Curiosity Did Not Kill the Robot: A Curiosity-based Learning System for a Shopkeeper Robot

MALCOLM DOERING, ATR, Japan and Osaka University, Japan
PHOEBE LIU, Figure Eight, U.S.A.
DYLAN F. GLAS, Futurewei Technologies, U.S.A
TAKAYUKI KANDA, ATR, Japan and Kyoto University, Japan
DANA KULIĆ, University of Waterloo, Canada
HIROSHI ISHIGURO, ATR, Japan and Osaka University, Japan

15

Learning from human interaction data is a promising approach for developing robot interaction logic, but behaviors learned only from offline data simply represent the most frequent interaction patterns in the training data, without any adaptation for individual differences. We developed a robot that incorporates both data-driven and interactive learning. Our robot first learns high-level dialog and spatial behavior patterns from offline examples of human–human interaction. Then, during live interactions, it chooses among appropriate actions according to its curiosity about the customer's expected behavior, continually updating its predictive model to learn and adapt to each individual. In a user study, we found that participants thought the curious robot was significantly more humanlike with respect to repetitiveness and diversity of behavior, more interesting, and better overall in comparison to a non-curious robot.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; • **Computing methodologies → Online learning settings**; • **Computer systems organization → Robotics**;

Additional Key Words and Phrases: Data-driven social interaction, curiosity-based learning

ACM Reference format:
Malcolm Doering, Phoebe Liu, Dylan F. Glas, Takayuki Kanda, Dana Kulić, and Hiroshi Ishiguro. 2019. Curiosity Did Not Kill the Robot: A Curiosity-based Learning System for a Shopkeeper Robot. *ACM Trans. Hum.-Robot Interact.* 8, 3, Article 15 (July 2019), 24 pages.
https://doi.org/10.1145/3326462

Phoebe Liu and Dylan F. Glas were at ATR, Japan when this work was conducted.
The research was supported by JST, ERATO, ISHIGURO symbiotic Human-Robot Interaction Project, Grant Number JPMJER1401.
Authors' addresses: M. Doering and T. Kanda, Kanda Laboratory, Department of Social Informatics, Graduate School of Informatics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan; emails: doering@robot.soc.i.kyoto-u.ac.jp, kanda@i.kyoto-u.ac.jp; P. Liu, Figure Eight, an Appen company, 940 Howard St., San Francisco, CA 94103, USA; email: phoebe.liu@figure-eight.com; D. F. Glas, Futurewei Technologies, 100 First Street, Suite 2250, San Francisco, CA 94105 USA; email: dylan.f.glas@gmail.com; D. Kulić, Department of Electrical and Computer Engineering, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1; email: dkulic@ece.uwaterloo.ca; H. Ishiguro, Intelligent Robotics Laboratory, Department of Systems Innovation, Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama Toyonaka Osaka 560-8531 Japan; email: ishiguro@irl.sys.es.osaka-u.ac.jp.

2573-9522/2019/07-ART15
https://doi.org/10.1145/3326462

1 INTRODUCTION

In recent years, opportunities for the public to interact with social robots have become more common, with robots appearing in education (Chang et al. 2010; Kanda et al. 2004), entertainment (Kozima et al. 2009; Michalowski et al. 2007), and in the service industry (Severinson-Eklundh et al. 2003; Triebel et al. 2016). Thanks to the increasing availability of big data, one promising approach to train conversational robots to interact autonomously with humans is to automatically learn replicable dialog patterns (e.g., question–answer) from observations of real human–human interaction, as demonstrated in prior work for robots in the role of a shopkeeper (Liu et al. 2016), an assistant (Breazeal et al. 2013), a bartender (Petrick and Foster 2012), and a storyteller (Leite et al. 2016).

While existing data-driven approaches enable a robot to generate socially appropriate behaviors, such learning-based techniques only learn from previously collected data (Admoni and Scassellati 2014; Breazeal et al. 2013; Leite et al. 2016; Liu et al. 2016; Petrick and Foster 2012; Young et al. 2013) but are not able to continuously adapt to human actions during live interaction once the initial interaction strategies have been learned. This can result in a robot that behaves the same way regardless of the human's behaviors, which sometimes yields monotonous, boring interactions. To illustrate, imagine the following conversation (Figure 1) between a customer and a shopkeeper in a camera shop:

> *Customer*: I am looking for a camera to take pictures of friends.
> *Shopkeeper*: What sort of camera did you have before?
> *Customer*: I just used my phone to take pictures.
> . . . after customer and shopkeeper talk for a while. . .
> *Customer*: So anyway, I'm looking for a camera to take pictures of friends.

In this example, the shopkeeper starts his line of questioning believing it will lead to a sale. However, when the sale does not occur, and the customer later restates his purpose, the interaction becomes a learning opportunity for the shopkeeper. At this point, the human shopkeeper, having already tried the most promising strategy to make a sale, will be driven by his *intrinsic motivation* to try different actions, which could possibly lead to a favorable outcome. From a learning perspective, this can be seen as an exploration of the interaction space, probing to see how the customer will react to various actions, such as presenting the customer with a camera or leaving the customer to look around on their own. In machine learning, the concept of "curiosity" has been defined to represent an intrinsic reward signal that motivates an agent to explore a feature space in such a way (Kaplan and Oudeyer 2011; Oudeyer et al. 2007; Schmidhuber 2013).

Inspired by our own human sense of curiosity, we propose an approach that enables a robot to learn continuously through interaction, driven by an entropy-based "curiosity" mechanism. Nonetheless, "curiosity" should not *kill* the robot. That is, the robot should not randomly or naively explore behaviors in an unconstrained space such that it fails its job as a shopkeeper (e.g., saying "goodbye" when the customer enters the shop). To ensure that the robot's responses remain task appropriate, the structure of common dialog patterns for reproducing high-level behavior can be learned *a priori* (Liu et al. 2016; Liu et al. 2017a). Then, during interaction, the robot can learn about the customer's individual differences, driven to explore different behaviors by the "curiosity" mechanism while being constrained by the appropriate pre-learned patterns. Our goal in this work is not to fully replicate human curiosity, not to exploit learned information in goal-driven behavior, nor to pursue specific customer information. Rather, our aim is to emulate curiosity sufficiently to drive the robot to explore a variety of behaviors and learn about customers' individual differences, to yield more humanlike, interesting human–robot interactions.

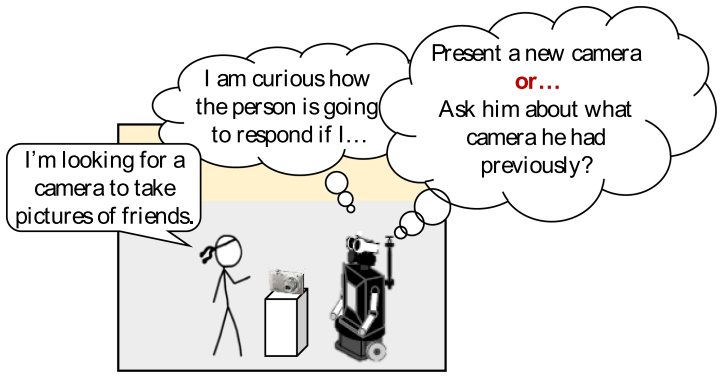


Fig. 1. A curious robot chooses among the appropriate actions to satisfy its curiosity about the customer's individual differences during interaction.

2 RELATED WORK

2.1 Data-driven Approaches for Conversational Agents and Social Robots

There exist some dialog systems that learn the state of a conversation and model it over time, so that they can always perform the most appropriate action; however, the states, which are formulated as slots and values (Williams et al. 2013; Williams and Young 2007), depend on human designers to specify domain-specific knowledge and annotate training data, so they are not suitable for unsupervised learning of robot behaviors. In other cases, the systems have not been evaluated in live interactions with humans (Breuing and Wachsmuth 2012; Jokinen et al. 1998). For social robots, several studies have aimed to learn robot behaviors for completing the same task as a human from data collected from online games (Breazeal et al. 2013; Chernova et al. 2011; Toris et al. 2014), remote web users (Leite et al. 2016), and real human interaction data (Admoni and Scassellati 2014; Foster et al. 2012; Nagai et al. 2008). Thomaz et al. developed a framework for online users to provide feedback to a Reinforcement Learning (RL) agent learning to perform tasks observed in a game (Thomaz and Breazeal 2006; Thomaz and Breazeal 2008). Leite et al. proposed a semi-situated learning method to crowdsource dialog lines from multiple authors, with each dialog line associated with a goal-directed action (Leite et al. 2016). Our work complements these approaches by considering data collected directly from human–human interaction in a physical environment; however, we are also interested in a robot that continues to learn based on its intrinsic motivation to better adapt to each user over time.

2.2 Curiosity-driven Learning

There is some work, especially in the field of sensorimotor learning, on generating autonomous agent behaviors based on intrinsic motivation, without pre-programmed goals (Kaplan and Oudeyer 2011; Oudeyer et al. 2007; Schmidhuber 2013). Oudeyer et al. developed a RL algorithm with the objective of maximizing learning progress (Oudeyer et al. 2007). Kaplan and Oudeyer demonstrated that a robot is able to explore within its repertoire of motor primitives (Kaplan and Oudeyer 2011). Action selection strategies based on information gain have also been applied for spatiotemporal exploration by mobile robots in continuously changing environments (Müller et al. 2014; Santos et al. 2017). The interactive art sculpture presented in Chan et al. (2015) is a distributed system with a large sensorimotor space that employed a similar concept of curiosity-based learning. Their experiment demonstrated that the sculpture gradually shifted to more exploratory actuation patterns as it learned about its own mechanisms and surroundings through self-experimentation and interaction.

Hester et al. presented an RL model that used intrinsic rewards to explore uncertain sensorimotor spaces and acquire new experiences for training the model (Hester and Stone 2017). It was tested on a robot that learned to hit a cymbal. Qureshi et al. introduced an intrinsic motivation system for a social robot to learn when to perform gestures and facial expressions in real-world interactions based on minimization of state prediction error via RL (Qureshi et al. 2018). Our work also employs the intrinsic drive of curiosity but for a social robot that generates verbal expressions and nonverbal motion.

