



PDF Download
3772064.pdf
13 January 2026
Total Citations: 0
Total Downloads: 103

 Latest updates: <https://dl.acm.org/doi/10.1145/3772064>

RESEARCH-ARTICLE

How Can Robots Acquire Venue-Specific Social Strategies? The Case of Robot Clerks Learning Sales Negotiations

RYOTARO MURAMATSU, Kyoto University, Kyoto, Kyoto, Japan

DRAŽEN BRŠČIĆ, Kyoto University, Kyoto, Kyoto, Japan

T. KANDA, Kyoto University, Kyoto, Kyoto, Japan

Open Access Support provided by:

Kyoto University

Published: 22 December 2025

Online AM: 18 October 2025

Accepted: 30 September 2025

Revised: 17 September 2025

Received: 08 March 2024

[Citation in BibTeX format](#)

How Can Robots Acquire Venue-Specific Social Strategies? The Case of Robot Clerks Learning Sales Negotiations

RYOTARO MURAMATSU, DRAŽEN BRŠČIĆ, and TAKAYUKI KANDA, Kyoto University, Kyoto, Japan

Social robots are gradually being integrated into diverse social spheres, assuming multifaceted roles such as store clerks. Many of the interactive tasks entrusted to these robots are intricate, requiring more than simply responding to requests, and thus are not easily preprogrammed. One such task is sales negotiation. Empowering robots with negotiation skills is challenging due to the nuanced, context-dependent nature of effective tactics, which typically vary across different stores. To tackle this, we explore the potential of reinforcement learning without demonstrations for robots to grasp complex social interactions and adapt to specific environments, using sales negotiation as a test case. Drawing insights from interviews with experienced sales clerks and observations of negotiation role-plays, we identified essential verbal and nonverbal features. Coupled with a fast-learning reinforcement method, we evaluated the system through interactions with human participants in two distinct store settings. Results show that the robot was able to acquire negotiation strategies tailored to each environment.

CCS Concepts: • **Human-centered computing** → *Empirical studies in HCI*; • **Computer systems organization** → *Robotics*;

Additional Key Words and Phrases: shopkeeper robot, reinforcement learning, sales negotiation

ACM Reference format:

Ryotaro Muramatsu, Dražen Bršćić, and Takayuki Kanda. 2025. How Can Robots Acquire Venue-Specific Social Strategies? The Case of Robot Clerks Learning Sales Negotiations. *ACM Trans. Hum.-Robot Interact.* 15, 2, Article 30 (December 2025), 21 pages.
<https://doi.org/10.1145/3772064>

1 Introduction

Improvements in robot technology are gradually leading to the practical application and real-world operation of social robots. Robots have been introduced at corporate reception desks [33, 49] or tourist information centers [47], and their use is expected to expand to a variety of settings in the future to help alleviate labor shortages. Currently, practical applications of robots mostly involve answering human questions or performing predefined tasks. For example, the robot may offer to give directions to visitors [5, 15], guide during museum tours [11, 34], or carry objects to designated locations [23].

This work was supported by JST, Japan (Moonshot R&D Grant JPMJMS2011) and by JSPS, Japan (KAKENHI Grants JP24H00722 and JP23K11271).

Authors' Contact Information: Ryotaro Muramatsu, Kyoto University, Kyoto, Japan; e-mail: muramatsu.ryotaro.84m@kyoto-u.jp; Dražen Bršćić (corresponding author), Kyoto University, Kyoto, Japan; e-mail: drazen@i.kyoto-u.ac.jp; Takayuki Kanda, Kyoto University, Kyoto, Japan; e-mail: kanda@i.kyoto-u.ac.jp.



This work is licensed under Creative Commons Attribution International 4.0.

© 2025 Copyright held by the owner/author(s).

ACM 2573-9522/2025/12-ART30

<https://doi.org/10.1145/3772064>

However, many interactive service tasks go beyond mere responses to verbal requests. They require a good understanding of the context of the situation and the ability to recognize the user's state, which often means being perceptive of implicit nonverbal signals from the user. Because of that, the skills for proficient execution of such tasks are often hard to articulate and teach to others. Instead, learning through hands-on practice is necessary for effective acquisition.

