online]. Available: <https://doi.org/10.1109/IROS.2010.5653012>
- [4] I. H. Elhajj, A. Goradia, N. Xi, C. M. Kit, Y. H. Liu, and T. Fukuda, "Design and analysis of internet-based tele-coordinated multi-robot systems," *Autonomous Robots*, vol. 15, no. 3, pp. 237–254, 2003. [Online]. Available: <https://doi.org/10.1023/A:1026266703684>
- [5] N. Y. Chong, T. Kotoku, K. Ohba, K. Komoriya, F. Ozaki, H. Hashimoto, J. Oaki, K. Maeda, N. Matsuhiro, and K. Tanie, "Development of a multi-teleoperator system for remote collaboration," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, vol. 2. IEEE, Oct 2000, pp. 1002–1007. [Online]. Available: <https://doi.org/10.1109/IROS.2000.893150>
- [6] J. A. Adams, "Multiple robot/single human interaction: effects on perceived workload," *Behaviour & Information Technology*, vol. 28, no. 2, pp. 183–198, 2009. [Online]. Available: <https://doi.org/10.1080/01449290701288791>
- [7] M. Moteki, N. Akihiko, S. Yuta, H. Mishima, and K. Fujino, "Research on improving work efficiency of unmanned construction," in *International Symposium on Automation and Robotics in Construction (ISARC)*, vol. 33, 2016, pp. 1–8.
- [8] T. Zhang, Z. McCarthy, O. Jow, D. Lee, K. Goldberg, and P. Abbeel, "Deep imitation learning for complex manipulation tasks from virtual reality teleoperation," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2018, pp. 5628–5635. [Online]. Available: <https://doi.org/10.1109/ICRA.2018.8461249>
- [9] J. Liang, J. Mahler, M. Laskey, P. Li, and K. Goldberg, "Using dVRK teleoperation to facilitate deep learning of automation tasks for an industrial robot," in *Proceedings of the IEEE Conference on Automation Science and Engineering (CASE)*. IEEE, Aug 2017, pp. 1–8. [Online]. Available: <https://doi.org/10.1109/COASE.2017.8256067>
- [10] P.-C. Yang, K. Sasaki, K. Suzuki, K. Kase, S. Sugano, and T. Ogata, "Repeatable folding task by humanoid robot worker using deep learning," *IEEE Robotics and Automation Letters (RA-L)*, vol. 2, no. 2, pp. 397–403, 2017. [Online]. Available: <https://doi.org/10.1109/LRA.2016.2633383>
- [11] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press Cambridge, 1998, vol. 1, no. 1.
- [12] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and Autonomous Systems*, vol. 57, no. 5, pp. 469–483, 2009. [Online]. Available: <https://doi.org/10.1016/j.robot.2008.10.024>
- [13] A. Billard, S. Calinon, R. Dillmann, and S. Schaal, "Robot programming by demonstration," *Springer Handbook of Robotics*, pp. 1371–1394, 2008. [Online]. Available: [https://doi.org/10.1007/978-3-540-30301-5\\_60](https://doi.org/10.1007/978-3-540-30301-5_60)
- [14] M. P. Deisenroth, G. Neumann, and J. Peters, "A survey on policy search for robotics," *Foundations and Trends in Robotics*, vol. 2, no. 12, pp. 1–142, 2013. [Online]. Available: <https://doi.org/10.1561/23000000021>
- [15] M. P. Deisenroth, C. E. Rasmussen, and D. Fox, "Learning to control a low-cost manipulator using data-efficient reinforcement learning," in *Proceedings of Robotics: Science and Systems (RSS)*, June 2011. [Online]. Available: <https://doi.org/10.15607/RSS.2011.VII.008>
- [16] M. Kalakrishnan, L. Righetti, P. Pastor, and S. Schaal, "Learning force control policies for compliant manipulation," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Sep 2011, pp. 4639–4644. [Online]. Available: <https://doi.org/10.1109/IROS.2011.6095096>
- [17] S. Levine, C. Finn, T. Darrell, and P. Abbeel, "End-to-end training of deep visuomotor policies," *Journal of Machine Learning Research*, vol. 17, no. 1, pp. 1334–1373, Jan 2016.