Madani et al. proposed a two-level cognitive system for learning language-to-perception groundings with adaptive visual perception cast as "perceptual curiosity" and an evolutionary system to learn grounding functions to guide "epistemic curiosity" (Madani et al. 2016). As the system learned, epistemic curiosity drove the robot to ask about objects with low-confidence groundings. In contrast to their approach, our work focuses on social behavior and uses and entropy-based metric to explore low-confidence social spaces.

2.3 Robots That Elicit Curiosity in Humans

Some studies have focused on eliciting curiosity in humans for the sake of incentivizing productivity or improving social interaction. For instance, eliciting curiosity in humans was investigated for a robot that generated unpredictable responses (Law et al. 2017) and for a robot that demonstrated curiosity in its verbal expressions (Gordon et al. 2015). Their results showed that such robot behaviors positively influenced the user's experience during interaction. Likewise, we expect that a curiosity-driven robot will have a similar effect. The originality of our work lies in generating curiosity-driven behaviors via continuous learning, rather than manually implementing pre-scripted verbal expressions.

2.4 Active Learning for Social Robots

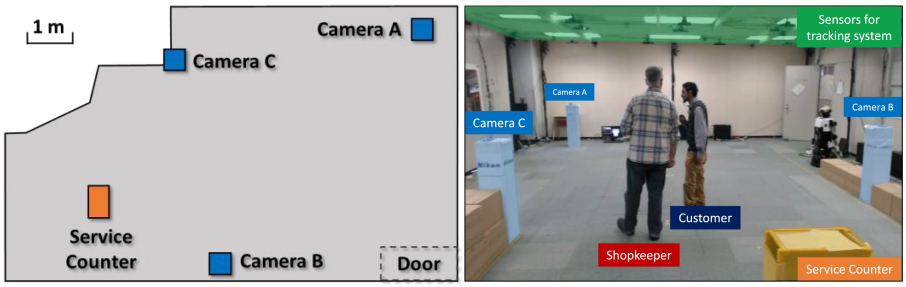
There is work in the area of active learning (Settles 2012) on how robots can increase their overall knowledge based on asking appropriate questions, choosing what questions to ask, and choosing when to ask things. For example, Cakmak and Thomaz identified the types of questions that a robot could ask to guide active learning of new skills, explored their use in human communication, and evaluated human perceptions of robots that ask such questions (Cakmak and Thomaz 2012). Thomason et al. has shown that an "inquisitive" robot that uses active learning to learn language-grounding functions is judged to be more fun to interact with when it asks off-topic questions (Thomason et al. 2017). The scenarios and goals in these works are different from our work, but there is overlap in choosing among robot actions to efficiently explore a space and improve a model based on interaction with humans.

3 HUMAN–HUMAN INTERACTION DATA COLLECTION

A dataset of human–human interactions, described below, was used for the purpose of training the data-driven robot behavior system, to be presented in Section 4. We used the dataset originally presented in Liu et al. (2017b).

3.1 Scenario

To observe typical interaction patterns, a camera shop environment was setup in an 8m × 11m experiment space with three camera models, each at a different location. For each interaction, one shopkeeper participant interacted with one customer participant. An interaction example is shown in Figure 2.



(Customer enters the shop and shopkeeper approaches)
C: I'm looking for a camera that is stylish, affordable, and easy to use. Do you have one of those?
S: Depends on what sort of pictures you take, we have two very stylish affordable cameras, one is $68 and one is over $500.
C: Oh I see, I think I might be looking to get the one that costs 68 dollars, can you please show me?
S: (Present Nikon) Sure, just over here this camera takes beautiful pictures but it's designed to be point-and-shoot, catch the moment.
C: Can you explain to me about the preset modes?
S: It has 18 different modes where you can tell it what sort of conditions you are in...
(Customer asks a few more questions about camera features and shopkeeper answers)

Fig. 2. An example interaction observed in our data collection.

3.2 Sensors

The participants' speech and position data were recorded as they interacted with each other. A sensor network with a human position tracking system recorded the participants' location data (Brscic et al. 2013). To capture the participants' speech, they were asked to speak into handheld smartphones running the Google speech recognition API. To detect the start and stop of speech activity, participants were required to touch the mobile screen to indicate the beginning and end of their speech.

3.3 Participants

Fluent English speakers were recruited as participants for the role of the customer. They had varied levels of knowledge about cameras. A total of 18 participants were employed (13 male, 5 female, average age 32.8, s.d. 12.4). To obtain a diverse set of behaviors, two different participants were chosen to role-play as the shopkeeper. They had very different interaction styles, one with a more outgoing personality (male, age 54) and another with a quieter disposition (female, age 25).

3.4 Procedure

The participants were encouraged to act naturally and focus discussion on the features listed on the camera spec sheets (8 to 10 features per camera). Customer participants were encouraged to play with the cameras, browse the shop, or ask camera-related questions. They role-played in different interactions as advanced or novice camera users to keep them interested and minimize fatigue. The shopkeeper participants were instructed to wait at the service counter at the start of the interaction, be polite, and behave according to their role (e.g., greetings and farewells, letting the customer browse, answering questions, or introducing products when appropriate). As the example in Figure 2 shows, participants used a variety of fillers (e.g., "you know" and "like") and backchannels (e.g., "I see") in their utterances. Therefore, we believe our setup elicited reasonably natural behaviors.

Each shopkeeper interacted with nine different customer participants. Each customer role-played 24 interactions (12 as advanced and 12 as novice) for a total of 432 interactions. Twenty-seven interactions were removed (16 due to technical failures and 11 due to customers who did

not follow instructions). In total, we collected 405 interactions, with 4,061 shopkeeper utterances and 4,115 customer utterances.[1]

4 PROPOSED TECHNIQUE

To develop a curious robot that continues to learn during interaction, we used a series of techniques that enable both behaviors and interaction logic to be directly learned from noisy sensor data without human intervention. The system architecture is shown in Figure 3.

First, using data abstraction techniques from previous work (Liu et al. 2016), the raw data were processed into a form suitable for input and output to two neural networks.

Then, although the behaviors of the robot are to be driven by curiosity, the robot should still follow the social rules observable in the human–human data. So we trained an *Appropriateness Learner* (the first neural network) on the processed human–human interaction data, for the purpose of constraining the robot's actions to a subset of socially appropriate actions that it can curiously explore in a particular situation.

Next, we developed a *Curiosity Learner* (the second neural network) to assign a curiosity score to each possible robot action. We model curiosity as the drive to minimize the variance of the prediction error of the consequence of the robot's actions. Similarly to the shopkeeper, a customer will generally behave within the norm of certain interaction patterns, which can be used as a prior for the *Curiosity Learner*. However, due to individual differences, different customers may react very differently to a given robot action, and it is these differences to which a curious robot must tailor its interaction dynamics. Thus, during live interaction, the *Curiosity Learner* is updated.

4.1 Data Abstraction and Representation

Behavior abstraction was applied to the raw interaction data to reduce the data's dimensionality and the effects of sensor noise, thus simplifying the learning problem. Behavior abstraction was accomplished by action segmentation, motion target and stopping location clustering, applying models of spatial formations, and action clustering, as originally presented in previous work (Liu et al. 2016). Here we describe how the data abstraction techniques were used to prepare the input vectors and training targets for the Appropriateness Learner and Curiosity Learner neural networks. An example of our abstracted representation is shown in Figure 4.

4.1.1 Motion and Spatial Formation Abstraction. The participants' common motion targets and stopping locations in the camera shop environment were discovered with unsupervised clustering. Spatial formations *presenting object, face-to-face*, and *waiting* were computed using pre-existing HRI models (in Yamaoka et al. (2008), Hall (1966), and Kitade et al. (2013), respectively). Below we refer to the formations as *spatial states*. Furthermore, a *state target* is specified for the *presenting object* formation, i.e., the object being presented.

4.1.2 Action Sequence Discretization. Each interaction was discretized into a sequence of alternating human and robot actions, with an action identified whenever a participant (1) speaks an utterance, (2) changes their motion target, or (3) yields their turn by allowing a period of time to elapse with no action, which we define as a yield action. We denote an interaction as $(h(t), r(t), h(t+1)\ldots)$, where $h(t)$ represents the human action and $r(t)$ represents the robot action at time t. Each action consists of the participant's speech, current location, movement arrival location, movement departure location, spatial state, and state target (in the case of *presenting object*).

4.1.3 Action Vectorization. Each action was represented as a vector for input to the *learners* and for action clustering. The vectorization procedure is the same as in Liu et al. (2016, 2017a).

[1]The dataset can be obtained here: http://www.geminoid.jp/dataset/camerashop/dataset-camerashop.htm.

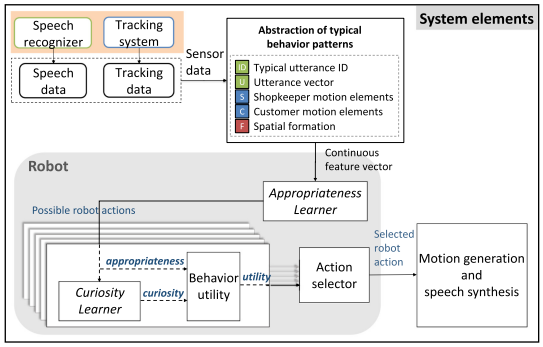


Fig. 3. Architecture of the proposed system elements.

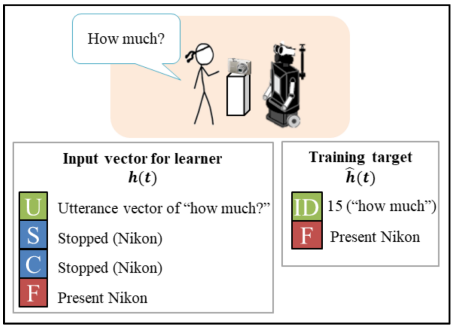


Fig. 4. Example of input vector and training target. Here, "S" and "C" denote shopkeeper and customer motion elements, "F" denotes spatial formation, and "ID" denotes typical utterance ID.