Sales negotiation is a representative example of such a task. In many retail stores, store employees are given the freedom to negotiate with customers and encourage them to buy by offering slightly discounted prices or other benefits [21, 42]. However, sales negotiation is not only about meeting customer demands. Stores cannot always comply with every customer's bargaining requests as it may not be profitable to do so. Instead, a clerk's negotiation needs to be based on an estimate of the extent to which the customer is willing to compromise on the price while still making the purchase. The bargaining request does not need to come from the customer; the store can offer discounts or other benefits to encourage hesitant customers to buy products, resulting in higher sales. For efficient sales negotiations, a clerk also often needs to undertake negotiation even if not asked to by the customers and may sometimes even refuse customer requests, all of which cannot be done by simply responding to human input.

The art of sales negotiation is not something that can be easily mastered. Even though expert negotiators can offer advice to novices, it typically requires practice to acquire the skill. Moreover, effective negotiation strategies vary from environment to environment. Each store handles different products and has a different clientele, and negotiation strategies reflect that. For example, stores that sell expensive products like home appliances or cars often negotiate prices more aggressively than stores that sell less expensive goods. Sales negotiations may also depend on culture and customs and vary between geographical areas. Therefore, programming a robot to handle negotiations would have to be done separately for each store, which is time-consuming and labor-intensive if done manually.

To automatically learn negotiation techniques appropriate for each store, a natural choice is to learn from experience using **Reinforcement Learning (RL)** [44]. In particular, task-performance-driven RL [1] allows a robot to iteratively improve and gradually learn suitable strategies for the environment it operates in, purely based on robot performance and without relying on demonstrations and explicit instructions from a teacher. Thus, to address the challenge of enabling robots to engage in effective sales negotiations, we designed a negotiation system for robot clerks based on task-performance-driven RL.

The main contributions of the work are:

- We design and implement a complete system for learning a complex and venue-dependent social task, integrating sensing, action selection, and learning. This includes discerning the necessary sensing features and robot actions by leveraging insights from human experts, and identifying effective learning methods using simulations.
- More specifically, we develop a system that allows robot clerks to learn how to successfully conduct sales negotiations in face-to-face interactions with customers. As such, the robots can acquire negotiation strategies that are adapted to the environment in which they are deployed. Unlike existing text-based negotiation systems, the system incorporates not only verbal information but also nonverbal features. Moreover, we make the learning sufficiently fast to make it viable for actual deployment in stores.
- To validate the system's adaptability, we evaluate the learning of sales negotiation from scratch in two diverse mock store settings with human participants. By manipulating the type of clientele with varying negotiation preferences, we confirmed that the system was able to acquire negotiation strategies that were appropriate to each store. (Notably, such testing of social interaction learning in multiple settings is still rare due to its inherent complexity.)

Together, these contributions demonstrate a complete system for real-time, socially situated learning, rather than proposing a new algorithm.

2 Related Work

2.1 Agents for Sales Negotiation

Gratch et al. [16] argued that negotiation tasks like bargaining represent an important challenge for autonomous agents as they involve three essential skills for interaction with humans: intelligence (such as reasoning), language, and embodiment (such as nonverbal communication). Still, the bulk of research on agent negotiation focused on dialog-based agents or online chatbots, and therefore only the first two skills. For example, textual chat systems that enable smooth price negotiation with people using natural language as input have been studied extensively, and many of these works focused on learning (non-embodied) negotiation strategies [25]. He et al. [18] devised a chat system using a dataset of price negotiation conversations between sellers and buyers in an online store, and Parvaneh et al. [35] created a price negotiation chat system that estimates the appropriate price based on product photos and information. The importance of detecting and considering (textual) nonverbal information from partners for negotiation was investigated in [6].