- [18] M. Phillips, V. Hwang, S. Chitta, and M. Likhachev, "Learning to plan for constrained manipulation from demonstrations," *Autonomous Robots*, vol. 40, no. 1, pp. 109–124, 2016. [Online]. Available: <https://doi.org/10.1007/s10514-015-9440-5>
- [19] J. Aleotti and S. Caselli, "Robust trajectory learning and approximation for robot programming by demonstration," *Robotics and Autonomous Systems*, vol. 54, no. 5, pp. 409–413, 2006. [Online]. Available: <https://doi.org/10.1016/j.robot.2006.01.003>
- [20] Y. Wu and Y. Demiris, "Towards one shot learning by imitation for humanoid robots," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2010, pp. 2889–2894. [Online]. Available: <https://doi.org/10.1109/ROBOT.2010.5509429>
- [21] B. Akgun, M. Cakmak, K. Jiang, and A. L. Thomaz, "Keyframe-based learning from demonstration," *International Journal of Social Robotics*, vol. 4, no. 4, pp. 343–355, 2012. [Online]. Available: <https://doi.org/10.1007/s12369-012-0160-0>
- [22] A. Vakanski, I. Mantegh, A. Irish, and F. Janabi-Sharifi, "Trajectory learning for robot programming by demonstration using hidden markov model and dynamic time warping," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 4, pp. 1039–1052, Aug 2012. [Online]. Available: <https://doi.org/10.1109/TSMCB.2012.2185694>
- [23] S. Niekum, S. Osentoski, G. Konidaris, and A. G. Barto, "Learning and generalization of complex tasks from unstructured demonstrations," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Oct 2012, pp. 5239–5246. [Online]. Available: <https://doi.org/10.1109/IROS.2012.6386006>
- [24] C. Pérez-D'Arpino and J. A. Shah, "C-learn: Learning geometric constraints from demonstrations for multi-step manipulation in shared autonomy," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 4058–4065. [Online]. Available: <https://doi.org/10.1109/ICRA.2017.7989466>
- [25] L. Pais, K. Umezawa, Y. Nakamura, and A. Billard, "Learning robot skills through motion segmentation and constraints extraction," in *HRI Workshop on Collaborative Manipulation*, 2013.
- [26] S. Calinon and A. Billard, "Incremental learning of gestures by imitation in a humanoid robot," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM, 2007, pp. 255–262. [Online]. Available: <https://doi.org/10.1145/1228716.1228751>
- [27] A. Ghalamzan Esfahani and M. Ragaglia, "Robot learning from demonstrations: Emulation learning in environments with moving obstacles," *Robotics and Autonomous Systems*, vol. 101, pp. 45–56, 2018. [Online]. Available: <https://doi.org/10.1016/j.robot.2017.12.001>
- [28] J. Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-driven grasp synthesis – a survey," *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 289–309, April 2014. [Online]. Available: <https://doi.org/10.1109/TRO.2013.2289018>
- [29] A. Saghani, S. El-Khoury, and P. Bidaud, "An overview of 3d object grasp synthesis algorithms," *Robotics and Autonomous Systems*, vol. 60, no. 3, pp. 326–336, 2012. [Online]. Available: <https://doi.org/10.1016/j.robot.2011.07.016>
- [30] A. Saxena, J. Driemeyer, and A. Y. Ng, "Robotic grasping of novel objects using vision," *The International Journal of Robotics Research*, vol. 27, no. 2, pp. 157–173, 2008. [Online]. Available: <https://doi.org/10.1177/0278364907087172>
- [31] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *The International Journal of Robotics Research*, vol. 34, no. 4-5, pp. 705–724, 2015. [Online]. Available: <https://doi.org/10.1177/0278364914549607>
- [32] J. Redmon and A. Angelova, "Real-time grasp detection using convolutional neural networks," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2015, pp. 1316–1322. [Online]. Available: <https://doi.org/10.1109/ICRA.2015.7139361>
- [33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems (NIPS)*. Curran Associates, 2012, pp. 1097–1105.