The human customer and shopkeeper utterances were vectorized using common text-processing techniques. First, stop words were removed from the utterances, a stemmer was applied, and n-grams were computed for $n = 1$, 2, and 3. Next, latent semantic analysis (LSA) (Landauer et al. 1998) was used to reduce the n-gram vectors to 1,000 dimensions. To emphasize important words, keywords were extracted from the utterances using a cloud-based API (now part of IBM Watson[2]) to create a separate keyword vector, which was then reduced to 200 dimensions using LSA. In this way, each utterance was represented with a 1,200-dimension vector. Customer and shopkeeper utterance vectorization were performed separately.

The shopkeeper and customer spatial information was vectorized into a vector containing the participant's location (7 dimensions), movement departure location (7 dimensions), and movement arrival location (7 dimensions). Furthermore, the shopkeeper and customer's joint spatial state, consisting of spatial formation (3 dimensions) and state target (4 dimensions), was also included. Thus, customer actions and shopkeeper actions were both 1,228 (m) dimensions.

4.1.4 Action Clustering. Each action was represented using a discrete ID, obtained by action clustering, for the output of the *learners*. The goal of action clustering was to group together similar actions from the human–human dataset to find sets of discrete customer and shopkeeper actions that commonly occurred during the interactions. The shopkeeper action cluster IDs were used as outputs for the Appropriateness Learner and the customer action cluster IDs were used as outputs for the Curiosity Learner.

Actions were clustered into clusters of similar actions by BIRCH clustering (Zhang et al. 1996). The number of clusters, K, was set to 800. The branching factor was set to 10 and the threshold was 0.005.

Since the system's output, a shopkeeper action cluster ID, must dictate the robot's behavior, which includes speech, destination location, and spatial formation, a *typical action* was automatically selected for each shopkeeper action cluster by finding the action that was most similar to all the other actions in the cluster (i.e., the medoid) using cosine distance. Since complete utterances with few ASR errors tended to share the most similarities with other utterances in the same cluster, the utterances of typical actions tended to be well formed and easy to understand. We refer to these as *typical utterances*. When the system outputs a shopkeeper action cluster, the robot speaks the *typical utterance* and moves based on the spatial portion of the *typical action*.

[2]https://www.ibm.com/watson/services/natural-language-understanding/.

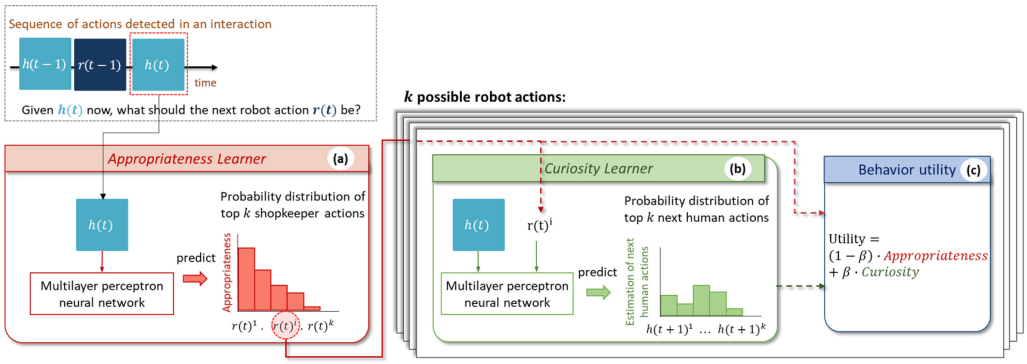


Fig. 5. Details of our system, both the Appropriateness Learner and Curiosity Learner is triggered when a human action (e.g., utterance or silence) is detected. (a) The Appropriateness Learner learns a set of k socially appropriate robot actions, and for each robot action (b) we query the Curiosity Learner to output a measure for curiosity; (c) finally, a utility measure is calculated.

The number of clusters, K, was set to be 800. This value was chosen to be larger than in previous work (Liu et al. 2016; Liu et al. 2017a) to obtain more fine-grained action clusters that account for more variability in speech. For example, a large number of shopkeeper clusters provides the shopkeeper with more possible subtle variations of behavior, such as saying, "This camera is five hundred dollars" versus "Let me tell you about the price of this camera, it's five hundred dollars." Moreover, a large number of customer clusters helps account for individual differences in phrasing, e.g., "What is the price?" versus "Tell me the price."

4.2 Learning from Example Interactions

Here, we describe the details of the individual components of the curiosity-based learning system, which is illustrated in Figure 5.

4.2.1 Appropriateness Learner. First, to learn the social appropriateness of a robot action, we applied a feed-forward multilayer perceptron neural network, which has the ability to learn a mapping, based on the relative importance of each input feature, to a discrete class, from examples in a dataset $\mathcal{D}$. Our training data for the neural network are composed of $(h(t), \hat{r}(t))$ action pairs, where $h(t) \in \mathbb{R}^m$ is the human action input vector and $\hat{r}(t) \in \{0,1\}^K$ is a target robot shopkeeper action, where K is equal to the total number of robot actions obtained from clustering. That is, if $\hat{r}(t)^i = 1$, then observation $h(t)$ maps to robot action i.

Based on the results from previous studies, we can interpret the neural network as learning a measure of how *appropriate* each robot action is (Liu et al. 2017a):

$$Appropriateness = p(r(t)^1), \ldots, p(r(t)^K). \tag{1}$$

During online interaction, we want to constrain the robot to a number of possible behaviors that are socially appropriate for a particular situation. Thus, we only select the top$-k$ most appropriate robot actions predicted by the neural network, among which the robot can freely explore using the *Curiosity Learner*.

4.2.2 Curiosity Learner. We model the curiosity of a robot action as trying to minimize the variance of the prediction error (Chan et al. 2015), i.e., the robot is curious about those actions for which it is uncertain how the customer will respond and less curious about actions for which it is confident it can predict what the customer will do next. Similarly to the *Appropriateness Learner*, we

can learn an initial estimation of potential next customer actions by applying a second multilayer perceptron neural network.

Considering a sequence of alternating actions $(h(t), r(t), h(t+1))$, the training input for the neural network is $(h(t), r(t))$, and the training target is the discretized value, $\hat{h}(t+1)$. The neural network learns a probability distribution over the set of human actions in the next timestep, $p(h(t+1)^1), \ldots, p(h(t+1)^K)$, where K is the total number of discrete human actions. Finally, only the top$-k$ most likely subsequent customer actions were used.

To measure the uncertainty of the customer's next action, we can calculate the entropy of the probability distribution that is output by the neural network (Wang 2008). Previous computational models have also incorporated such uncertainty-based strategies, generating biases toward actions or states that have high entropy (Cohn et al. 1996; Rothkopf and Ballard 2010). A high entropy value means that the robot is unsure what the customer will do as a result of its own action, while a low entropy value means that the robot is fairly confident of what the customer will do next. Within this framework, the robot is encouraged to take actions that result in states that are deemed surprising—i.e., actions for which the robot is unsure what the customer will do next.

Thus, the *curiosity* measure is the normalized entropy of the probability distribution:

$$Curiosity = \frac{-\sum_{i=1}^{k} p'(h(t+1)^i)\ln(p'(h(t+1)^i)}{\ln(k)}, \tag{2}$$

where $p'(h(t+1)^1), \ldots, p'(h(t+1)^k)$ is the probability distribution over the top$-k$ most likely subsequent customer actions, normalized to sum to 1.

4.2.3 Behavior Utility. For each potential robot action, a behavior utility function is evaluated, which combines the factors of *social appropriateness* and *curiosity*:

$$Utility = (1-\beta) \cdot Appropriateness + \beta \cdot Curiosity, \tag{3}$$

where β is a tuning parameter that is adjustable. A high β biases the robot to be more curious while a low β biases the robot to be more socially appropriate during interaction.

4.2.4 Action Selector. To select a behavior for a robot, the behavior utility function is evaluated for each of the potential actions the robot can perform. The action selector then executes the robot action with the highest utility, which is a discrete action, consisting of a typical utterance and a target spatial formation.

4.3 Adaptation to Individuals

As the robot continues to interact with a customer, it should come to better understand how the customer will respond to its actions. Thus, it will tend to be less curious about actions it has taken previously. To reflect this new observation in the *Curiosity Learner*, during live interaction, we can update the weights of the neural network in the *Curiosity Learner* through backpropagation. Backpropagation is used to modify the synaptic weights of the internal (hidden) and output layers of the neural network (Rojas 1996) by trying to minimize the loss between the target and the predicted value. In this way, the input-output mapping of the neural network can be dynamically updated to reflect new observations.