Fewer works considered sales negotiation with embodied agents, such as robots or virtual agents. Aydoğan et al. proposed to let robots negotiate prices by deploying typical speech-based strategies [3] or extending them also to use expressive gestures [4]. Other works looked at embedded nonverbal behavior of a negotiating virtual agent, such as turn-taking skills DeVault et al. [12], the influence of the agent's voice [29], or politeness strategies [45]. In these works the negotiation strategies were mostly manually designed and not learned, as done in this study. While agents that autonomously learn how to negotiate have also been studied [14, 19, 36], these works were mainly limited to text as the medium for interaction and did not take into account human behavior other than linguistic information. In contrast, we used a robot to learn face-to-face negotiations with a customer, utilizing both verbal and nonverbal features.

Since negotiation inherently involves persuasive intent, our work also connects to research on persuasive strategies used by social robots. Prior studies have shown how robots can influence human decisions through combinations of gaze, gesture, vocal tone, and timing [9, 17]. Winkle et al. [48] conducted an empirical study on persuasive strategies in socially assistive robots, finding that strategies such as goodwill and similarity influenced user compliance. Our work builds on this line by exploring how such strategies can be learned and adapted in real time to different social contexts, rather than being pre-scripted.

Other service robot domains—such as robot waiters, receptionists, and delivery agents—have explored human-facing interactions, often focusing on task execution, polite dialogue, or long-term engagement [8, 15, 43]. Our work complements these by addressing the learning of adaptive persuasive strategies in face-to-face negotiation.

2.2 Learning of Robot's Social Interactions

Many works studied the learning of cooperative and interactive behaviors between robots and humans. One possible approach to do that is through learning from demonstration, where a human demonstrates the relevant action which the robot then imitates [2, 40]. While this is frequently applied to teaching a robot's movements or physical interactions (e.g., [24]), it can also be specifically applied to learning socially interactive behaviors [7, 28]. Yet, providing efficient demonstrations for an intricate and context-dependent task such as negotiation is often difficult to do. In some cases, it could be easier or more natural to provide demonstrations by a human teleoperating the robot, as in [22, 41], but even that has challenges, in particular the need for appropriate teleoperation interfaces.

A related approach to acquiring interaction skills is to use imitation learning of previously collected human–human interaction data, which has been successfully used for a robot to learn, e.g., shopkeeper’s interactive behaviors in [13, 27, 32]. The bottleneck of that approach is the need to collect a sufficient amount of human–human data in advance, which can often be challenging.

As an alternative, learning social interactions between a robot and a human using RL [44] has proven effective in many studies [1]. With RL the robot does not need human demonstrations or pre-collected datasets; instead, it learns while doing its task, gradually improving its performance. RL approaches have been shown to work in real-world **Human–Robot Interactions (HRIs)**. For instance, Qureshi et al. [37, 38] conducted experiments in which a social robot deployed in a public space learned appropriate ways to perform interactive behaviors, such as doing handshakes, and Raggioli et al. [39] used RL to adapt the navigation behavior of a social robot during interactions with elderly individuals. Notably, the ability of RL to adapt the agent’s behavior to the specifics of its environment makes it suitable for acquiring social strategies that are venue-specific. However, we are not aware of works that evaluated this ability by deploying the learning social robot in more than one environment, which is what we did in this work.

3 How People Negotiate

To our knowledge, human–human sales negotiations have never been described in the literature in enough detail such that we could simply derive the rules from there. Therefore, to understand how people conduct sales negotiations, we first conducted interviews with actual store clerks to investigate how they negotiate prices. This helped us obtain insight into what clerks are aware of when negotiating and what strategies they use. Building upon that, we ran role-play experiments on sales negotiation in a laboratory environment where we were able to observe the specific flow of the interaction including the prominent features, as well as the actions that are typically undertaken.

3.1 Interviews with Experts

We interviewed eight people who had experience working as store clerks at electronics retailers. Their ages ranged from 26 to 52 ($M = 36.0$, $SD = 10.3$), with sales experience ranging from 2 to over 10 years. There were six male and two female participants. They all worked in different stores. They were recruited nationwide through a recruiting agency and interviewed online.

The interview was semi-structured. In addition to questions about which store they worked at and for how long, we asked them about customer behavior during shopping and what they pay attention to when serving them (e.g., reactions or expressions), as well as about their experience with price negotiations: how often they occur and what is the process.

