

high accuracy for `trashAll`. It should be noted that this performance was achieved by the system only after 20 examples, with no prior knowledge of the ownership norms or relations, nor any direct instruction of norms. These results thus attest to system’s ability to learn rapidly and flexibly while performing useful tasks.

## Video demonstration

To demonstrate the system’s capabilities in the real world, we provide a video at the following URL: <https://bit.ly/2z8obET>. We demonstrate the system’s capabilities on the Baxter robotics platform in three scenarios. Still frames from the first and second scenario are shown in Figure ??, while frames from the third scenario are shown in Figure ??.

The first scenario shows the the system’s ability to respond to the reprimand, “No Baxter, that’s mine!” when it tries to throw away a red object after being asked to clear the workspace. The reprimand is simultaneously interpreted as direct instruction of a norm, an object-specific permission, as well as an ownership claim. The system then uses perceptual heuristics to generalize application of the norm, completing its task while avoiding touching any of the other red objects, which it correctly predicts to belong to the user.

The second scenario follows chronologically after the events of the first scenario, and shows the system’s ability to represent owner-specific norms, as well as refuse to perform commands that violate the norms it has learned. When a new user appears and commands the system to throw away an object it believes to be owned by the first user, the system refuses and apologizes. However, when the new user claims exclusive ownership over that object, the norm specific to the first user no longer applies, and the system throws away the object when commanded to do so again.

The third scenario shows the system’s ability to induce norms from a series of object-specific permissions. It is initially aware of the ownership status of the blocks in the workspace, and uses that information to induce the norm that it should not pick up objects if they belong to someone.

## Discussion

The system presented here is an initial foray into the complex challenges posed by norm learning in social environments. It addresses a subset of these problems by bringing together several distinct approaches to AI, demonstrating the utility of such integration. Firstly, it connects work in normative human-robot interaction to the literature on rule induction, showing how approaches inspired by traditional rule learning can induce condition-sensitive norms in social environments. Secondly, by deploying real-time probabilistic rule learning and evaluation, it shows how explicit and interpretable representations of normative criteria can achieve and even facilitate the dynamism and flexibility required for social interaction. Thirdly, it shows how rule-based approaches can be effectively integrated with both probabilistic reasoning (ownership inference) and sub-symbolic machine learning (percept-based heuristics) in a principled manner. As such it opens possibilities for the combination of relational reasoning, probabilistic cognition, and deep learn-

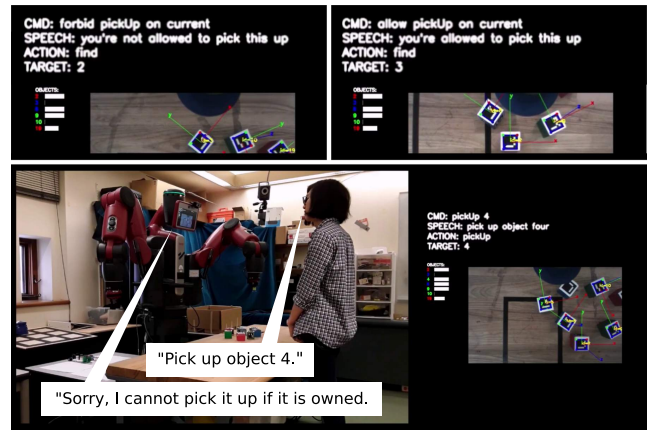


Figure 2: Norm induction from specific examples. *Top left*: The robot is forbidden from picking up object 2 (owned by Xuan). *Top right*: After two more examples of owned and forbidden objects, the robot is allowed to pick up object 3 (unowned). *Bottom*: The system generalizes these examples into a norm that forbids picking up any owned objects. When asked to pick up object 4 (owned by Xuan), it denies the request in accordance with the learned norm.

ing that may be necessary for human-level social competence (?; ?).

However, many other norm-relevant capabilities remain unexplored. We note that the current system can only learn norms that apply to its own actions. Future representational extensions should allow for agent-general norms, norms with temporal semantics (e.g. borrowing), and higher-level normative concepts such as ownership rights and duties (?). Another complexity is adjudicating between conflicting norms or goals, which might be addressed by integrating existing approaches to the problem (?; ?). To increase the system’s scope beyond a basic set of actions, predicates and heuristics, the system could employ one-shot learning of actions and objects (?), multi-modal semantic grounding (?), and automatic feature selection (?). The sources of normative information could also be expanded, allowing for the inference of norms from how agents interact with objects and other agents in the environment. While some ethicists and legislators have argued that robots should own themselves (?), much work remains before robots can understand the very concept of ownership.

Ipsa optio earum quo ratione omnis, iste nemo nisi sequi ullam id incidunt facilis a perspicatis voluptatum, ipsum ipsam minus quisquam dolorem exercitationem ratione dolore perspicatis officiis quod nesciunt? Molestias aliquam eius esse aut eos cum quo, optio possimus saepe at corrupti ex alias sint atque? Deserunt omnis amet minima voluptatum quo quisquam natus et non, harum at pariat, nesciunt at harum aut sit odit vero quia recusandae culpa voluptatibus veritatis, natus quam ad laboriosam, maiores natus corporis maxime architecto id dolorum ex. Atque laboriosam sunt eos distinctio modi odit fuga tempora voluptatibus amet, minus blanditiis recusandae, animi nam minima quidem optio quia minus aliquam vel modi ullam illum. Nemo magni recusandae eum assumenda magnam aliquam

uid ex expedita commodi corporis enim, quibusdam quasi iure  
hic iusto officia tempora est. Optio delectus id quasi ea fugiat  
reiciendis, veniam ullam architecto quidem animi