To update the weight of the neural network, the newly observed human action is first mapped to an action cluster, $\hat{h}(t)$, using the nearest-neighbor algorithm. Then, this action is used as the target for backpropagation, with the previous customer action $h(t-1)$ and robot action $r(t-1)$ as inputs. We used the cross-entropy function to compute the loss. To control how quickly the neural network learns the observed human action, we backpropagate the newly observed interaction data through the neural network over several epochs (to a maximum of l) until the cross-entropy loss

Algorithm 1 Robot action selection and online adaptation to customer action

Initialize β, th, l

When new human action $h(t)$ is detected:

// Compute the next robot action

$h(t) \leftarrow vectorize\big(h(t)\big)$ // Vectorize for input to the *Appropriateness Learner*

$r(t)^1, \ldots, r(t)^k \leftarrow AppropriatenessLearner\big(h(t)\big)$ for k robot actions with highest $appropriateness$

for $i \in [1, k]$ **do:** // For each of the most appropriate robot actions

  $r(t)^i \leftarrow vectorize(r(t)^i)$ // Vectorize for input to the *Curiosity Learner*

  $h(t+1)^1, \ldots, h(t+1)^k \leftarrow CuriosityLearner(h(t), r(t)^i)$ for k most likely next human actions

  $curiosity(r(t)^i) \leftarrow \frac{-\sum_{i=1}^{K} p(h(t+1)^i)\log(p(h(t+1)^i)}{ln(K)}$ // where $p(h(t+1)^i)$ is the prob. output by *Curiosity Learner*

  $utility(r(t)^i) \leftarrow (1-\beta) \cdot appropriateness(r(t)^i) + \beta \cdot curiosity(r(t)^i)$

$r(t) \leftarrow \max_{i \in k}(utility\ (r(t)^i))$ // Find the robot action with highest utility and execute

// Do the online learning to update the *Curiosity Learner*

$\hat{h}(t) \leftarrow NearestNeighbor$ of all possible K human actions to $h(t)$ // Find the most similar customer action cluster to $h(t)$

do: // Iteratively train *Curiosity Learner* until stopping condition is met

  Update *Curiosity Learner* by backpropagating with input ($previous\ customer\ input$, $previous\ robot\ input$) and training target $\hat{h}(t)$

  if $training\ loss < th$ or $num.iterations > l$: **break** // Update until training loss thresh. or max. num. iterations

$previous\ customer\ input \leftarrow h(t)$ // update for the next round of online learning

$previous\ robot\ input \leftarrow r(t)$ // "

Fig. 6. Pseudocode for the curiosity-based learning algorithm for robot action selection and online adaptation to the individual customer.

is below a certain threshold, th. This allows the recently observed human action to immediately become a much more likely prediction for that prompt.

Because the Curiosity Learner network is trained on only a single training example for a limited number of epochs, the online learning process does not impede the runtime performance. An analysis of the runtime performance is presented in Section 5.4.3.

The overall process for the curiosity-based learning system is described in the pseudocode in Figure 6.

4.4 Model Parameters

4.4.1 For Learning from the Human–Human Dataset. To find the ideal parameters for training the neural networks on the human–human interaction dataset, we iteratively tested different parameter values. The parameters were adjusted to improve the Appropriateness Learner's shopkeeper action prediction accuracy and the Curiosity Learner's customer action prediction accuracy.

The architectures of both neural networks in the Appropriateness Learner and the Curiosity Learner are the same, consisting of an input layer, followed by three leaky rectified hidden layers, and a softmax output layer. The input to the Appropriateness Learner is the human action vector of dimension m and the input to Curiosity Learner consists of both a human and a robot action vector with total dimension $2m$. Each hidden layer consists of 800 neurons.

Both neural networks were trained using momentum-based mini-batch stochastic gradient descent, with a batch size of 128, a learning rate of 0.0005, and a momentum coefficient of 0.9. Normalized initialization, described in Ioffe and Szegedy (2015), was used to initialize the neural network. The network was trained to minimize the cross-entropy loss for 2,000 epochs between the observed target action and the predicted action for the entire training set.

4.4.2 For Online Learning. The Curiosity Learner's parameters for online learning during interaction were tested in simulation, where a human user was able to choose the actions of the customer and observe the proposed system's responses.

We found 0.05 to be a good threshold, th, and 10 to be a good maximum number of iterations, l, for terminating the iterative backpropagation process. Setting th too high or l too low causes online learning to halt before the Curiosity Learner learns the customer's individual differences. Setting th too low or l too high causes overfitting to the most recently observed customer action, such that the network predicts that customer action regardless of the robot's previous action.

For behavior generation, we tested several values for k, the number of possible robot actions, and found $k = 5$ constrained the exploration space for curiosity to be within the realm of socially appropriate behaviors. Increasing k allows for a wider variety of robot behaviors but increases the rate of socially inappropriate behaviors.

We also tested several β values for the behavior utility function and found 0.9 to be a good balance for executing behavior that is both socially appropriate and curious. Detailed analysis of the influence of the parameter beta is provided in the appendix.

5 EVALUATION OF THE CURIOSITY LEARNER

We conducted an evaluation of the Curiosity Learner's ability to adapt to individual customer's behaviors.

There are two learning phases for the Curiosity Learner. The first learning phase is when the Curiosity Learner (and the Appropriateness Learner) are trained on the human–human dataset (Section 4.3) – i.e., offline learning. The second phase of learning occurs in *real time*, i.e., online learning, when the proposed system interacts with a real human customer (Section 4.4). Recall that the main task of the Curiosity Learner is to predict the customer's next action $h(t + 1)$ based on the current customer and shopkeeper actions $h(t), r(t)$. During the offline learning phase, the Curiosity Learner learns to predict the *most frequent* customer action that followed those customer and shopkeeper actions in the training dataset. However, during the second, online, learning phase, the Curiosity Learner learns to more accurately predict the actions of the *individual* customer who is currently interacting with the robot. This is accomplished by training the Curiosity Learner on the observed individual customer's actions, such that the probability of predicting the previously observed actions increases.

The Curiosity Learner's ability to learn from the customer's individual behaviors is critical for generating more varied, interesting interactions. When the Curiosity Learner's ability to predict the customer's behavior in response to certain robot actions improves, the robot's "curiosity" drives it to perform other actions, for which the customer's reaction is less predictable.

Thus, the evaluation described below focuses on the online learning phase. To more thoroughly evaluate the online learning phase, an additional interaction dataset was created via simulation. Subsequently, the Curiosity Learner's customer action prediction accuracy and the curiosity scores of the simulated robot actions were measured before and after online learning.

5.1 Simulated Interactions

We simulated interactions between a robot shopkeeper and a human customer to create a dataset for evaluation separate from the human–human dataset (Section 3), since the human–human dataset was used to train the Curiosity Learner in the first, offline, learning phase (Section 4.5). Interactions were simulated using a graphical user interface that displayed the customer and shopkeeper's locations on a map of the camera shop. A human user was able to control the customer's speech by text input and the customer's location by mouse click. The shopkeeper's actions (speech text, spatial state, and state target) were automatically selected in response to the customer's

actions by the Appropriateness Learner (Section 4.3) and displayed to the user. Using the simulation tool, 15 interactions were generated by a user playing various types of customers (a professional photographer, an art student, a frequent family-vacationer, a silent browsing customer, etc.). Two to three interactions were role played for each customer type. The 15 interactions contained 205 customer–robot action turns in total. The average interaction length was 13.7 turns (s.d. 2.9).

5.2 Simulated Evaluation Setup

The experiment consisted of comparing the Curiosity Learner's customer action predictions under two experimental conditions. Predictions were made for each of the 205 simulated customer actions, $h(t)$, given the context $h(t-1), r(t-1)$. The conditions were as follows:

- **No Adaptation:** The Curiosity Learner predicted the customer's actions *without online learning*.
- **With Adaptation:** The Curiosity Learner's predictions of all customer actions $h(t)$ in an interaction i were made after sequentially online-learning from each of $h(t-1), r(t-1) \rightarrow h(t)$ in interaction i. This condition shows the Curiosity Learner's capability to learn and remember the customer's behaviors over the entire duration of the interaction.

5.3 Evaluation Metrics

5.3.1 Customer Action Prediction Accuracy. To evaluate the effectiveness of online learning, the customer action prediction accuracy on each of the 205 customer actions was measured in both of the experimental conditions. To determine whether a predicted customer action $h'(t)$ matches the simulated ground-truth customer action, the ground-truth action is first matched to a customer action cluster $\hat{h}(t)$ using the nearest-neighbor classifier, as described in Section 4.3, and if $h'(t) = \hat{h}(t)$, then it is considered a correct prediction.

Furthermore, we looked at both whether the ground-truth-matching customer action $\hat{h}(t)$ was the Curiosity Learner's top prediction ("top-one accuracy") and whether it was among the top five predictions ("top-five accuracy").

5.3.2 Average Curiosity Scores. While the customer action prediction accuracy shows how well the Curiosity Learner learns about individual customers, the curiosity score, which is computed from the Curiosity Learner's predictions, has a more direct effect on the robot's behavior. When the curiosity score for a shopkeeper action is high, the robot is more likely to take that action. Furthermore, after observing a customer action $h(t)$ in response to a previous robot action $r(t-1)$, the curiosity score for $r(t-1)$ should decrease in the context of actions like $h(t-1)$, such that if the customer performs a similar action to $h(t-1)$ subsequently, the robot will be more likely to explore a different shopkeeper action. An example of this process is presented in Figure 11.

Average curiosity scores for both experimental conditions were computed over each of the 205 $h(t-1), r(t-1)$ action pairs in the simulated dataset. That is, each $h(t-1), r(t-1)$ was fed into the Curiosity Learner in both the *no adaptation* and *with adaptation* conditions, and curiosity scores were computed from the learners' outputs, in the form of probability distributions over the set of possible customer actions $p(h(t+1)^1), \ldots, p(h(t+1)^K)$, using Equation (2).

5.4 Evaluation Results

The evaluation's results are shown in Figures 7 and 8.

5.4.1 Customer Action Prediction Accuracy Improves After Adaptation. The results in Figure 7 show that the Curiosity Learner with adaptation, i.e., the online learning phase—based on individual customer behavior—was better able to predict the customer's actions. The Curiosity Learner

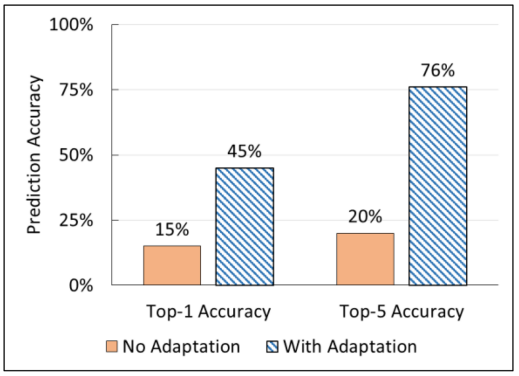


Fig. 7. The Curiosity Learner's customer action prediction accuracy with and without adaptation to individual differences.