First, their responses to the question of how many customers negotiate prices revealed differences between store clientele, which seemed to depend on the store location. In particular, two interviewees indicated that in the Kansai region of Japan, where our institution is located and which is reputed for people’s willingness to haggle, around 70%–80% of customers negotiated prices, and that this percentage was lower in other regions in which they worked. One respondent also observed differences within a geographic region, stating that, in areas with high land prices, customers bargained less (about 50%) than in other areas where they had worked (about 80%). Another respondent, who had worked in multiple stores in a rural region, reported that 95% of the customers negotiated and there was not much difference between stores.

The interviewees also mentioned that the required negotiation strategies from the store staff differed depending on the store and clientele. For example, in stores where few people asked for discounts, the stores also offered fewer discounts. In rural regions, where the number of customers is small, stores place importance on repeat customers, so they sometimes offer discounts even when it results in losses.

When asked about the strategies used for sales negotiation, none of the interviewees thought they could provide a simple set of rules to use. However, the majority mentioned that they tried to gauge the customers' willingness to buy: if customers did not seem likely to buy, they spent less time with them. Conversely, when dealing with customers displaying a genuine interest in buying, the interviewees actively offered discounts and services.

Since inferring the customers' intentions to purchase seemed to be important for negotiations, we also asked the interviewees about how they derived these inferences. They responded that, once accustomed to serving customers, they become adept at estimating a customer's purchase intent based on the customer's demeanor. Six respondents explicitly mentioned nonverbal information as a means of predicting intent to purchase. Three cited the customers' reactions to greetings and product descriptions, eye contact, and body orientation as factors to measure purchase intent. Two also talked about the customer's attire, the time spent looking at a single product, and demographics. For example, customers who responded well to a greeting or question were highly motivated to buy. Customers who were determined to purchase a product did not take much time to decide on a product and, if they looked at one product for a long time, they were likely to purchase it. Customers who were not interested in a product were more likely to look away and, if their face and body were facing the clerk, it was probable that they would buy the product. Finally, well-dressed people and families were more likely to buy.

In summary, the interviews have clearly shown that the required negotiation strategies depend on the store and the type of customers that frequent it. Moreover, we were able to identify the important elements for conducting sales negotiations. In particular, if we exclude the information that is hard to detect from a robot (e.g., specific behavioral responses to greetings) or that is strongly dependent on a store (e.g., the customer's dressing style), the features that appear to be most informative about the state of the negotiation process are the following: (1) the customer's previous utterances, (2) where they are facing, and (3) how quickly they are to respond. In addition, (4) how much discount has been offered is obviously another important factor affecting the future flow of the negotiations.

3.2 In-Lab Sales Negotiations Role-Play

While the interview results gave us a general idea of the factors that are important in negotiations, this was still not enough to fully specify the features and robot actions to be used by the learning system. Therefore, we additionally ran a series of simple observation experiments of sales negotiations in mock store settings prepared in the lab. This scenario-based approach aligns with prior work in HRI that has used role-playing and imaginative design to support the early development of robot behaviors [10].

The rationale for using role-play in the lab was that it allowed us to easily observe the participants engaged in uninterrupted one-on-one sales negotiation. Achieving such isolation in actual stores would be much more difficult since they typically sell multiple similar products, the shopkeepers have several tasks they do, and customers often come in groups. For example, when a customer is undecided about which product to buy, the bargaining can become entangled with product selection. The controlled lab environment allowed us to circumvent these complexities.

First, we observed negotiations between humans. The setup of the experiment was the negotiation during the purchase of a PC at a retail store. The store clerk was played by a member of our lab. Another two colleagues acted as customers and played one of four customer types: "customer with sufficient budget," "customer with a small budget who bargains aggressively," "customer with a small budget who is reluctant to bargain," or "customer with little intention of buying." Second, we conducted an experiment with the same setup but with a robot serving as the clerk. The robot was remotely controlled by the same colleague who played the clerk in the former experiment, whereas three colleagues from our laboratory played out the various types of customers.