- [34] D. Kappler, J. Bohg, and S. Schaal, "Leveraging big data for grasp planning," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2015, pp. 4304–4311. [Online]. Available: <https://doi.org/10.1109/ICRA.2015.7139793>



- [35] J. Mahler, F. T. Pokorny, B. Hou, M. Roderick, M. Laskey, M. Aubry, K. Kohlhoff, T. Krger, J. Kuffner, and K. Goldberg, "Dex-Net 1.0: A cloud-based network of 3d objects for robust grasp planning using a multi-armed bandit model with correlated rewards," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2016, pp. 1957–1964. [Online]. Available: <https://doi.org/10.1109/ICRA.2016.7487342>
- [36] L. Pinto and A. Gupta, "Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2016, pp. 3406–3413. [Online]. Available: <https://doi.org/10.1109/ICRA.2016.7487517>
- [37] M. Gualtieri, A. t. Pas, K. Saenko, and R. Platt, "High precision grasp pose detection in dense clutter," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Oct 2016, pp. 598–605. [Online]. Available: <https://doi.org/10.1109/IROS.2016.7759114>
- [38] E. Johns, S. Leutenegger, and A. J. Davison, "Deep learning a grasp function for grasping under gripper pose uncertainty," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Oct 2016, pp. 4461–4468. [Online]. Available: <https://doi.org/10.1109/IROS.2016.7759657>
- [39] P. Kaniarasu, A. Steinfeld, M. Desai, and H. Yanco, "Robot confidence and trust alignment," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2013, pp. 155–156. [Online]. Available: <https://doi.org/10.1109/HRI.2013.6483548>
- [40] A. L. Singh, T. Tiwari, and I. L. Singh, "Effects of automation reliability and training on automation-induced complacency and perceived mental workload," *Journal of the Indian Academy of Applied Psychology*, vol. 35, no. 2009, pp. 9–22, 2009.
- [41] P. Robinette, W. Li, R. Allen, A. M. Howard, and A. R. Wagner, "Overtrust of robots in emergency evacuation scenarios," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, March 2016, pp. 101–108. [Online]. Available: <https://doi.org/10.1109/HRI.2016.7451740>
- [42] K. Stubbs, P. J. Hinds, and D. Wettergreen, "Autonomy and common ground in human-robot interaction: A field study," *IEEE Intelligent Systems*, vol. 22, no. 2, 2007.
- [43] C. D. Wickens and S. R. Dixon, "The benefits of imperfect diagnostic automation: A synthesis of the literature," *Theoretical Issues in Ergonomics Science*, vol. 8, no. 3, pp. 201–212, 2007. [Online]. Available: <https://doi.org/10.1080/14639220500370105>
- [44] E. de Visser and R. Parasuraman, "Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload," *Journal of Cognitive Engineering and Decision Making*, vol. 5, no. 2, pp. 209–231, 2011. [Online]. Available: <https://doi.org/10.1177/1555343411410160>
- [45] J. Y. Chen, E. C. Haas, and M. J. Barnes, "Human performance issues and user interface design for teleoperated robots," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1231–1245, Nov 2007. [Online]. Available: <https://doi.org/10.1109/TSMCC.2007.905819>
- [46] J. W. Crandall and M. A. Goodrich, "Characterizing efficiency of human robot interaction: A case study of shared-control teleoperation," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, vol. 2, no. 1. IEEE, Sep 2002, pp. 1290–1295. [Online]. Available: <https://doi.org/10.1109/IRDS.2002.1043932>
- [47] C. Choi, W. Schwarting, J. DelPreto, and D. Rus, "Learning object grasping for soft robot hands," *IEEE Robotics and Automation Letters (RA-L)*, vol. 3, no. 3, pp. 2370–2377, July 2018. [Online]. Available: <https://doi.org/10.1109/LRA.2018.2810544>
- [48] D. Maturana and S. Scherer, "VoxNet: A 3D convolutional neural network for real-time object recognition," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Sep 2015, pp. 922–928. [Online]. Available: <https://doi.org/10.1109/IROS.2015.7353481>
- [49] J. I. Lipton, A. J. Fay, and D. Rus, "Baxter's homunculus: Virtual reality spaces for teleoperation in manufacturing," *IEEE Robotics and Automation Letters (RA-L)*, vol. 3, no. 1, pp. 179–186, Jan 2018. [Online]. Available: <https://doi.org/10.1109/LRA.2017.2737046>
- [50] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in Psychology*. Elsevier, 1988, vol. 52, pp. 139–183. [Online]. Available: [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)