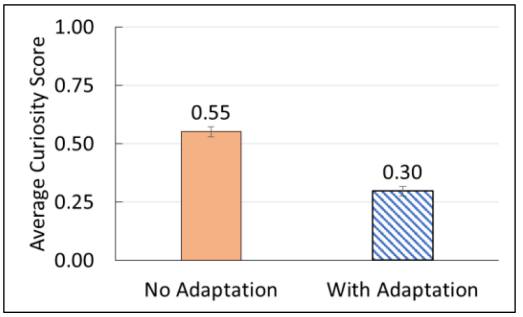


Fig. 8. Average curiosity scores of each simulated turn with and without adaptation.

had top-one accuracy of 45% with adaptation and 15% with no adaptation. The customer action prediction accuracy matters, because when the Curiosity Learner predicts a previously observed action with high probability, this leads to a low curiosity score for robot actions that might cause the customer to respond in the same way again. This enables the robot to explore different actions, resulting in more varied, interesting interactions. For example, this is the case for robot action R3 in the example interaction shown in Figure 11.

To compute the curiosity score, the top-k predictions are used. In our implementation, k was set to 5. Therefore, it is also useful to evaluate how often the correct customer action is among the Curiosity Learner's top five predictions. The Curiosity Learner had top-five accuracy of 76% with adaptation and 20% with no adaptation. This means that for about 76% of the instances the Curiosity Learner learned sufficiently to try performing a different, curious action if the robot were to encounter a similar situation again.

5.4.2 Curiosity Scores Decrease with Adaptation. Figure 8 shows how the average curiosity scores of the simulated robot actions $r(t)$ changed as a result of adaptation (with error bars showing the standard error). The average curiosity score without adaptation was 0.55 and dropped to 0.30 with adaptation. This is a consequence of learning the customer's action $h(t+1)$, as demonstrated in Section 5.4.1. When the Curiosity Learner adapts to the customer's behavior, it becomes more "confident" in its prediction of the customer action, which is reflected in a less uniform distribution over the possible customer actions. As a result, the entropy, and thus the curiosity score, is lower. This yields a lower utility for the action $r(t)$, allowing the robot to explore other actions instead.

5.4.3 Runtime Performance. The online learning is a very fast process. Each of the 205 simulated instances of online learning completed in an average of 55.7 ms (s.d. 33.1) when run with a GeForce GTX 980 Ti graphics card. The runtime performance during the user study, presented in Section 6, was similar and did not delay the robot's response time.

In summary, these evaluations demonstrate that the Curiosity Learner is able to effectively learn an individual customer's behaviors through interaction. During real time interaction, the Curiosity Learner's ability to learn is critically important, since curiosity scores are assigned to robot actions based on the learner's output. Thus, adaptation allows the proposed technique to vary its behavior based on what has been learned about the customer.

6 USER STUDY

6.1 Conditions

To observe the effect of the proposed techniques during live interaction, we conducted a user study to compare the two conditions:

- **Curious robot:** uses both the *Appropriateness Learner* and the *Curiosity Learner* for generating robot behaviors
- **Non-curious robot:** uses only the *Appropriateness Learner*.

The experiment used a within-participants design and the order of the conditions was counterbalanced to avoid ordering effects.

6.2 Hypothesis and Prediction

We made the following hypotheses about the effects of our proposed techniques:

- Driven by curiosity, the curious robot will perform more actions it has not taken before, mimicking humanlike-ness. Thus, it will be perceived as being more *humanlike*.
- The curious robot will be able to adapt its behaviors to some individual customers' differences, creating opportunities for interactions to develop in diverse ways and resulting in more *interesting* interactions.
- The curious robot will exercise curiosity within a set of appropriate actions, thus *social appropriateness* should be similar for both the non-curious and the curious robot.
- Providing a more individualized interaction will lead to more enjoyable experiences for the customer, thus the curious robot will be perceived to have *better interactions* overall.

6.3 Experiment Setup

6.3.1 Participants. A total of 16 paid participants (10 male and 6 female, average age 34.9, s.d. 9.0) were recruited for this experiment. All of them were fluent English speakers and had no previous familiarity with the robot.

6.3.2 Environment. The experiment was conducted in the same camera shop setting used for the human–human data collection, with three cameras displayed in an 8m × 11m experiment space. The same sensor network was used for tracking, and the participants communicated with the robot using a handheld smartphone for speech recognition.

6.3.3 Robot Platform. For this experiment, we used a humanoid robot with a 3-degree-of-freedom (DOF) head, two 4-DOF arms, and a wheeled base, capable of moving at a speed of 0.7m/s. We implemented the proposed techniques in the robot, enabling it to autonomously generate behavior based on inputs from the sensor network and speech recognition results. The dynamic window approach (DWA) was used to avoid obstacles (Fox et al. 1997), and the speech synthesis system described in Kawai et al. (2004) was used to generate utterances.

To make the interactions more natural, idling behavior was implemented in the robot for both conditions, in which the robot makes small arm and head movements while idling, speaking, and moving (Shi et al. 2010). Automatic head-tracking of the robot's interaction partner was also implemented, and the robot followed the customer with its gaze during all interactions.

6.3.4 Procedure. Since the goal of this study was to evaluate a curiosity-driven robot, we asked the participants to role-play as a customer and to observe and interact with the robot with the purpose of evaluating the quality of the interaction. Each participant interacted with both robots. To test the robot, we suggested the participants could ask about the same camera features twice,

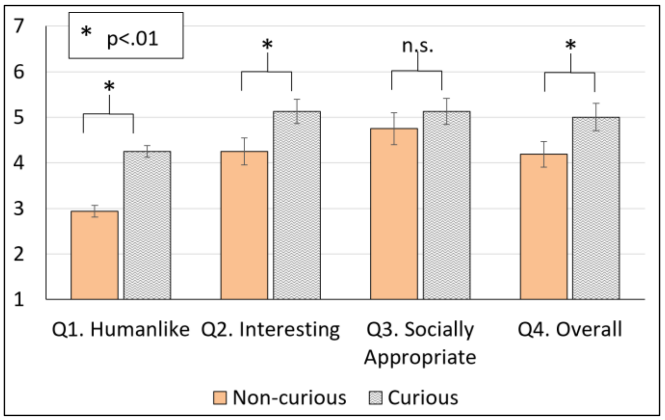


Fig. 9. User study questionnaire responses.

remain silent by playing with the camera, or acknowledge the robot with some simple utterances (e.g., "ok"). Participants freely chose when to end each interaction, by thanking the robot and leaving the shop.

As in our human–human data collection, before the start of the experiment, we explained to the participants what a typical interaction was like, asked them to become familiar with the smartphone interface, and confirmed their understanding of the instructions.

After each participant had interacted in one condition, a questionnaire was administered, and the procedure was repeated with the remaining condition (curious or non-curious). Half the participants interacted with the curious robot first then the non-curious robot, and the other half interacted in the opposite order.

6.4 Measurement

After each condition, participants filled out a written questionnaire, rating the following items on a 1–7 scale (1 being very negative and 7 being very positive):

- How humanlike did the robot's behavior seem to you, considering the repetitiveness and diversity of its behaviors?
- How interesting was your interaction with the robot?
- Was the robot's behavior socially appropriate for its role as a shopkeeper?
- Overall evaluation

After the questionnaire was completed, the participants were briefly interviewed.

6.5 Results

6.5.1 Questionnaire Results. To compare each rating between the curious and the non-curious robot, we conducted a paired t-test for each of the four questions, the results of which are shown in Figure 9.

We verified that all of our hypotheses were supported, as the analysis found significant differences between the conditions for ratings: Humanlike ($t(15) = 3.748$, $p = 0.002$), Interesting ($t(15) = 3.050$, $p = 0.008$), and Overall evaluation ($t(15) = 3.896$, $p = 0.001$). We did not find a significant difference for Socially appropriate ($t(15) = 1.695$, $p = 0.111$).

Thus, the results support our predictions: The curious robot was perceived to be more *humanlike* with respect to repetitiveness and diversity of behavior and more *interesting* than the non-curious

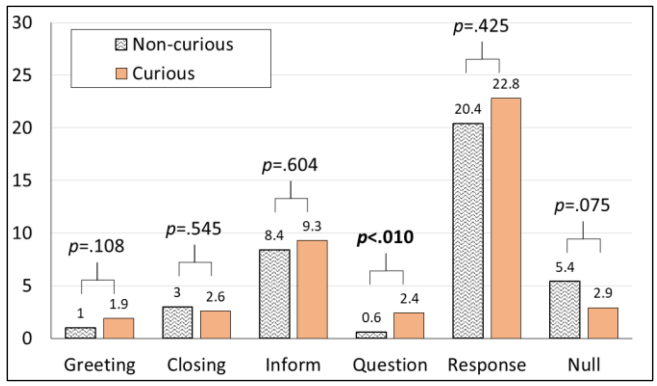


Fig. 10. Frequency of robot utterance types. The curious robot asks significantly more questions than the non-curious robot.

robot; there was no significant difference between the perceived *social appropriateness* between two robots; and the curious robot was rated to have a *better overall interaction* than the non-curious robot.

6.5.2 The Curious Robot Asks More Questions. We conducted an analysis of how many questions occurred in each condition of the user study (Figure 10). Each robot action in the user study logs from 10 of the experiment sessions (only 10 of 16 participants were analyzed due to failures in recording in the other 6 trials) was labeled by an expert annotator as either *greeting, closing, inform, question, response*, or *null* (actions where the robot remained silent—e.g., waiting at the service counter or silently standing by the customer). Then, the mean number of occurrences per interaction were computed for each condition and a *t*-test was performed.

The only utterance type to have a statistically significant difference between the two conditions was *question*. In the *non-curious* condition, the robot asked the customer an average of 0.6 questions per interaction. In the *curious* condition, the robot asked the customer an average of 2.4 questions per interaction. Thus, the *curious* robot asked four times as many questions as the *non-curious* robot.

Some of the questions asked by the curious robot were, "What sort of pictures do you like to take?," "What sort of camera did you have before?," and "What kind of camera are you looking for?" This may be one reason that participants observed more humanlike diversity of behavior in the curious robot and why they found those interactions to be more interesting.

6.5.3 Qualitative Observations. During the experiment we observed that the curious robot generated behaviors with a wider variety of actions, which improved the quality of the interactions in these situations: (1) when the customer repeatedly responded to the robot with backchannels, the robot provided a variety of information rather than repeating itself; (2) when the customer was silent, the robot attempted various behaviors rather than repeating the same behavior; and (3) when the customer repeated the same question more than once, the robot was able to respond with different phrasings. Considering the third case, in a real interaction a customer may repeat himself because he did not understand what the robot said the first time. In this case, it is helpful for the robot to try rephrasing its response.

The curious robot often performed several types of behaviors: (1) rephrasing a response; (2) elaborating on basic responses by providing additional information; (3) performing proactive behaviors, such as making camera recommendations or guiding the customer to different cameras; and

(4) asking the customers questions about their preferences. For these reasons, the curious robot appeared to be more engaging to the customers than the non-curious robot. Some participants stated that the curious robot had more salesmanlike qualities than the non-curious robot, and many participants preferred interacting with the curious robot.

We found that the curious robot spoke more unique utterances than the non-curious robot. On average, the curious robot spoke 30.6 (s.d. 5.5) unique utterances, while the non-curious robot only spoke 21.3 (s.d. 9.0) unique utterances ($t(9) = 3.274$, $p = .010$). To see if there was a difference in customer behavior, we analyzed the total number of utterances spoken by each customer. Although not significant, ($t(9) = 1.336$, $p = .214$), the customers spoke a slightly higher average of 32.4 utterances (s.d. 13.3) to the curious robot, compared with 27.7 utterances (s.d. 7.3) to the non-curious robot.

6.6 Case Study

Using the proposed techniques, we observed that the robot's behaviors appeared to be driven by curiosity. Here we present two examples.

6.6.1 Example 1. Figure 11 illustrates an example where the curious robot was able to continuously adapt to the customer's responses during live interaction (i.e., exploring different actions given the same customer input), rather than always using the same "default" interaction style learned from the off-line training.

At first, the customer is preoccupied with the Canon camera and yields her turn by remaining silent for a period of time. Once the robot detects a yield action (highlighted in blue), it queries the *Appropriateness Learner* to output the top fiv most appropriate robot actions, with an appropriateness value of around 0.05 for each of the five actions. For each robot action, the *Curiosity Learner* predicts a probability distribution for the top five subsequent customer actions, for which the *curiosity* value is computed. Here, Action R1 has the highest *curiosity* value of 0.567, yielding a total *utility* value of 0.205. The robot thus executes Action R1 and asks, "Can I ask what sort of pictures you take?"

The customer answers that she is looking to take pictures of family and friends, and this utterance is matched to the nearest customer action cluster. Once the robot observes this customer action, it updates the weights of the neural network in the *Curiosity Learner*. The robot then talks about the Canon camera, after which the customer ignores the robot and continues to remain silent (highlighted in brown). Given the same input, the *Appropriateness Learner* outputs the same five possible robot actions as before, but because the *Curiosity Learner* has been updated to reflect the customer's behavior, the customer action probabilities have changed, and the *curiosity* value for Action R1 has decreased. As a result, the robot decides to take Action R3, which now has the highest *utility* value among the possible actions, with a value of 0.301. The robot then introduces additional features of the Canon camera, "Full frame it's same size as a piece of thirty-five-mil film that's the standard for a top-end camera."

6.6.2 Example 2. Figure 12 illustrates an example of how the curious robot adapts to engaged and unengaged customers. Both customers perform the same action, but because the Curiosity Learner has adapted based on each customer's previous behaviors, it is able to provide differing actions, tailored to their individual differences.

The left side of Figure 12 shows the interaction history of an engaged customer, who is asking the robot many questions. In customer action C6, the engaged customer responds to the robot with a backchannel, "Okay." Based on the customer's previous actions, the Curiosity Learner has learned that he is likely to continue the conversation with any of a number of proactive actions.

Interaction example

(Present Canon)

C1: *(plays with the camera and remains silent)*
R1: Can I ask what sort of pictures you take?
C2: I am looking to take pictures of family and friends.
R2: It's the top end professional camera comes in black.
C3: *(plays with the camera and remains silent)*
R3: Full frame it's same size as a piece of 35 mil film that's the standard for a top end camera.

	Possible robot actions at C1	Uti	App	Cur	Possible customer actions at $h(t+1)$: C2
R1	**(Present Canon) Can I ask what sort of pictures you take?**	**0.205**	**0.051**	**0.567**	0.507 I'm going on vacation soon just looking for a small camera 0.054 I'm just looking for a basic camera to take selfies with 0.046 oh yeah actually it was but I need the previous model 0.046 okay thank you how 0.026 I'm looking for a camera that will take automatic pictures
	(Present Canon) Full frame it's same size as a piece of 35 mil film that's the standard for a top end camera.	0.052	0.049	0.060	0.970 (Customer silent) 0.008 ok 0.002 okay alright thank you have a look around 0.002 yeah I think you're right now 0.002 how much does this cost
	(Present Canon) Some people think it's a bit heavy it weighs one kilogram but that is what you get when you buy a top end camera.	0.037	0.053	0.000	...
	(Present Canon) You Can do everything in manual mode so it will also give you software style graph that you can manually set to adjust histograms and tone within the camera manually.	0.041	0.053	0.013	...
	(Present Canon) You have full creative control it has every possible manual setting	0.034	0.048	0.001	...

	Possible robot actions at C3	Uti	App	Cur	Possible customer actions at $h(t+1)$: C4
	(Present Canon) Can I ask what sort of pictures you take?	0.006	0.054	0.001	0.959 I'm looking for a camera to take pictures of the family 0.025 I'm going on vacation soon just looking for a small camera 0.002 oh yeah actually it was but I need the previous model 0.002 I'm looking to take pictures of trains 0.001 I'm looking for a camera that will take automatic pictures
R3	(Present Canon) Full frame it's same size as a piece of 35 mil film that's the standard for a top end camera.	**0.546**	**0.051**	**0.601**	0.275 I'm looking for a camera to take pictures of the family 0.110 ok 0.080 I'm going on vacation soon just looking for a small camera 0.074 okay alright thank you have a look around 0.064 how much does this cost
	⋮				

Fig. 11. An example interaction of the curiosity-driven robot. The columns (e.g., Uti, App, and Cur) are the respective *utility, appropriateness,* and *curiosity* values. The probability distribution of the next customer action (e.g., Prob C2), $h(t+1)$ is also shown. For brevity, the predicted customer actions are only shown for the relevant robot actions and only two of the five possible robot actions are shown for C3.

So, the robot responds to the backchannel phrase in an engaged way, "You might like to pick it up and try taking a few shots."

In contrast, the right side of Figure 12 shows the interaction history of an unengaged customer. In this interaction the robot tries offering information or asking questions, but is answered with short, disinterested responses, e.g., "Not sure" and "I see," or is ignored by the customer. Finally, when the customer responds to the robot's offered information with "Okay," the Curiosity Learner, having learned that the customer is likely to give some short response or to remain silent, assigns the highest curiosity score, 0.141, to the action "If you have any other question, please let me know" with the robot returning to the service counter. This action is predicted to possibly result in the customer going to exoplore the other cameras in the shop, so in that way it is the most curious action. Conversely, had the robot selected the same action as for the engaged customer, it would almost certainly have been met by silence from the customer.

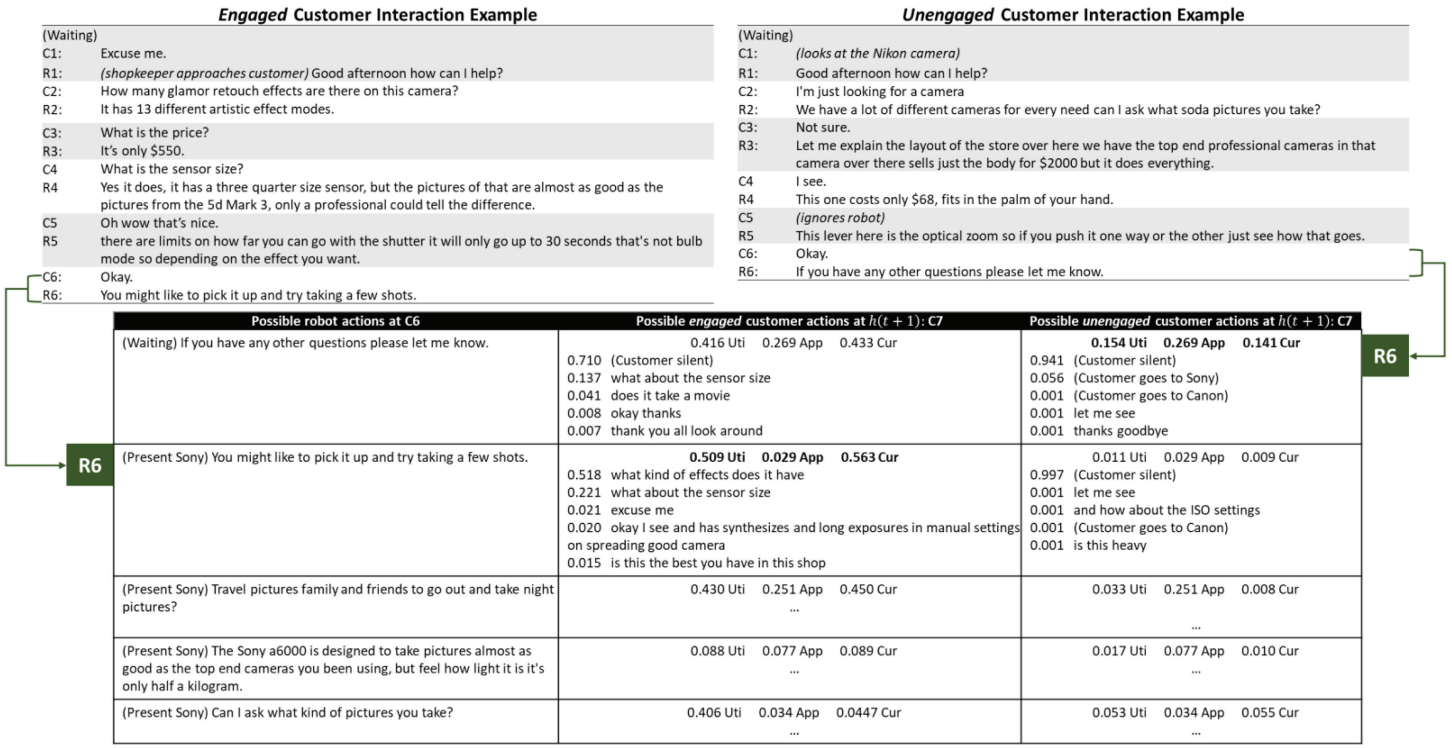


Fig. 12. An example interaction of the curiosity-driven robot. Uti, App, and Cur are the respective *utility, appropriateness* and *curiosity* values. The probability distribution of the next customer action (e.g. Prob C2), $h(t+1)$ is also shown. For brevity, the predicted customer actions are only shown for the relevant robot actions.

7 DISCUSSION

7.1 Curiosity and Individual Differences

We found that the "curious" robot was able to adapt its behaviors to some individual customer differences (e.g., interested versus uninterested customers) rather than always using the same default behaviors it learned from the off-line training. For example, when an uninterested customer continued to ignore the curious robot for some time, the robot would often go back to the service counter, saying "I will be at the service counter if you need any more help." While this was not the most proactive, salesmanlike behavior, it had the highest curiosity value for that particular situation, due to the fact that the robot had "lost curiosity" (i.e., the entropy-based curiosity score decreased) about previous actions (e.g., presenting features), since those actions did not elicit any unanticipated customer responses. In contrast, when the curious robot was interacting with an interested customer who had many questions, the robot would usually continue answering the customer's questions and would not leave the customer alone. Given the same situation, the non-curious robot would simply continue to respond with the same "default" behavior regardless of whether the customer was interested or uninterested, potentially resulting in a less ideal interaction than if it had adapted to the individual's needs.

7.2 Perception of Curiosity

The "curious" robot can only exhibit behaviors that are perceived as curious if such behaviors occurred in the human–human dataset, from which the robot learns. The camera shop interaction scenario, on which we demonstrated our proposed system, contains many curious behaviors, since the shopkeeper tried to learn about the customers. For example, *questions* are one type of behavior performed by the human shopkeeper that the robot learned to imitate. When the robot asks the customer a question, it is naturally perceived as being curious. If, however, the proposed system were to be applied to a scenario in which the target human is not curious, then it is unlikely that the robot would perform actions that exhibit curiosity.

However, it is possible that a robot trained on a dataset without curious behaviors can still learn about the humans it interacts with. This is because, at a fundamental level, the mechanism that drives the robot's behaviors will always result in robot actions that lead to uncertain human responses, such that the robot can learn more about the human. For example, sometimes the shopkeeper remaining silent and letting the customer take initiative creates a greater opportunity for the shopkeeper to learn about the customer, even though this might seem like a passive, uninterested behavior. Thus, there may be some benefit to applying the curious system to training datasets regardless of the quantity of curious behaviors they contain.

In our user study, we naturally wondered whether the attribute of "curiosity" of our robot was directly perceivable by the participants. Our interviews with the participants showed that some did perceive "curiosity" in the robot, while other participants could not tell the difference. Indeed, one common reason for the perception of "curiosity" in the robot was that the robot asked questions (e.g., "what sort of pictures do you like to take?"), which is one quantitative aspect of curiosity (Langevin 1971; Sinha et al. 2017). Our *Curiosity Learner* occasionally selects a robot utterance that is a question, and questions often lead to more variety in customer behavior. Note that our system does not have semantic understanding of which utterances constitute questions. For future work, it would be interesting to incorporate semantic understanding to better model curiosity for a conversational robot.

7.3 Generalizability

In this study, participants were instructed to focus only on camera-related conversation, which is not fully natural. It is uncertain how well the proposed techniques could generalize to more

natural conversation with a wider variety of topics. Currently, one limitation of our approach is that the robot cannot explicitly act on learned information, e.g., for the purpose of goal-directed behavior. However, we anticipate that additional mechanisms, such as incorporation of time series data, would enable a curious robot to exploit information it had learned about the customer earlier in the interaction (e.g., "Are you planning to travel soon?") and pursue specific goals (e.g., "This camera is great for travel").

In principle, we expect that the proposed approach can be applied to any domain characterized by repetitive, formulaic actions but also containing opportunities for individualized interaction (e.g., a waiter, receptionist, or travel agent). While the simple *Appropriateness Learner* can learn the repetitive behaviors, the *Curiosity Learner* can discover which behaviors are likely to lead to individual variation in customer responses. By guiding the interaction toward these behaviors, the curious robot creates opportunities for interactions to develop in diverse ways, opening up paths in the dialog that have the potential to branch out according to an individual's interests or needs. For example, a conversation with a curious robot that generates the curious action "What kind of pictures do you like to take?" may lead to topics relating to the customer's hobbies and end with the shopkeeper recommending a camera. In contrast, the non-curious robot would never have even asked the question in the first place.

8 CONCLUSION

In this work, we have presented a curiosity-based system for generating interactive behavior for a social robot. To the best of our knowledge, this study is the first to apply curiosity-based learning to this domain and to drive adaptation and individualization of dialogue. Our curious robot initially learns socially appropriate behavior by imitation from offline data and then continues to learn online, during live interaction with humans. The system adapts to customers' reactions in real time by choosing its own actions to explore and satisfy its curiosity about the customers' individual differences. In a user study we found that participants rated the curious robot to be significantly more humanlike with respect to repetitiveness and diversity of behavior, interesting, and better overall in comparison to a non-curious robot. The proposed techniques contribute a step toward advancement in the area of intrinsic motivation and curiosity-based learning for social robots. In future, it would be interesting to explore the application of the proposed technique on larger datasets and to other scenarios.

A APPENDIX

A.1 Analysis of the Curiosity Parameter

The curiosity parameter β is intended to control how curious the robot is when choosing an action. Specifically, it determines how much weight the robot places on the curiosity scores and social appropriateness scores of possible robot actions. We analyzed the effect of the curiosity parameter's value on the robot's behavior by comparing the robot's output action before and after adaptation at various curiosity parameter values.

A.1.1 Exploration Rate Metric. To measure the effect of the curiosity parameter value on the robot's behavior, we define *exploration rate* as the percent of instances where the robot's action in response to $h(t)$ without adaptation, $r_{no\ adaptation}(t)$, is different from its response to $h(t)$ after adaptation, $r_{no\ adaptation}(t)$. Formally,

$$exploration\ rate = \frac{\sum_{n=1}^{N} \sum_{t=1}^{numturns_n} (r_{no\ adaptation}(n,t) \neq r_{with\ adaptation}(n,t))}{\sum_{n=1}^{N} \sum_{t=1}^{numturns_n} 1}, \tag{4}$$

where N is the number of simulated interactions ($N = 15$), and $r_{no\ adaptation}(n, t)$ and $r_{with\ adaptation}(n, t)$ are the robot responses to customer action $h(t)$ at timestep t in interaction n. The *no adaptation* system used the Curiosity Learner before learning individual customer behaviors, and the *with adaptation* system used the Curiosity System that had learned about individual customer behaviors (as described in Section 5.2).

Thus, a high exploration rate means the robot changes its behavior more in response to learning about individual customer behaviors.

A.1.2 Effect of Curiosity Parameter on Exploration Rate. The exploration rate is the most direct method of measuring the effect of learning individual customer behaviors on the robot's behavior. The results in Figure 13 show that the effect of adaptation on the robot's actions increases as the curiosity parameter β increases. When $\beta = 0$, adaptation has no effect on the robot's behavior, because in this case the system does not use the output of the Curiosity Learner. But, when $\beta = 1.0$ the robot explores a different action in 77% of the instances after adaptation than it would have performed before adaptation. This demonstrates the effectiveness of the Curiosity Learner in adapting the robot's behaviors to individual customer behaviors.

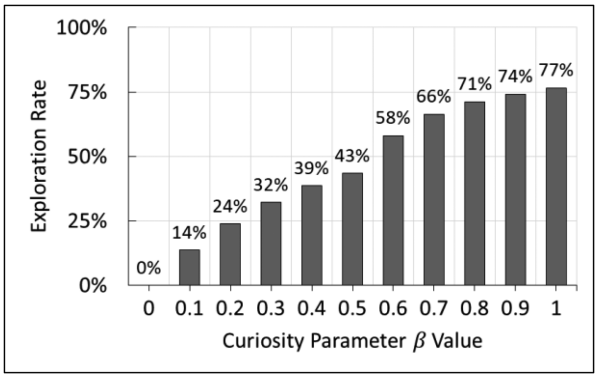


Fig. 13. The effect of the curiosity parameter on how frequently the robot changes is response to customer actions after adaptation.

REFERENCES

H. Admoni and B. Scassellati. 2014. Data-driven model of nonverbal behavior for socially assistive human-robot interactions. In *Proceedings of the ACM International Conference on Multimodal Interaction (ICMI'14)*. ACM, New York, NY, 196–199.

C. Breazeal, N. Depalma, J. Orkin, S. Chernova, and M. Jung. 2013. Crowdsourcing human-robot interaction: New methods and system evaluation in a public environment. *J. Hum.-Robot Interact.* 2, 1 (2013), 82–111.

A. Breuing and I. Wachsmuth. 2012. Let's talk topically with artificial agents! Providing agents with humanlike topic awareness in everyday dialog situations. In *Proceedings of the International Conference on Agents and Artificial Intelligence (ICAART'12)*.

D. Brscic, T. Kanda, T. Ikeda, and T. Miyashita. 2013. Person tracking in large public spaces using 3-D range sensors. *IEEE Trans. Hum.-Mach. Syst.* 43, 6 (2013), 522–534.

M. Cakmak and A. L. Thomaz. 2012. Designing robot learners that ask good questions. In *Proceedings of the 7th Annual ACM/IEEE International Conference on Human–Robot Interaction*. ACM, New York, NY, 17–24.

M. T. Chan, R. Gorbet, P. Beesley, and D. Kulič. 2015. Curiosity-based learning algorithm for distributed interactive sculptural systems. In *Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'15)*. IEEE, Los Alamitos, CA, 3435–3441.

C.-W. Chang, J.-H. Lee, P.-Y. Chao, C.-Y. Wang, and G.-D. Chen. 2010. Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *Educ. Technol. Soc.* 13, 2 (2010), 13–24.

S. Chernova, N. Depalma, E. Morant, and C. Breazeal. 2011. Crowdsourcing human-robot interaction: Application from virtual to physical worlds. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'11)*. IEEE, 21–26.

D. A. Cohn, Z. Ghahramani, and M. I. Jordan. 1996. Active learning with statistical models. *J. Artif. Intell. Res.* 4 (1996), 129–145.

M. E. Foster, S. Keizer, Z. Wang, and O. Lemon. 2012. Machine learning of social states and skills for multi-party human-robot interaction. In *Proceedings of the Workshop on Machine Learning for Interactive Systems (MLIS'12)*. 9.

D. Fox, W. Burgard, and S. Thrun. 1997. The dynamic window approach to collision avoidance. *IEEE Robot. Autom. Mag.* 4, 1 (1997), 23–33.

G. Gordon, C. Breazeal, and S. Engel. 2015. Can children catch curiosity from a social robot? In *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, New York, NY, 91–98.

E. T. Hall. 1966. *The Hidden Dimension*. Doubleday, Garden City, NY.

T. Hester and P. Stone. 2017. Intrinsically motivated model learning for developing curious robots. *Artif. Intell.* 247 (2017), 170–186.

S. Ioffe and C. Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the International Conference on Machine Learning*, 448–456.

K. Jokinen, H. Tanaka, and A. Yokoo. 1998. Context management with topics for spoken dialogue systems. In *Proceedings of the International Conference on Computational Linguistics (COLING'98)*. ACL, New York, NY, 631–637.

T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro. 2004. Interactive robots as social partners and peer tutors for children: A field trial. *Hum.-Comput. Interact.* 19, 1 (2004), 61–84.

F. Kaplan and P.-Y. Oudeyer. 2011. *From Hardware and Software to Kernels and Envelopes: A Concept Shift for Robotics, Developmental Psychology, and Brain Sciences*. Cambridge University Press, Cambridge.

H. Kawai, T. Toda, J. Ni, M. Tsuzaki, and K. Tokuda. 2004. XIMERA: A new TTS from ATR based on corpus-based technologies. In *Proceedings of the 5th ISCA Workshop on Speech Synthesis*.

T. Kitade, S. Satake, T. Kanda, and M. Imai. 2013. Understanding suitable locations for waiting. In *Proceedings of the 8th ACM/IEEE International Conference on Human–Robot Interaction* IEEE Press, Los Alamitos, CA, 57–64.

H. Kozima, M. P. Michalowski, and C. Nakagawa. 2009. Keepon. *Int. J. Soc. Robot.* 1, 1 (2009), 3–18.

T. K. Landauer, P. W. Foltz, and D. Laham. 1998. An introduction to latent semantic analysis. *Disc. Process.* 25, 2–3 (1998), 259–284.

R. Langevin. 1971. Is curiosity a unitary construct? *Can. J. Psychol.* 25, 4 (1971), 360.

E. Law, V. Cai, Q. F. Liu, S. Sasy, J. Goh, A. Blidaru, and D. Kulic. 2017. A wizard-of-oz study of curiosity in human-robot interaction. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'17)*.

I. Leite, A. Pereira, A. Funkhouser, B. Li, and J. F. Lehman. 2016. Semi-situated learning of verbal and nonverbal content for repeated human-robot interaction. In *Proceedings of the ACM International Conference on Multimodal Interaction (ICMI'14)*. ACM, New York, NY, 13–20.

P. Liu, D. F. Glas, T. Kanda, and H. Ishiguro. 2016. Data-driven HRI: Learning social behaviors by example from human–human interaction. *IEEE Trans. Robotics* 32 (2016), 988–1008.

P. Liu, D. F. Glas, T. Kanda, and H. Ishiguro. 2017a. Learning proactive behavior for interactive social robots. *Auton. Robots* 42, 5 (2017), 1067–1085.

P. Liu, D. F. Glas, T. Kanda, and H. Ishiguro. 2017b. Two demonstrators are better than one-a social robot that learns to imitate people with different interaction styles. *IEEE Trans. Cogn. Dev. Syst.* Early access.

K. Madani, C. Sabourin, and D. M. Ramík. 2016. Artificial curiosity emerging human-like behavior: Toward fully autonomous cognitive robots. In *Computational Intelligence* Springer, Berlin, 501–516.

M. P. Michalowski, S. Sabanovic, and H. Kozima. 2007. A dancing robot for rhythmic social interaction. In *Proceedings of the 2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI'07)*. IEEE, Los Alamitos, CA, 89–96.

S. Müller, S. Sprenger, and H.-M. Gross. 2014. Online adaptation of dialog strategies based on probabilistic planning. In *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'14)*. IEEE, Los Alamitos, CA, 692–697.

Y. Nagai, C. Muhl, and K. J. Rohlfing. 2008. Toward designing a robot that learns actions from parental demonstrations. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'08)*. IEEE, Los Alamitos, CA, 3545–3550.

P.-Y. Oudeyer, F. Kaplan, and V. V. Hafner. 2007. Intrinsic motivation systems for autonomous mental development. *IEEE Trans. Evol. Comput.* 11 (2007), 265–286.

R. P. Petrick and M. E. Foster. 2012. What would you like to drink? Recognising and planning with social states in a robot bartender domain. In *Proceedings of the International Workshop on Cognitive Robotics (CogRob'12)*.

A. H. Qureshi, Y. Nakamura, Y. Yoshikawa, and H. Ishiguro. 2018. Intrinsically motivated reinforcement learning for human–robot interaction in the real-world. *Neur. Netw.* 107 (2018), 23–33.

R. Rojas. 1996. The backpropagation algorithm. In *Neural Networks*. Springer, Berlin, 149–182.
C. A. Rothkopf and D. H. Ballard. 2010. Credit assignment in multiple goal embodied visuomotor behavior. *Front. Psychol.* 1, 173 (2010).
J. M. Santos, T. Krajník, and T. Duckett. 2017. Spatio-temporal exploration strategies for long-term autonomy of mobile robots. *Robot. Auton. Syst.* 88 (2017), 116–126.
J. Schmidhuber. 2013. Maximizing fun by creating data with easily reducible subjective complexity. In *Intrinsically Motivated Learning in Natural and Artificial Systems*. Springer, 95–128.
B. Settles. 2012. Active learning. *Synth. Lect. Artif. Intell. Mach. Learn.* 6, 1 (2012), 1–114.
K. Severinson-Eklundh, A. Green, and H. Hüttenrauch. 2003. Social and collaborative aspects of interaction with a service robot. *Robot. Auton. Syst.* 42, 3 (2003), 223–234.
C. Shi, T. Kanda, M. Shimada, F. Yamaoka, H. Ishiguro, and N. Hagita. 2010. Easy development of communicative behaviors in social robots. In *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'10)*. 5302–5309.
T. Sinha, Z. Bai, and J. Cassell. 2017. Curious minds wonder alike: Studying multimodal behavioral dynamics to design social scaffolding of curiosity. *Arxiv Preprint Arxiv:1705.00204*.
J. Thomason, A. Padmakumar, J. Sinapov, J. Hart, P. Stone, and R. J. Mooney. 2017. Opportunistic active learning for grounding natural language descriptions. In *Proceedings of the Conference on Robot Learning*. 67–76.
A. L. Thomaz and C. Breazeal. 2006. Reinforcement learning with human teachers: Evidence of feedback and guidance with implications for learning performance. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI'06)*. 1000–1005.
A. L. Thomaz and C. Breazeal. 2008. Teachable robots: Understanding human teaching behavior to build more effective robot learners. *Artif. Intell.* 172, 6–7 (2008), 716–737.
R. Toris, D. Kent, and S. Chernova. 2014. The robot management system: A framework for conducting human-robot interaction studies through crowdsourcing. *J. Hum.-Robot Interact.* 3, 2 (2014), 25–49.
R. Triebel, K. Arras, R. Alami, L. Beyer, S. Breuers, R. Chatila, M. Chetouani, D. Cremers, V. Evers, and M. Fiore. 2016. Spencer: A socially aware service robot for passenger guidance and help in busy airports. In *Field and Service Robotics*. Springer, Berlin, 607–622.
Q. A. Wang. 2008. Probability distribution and entropy as a measure of uncertainty. *J. Phys. A: Math. Theor.* 41, 6 (2008), 065004.
J. D. Williams, A. Raux, D. Ramachandran, and A. W. Black. 2013. The dialog state tracking challenge. In *Proceedings of the Special Interest Group on Discourse and Dialogue Conference (SIGdial'13)*. 404–413.
J. D. Williams and S. Young. 2007. Partially observable markov decision processes for spoken dialog systems. *Comput. Speech Lang.* 21, 2 (2007), 393–422.
F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita. 2008. How close?: Model of proximity control for information-presenting robots. In *Proceedings of the 3rd ACM/IEEE International Conference on Human–Robot Interaction*. ACM, New York, NY, 137–144.
J. E. Young, E. Sharlin, and T. Igarashi. 2013. Teaching robots style: Designing and evaluating style-by-demonstration for interactive robotic locomotion. *Hum.–Comput. Interact.* 28, 5 (2013), 379–416.
T. Zhang, R. Ramakrishnan, and M. Livny. 1996. BIRCH: An efficient data clustering method for very large databases. In *ACM Sigmod Record*, Vol. 25. ACM, New York, NY, 103–114.

Received June 2018; revised January 2019; accepted April 2019