The colleagues who participated were not familiar with the purpose of our study. They were told to imagine that they were interested in buying one specific product. The customer types (such as “with sufficient budget,” Section 3.2) were manipulated by defining the product price and customer’s budget, and by asking the participants to imagine the specified propensity for bargaining.

One of the authors performed a visual analysis of the video recordings of the experiment to identify features of the customer’s demeanor that prominently changed with the customer type. It was found that: (1) When customers are wondering whether to make a purchase, the time to respond to the clerk tends to be longer. (2) Customers with low purchase intent tend to look at the clerk less. In addition to these features, based on the videos, we established that the clerks’ action choice also largely depends on the action that the customer made as well as the current discount amount.

In addition, we found that the negotiations-related actions of the customer can be divided into a few main categories, such as asking for a discount. Similarly, the clerk’s actions were also limited to a few types, such as offering a small discount or refusing to bargain. Based on these observations we defined the RL states and actions, with the full lists given below in Sections 4.3 and 4.4.

4 Negotiation Learning Method

4.1 Overview

In RL, a learning agent performs actions according to the state of the environment and is rewarded for these actions. Through repeated trials, the agent gains experience, gradually learning which actions to take under which conditions and thus acquiring a strategy that yields high rewards. Therefore, to define the learning task, we must define the features representing the *state* S , the *actions* A that can be taken by the agent, and the *rewards* R for those actions before learning.

4.2 Rewards

Since RL optimizes the action choices to obtain the highest rewards, the reward needs to match the learning agent’s objectives. A reasonable choice for sales negotiation, used also in previous research such as [20], is to use the profit that the store gained from a sold product. If the product does not sell, the reward will be zero. If a discount is made during negotiations, the profit will be reduced accordingly.

Assuming that the price at which the product was sold is $PRICE$, and the cost that the store paid for it is $VALUE$, the basic profit can be expressed as $(PRICE - VALUE)$ monetary units. Since prolonged exchanges with a customer can be considered a lost opportunity to serve other customers, it is preferable to take fewer steps during negotiations. This can be encoded by subtracting a specific fixed value $STPEPCOST$ from the reward for each negotiation step. Thus, if the negotiation episode took N steps, the total accumulated reward R will be:

$$R = \begin{cases} (PRICE - VALUE) - STEPCOST \times N & \text{if sold} \\ -STEPCOST \times N & \text{otherwise} \end{cases} \quad (1)$$

4.3 States

Based on the interviews and observations of human–human negotiations described in the previous sections, we defined four features that represent the state used in RL for sales negotiations:

- *Customer’s action* describes the last utterance that the customer made, and can take one of the following values:
 - *haggle*: Asking for a discount;
 - *complain*: Complaining about the price;
 - *think*: Taking time to think;
 - *demand*: Expressing dissatisfaction about things other than the price;

- *buy*: Purchase and end the conversation;
- *quit*: End the conversation without purchase.
- *Discount* encodes the current amount of the offered discount. We implemented the maximum discount to be offered as the amount at which the store gets zero profit, i.e., $(PRICE - VALUE)$. The discount is offered in steps, with each step being $1/5$ of the maximum discount, i.e., $(PRICE - VALUE)/5$. The discount feature counts the number of discount steps, so it takes the values from zero (no discount offered) to five (maximum discount).
- *Interval* describes the length of the pause between the end of the last robot's utterance and the customer's speech, and it can be either "short" or "long."
- *Gaze* is the direction in which the customer is looking. It can take one of two values: "robot" or "other."

The first two features can be extracted from verbal data, whereas the other two contain nonverbal information. While the value of Discount is updated internally based on the previous actions of the robot, the other three features come from the observations of the environment. The details are described in Section 5.2.

4.4 Actions

The set of actions that the learning system can choose for the robot was also decided based on the interviews with clerks and the role-play experiments. They are as follows:

- *small_discount*: Offer discount by a one discount step ($1/5$ th of the maximum discount);
- *large_discount*: Offer the maximum discount;
- *appeal*: Appeal with something other than price, such as special offers and points;
- *refuse*: Refuse the customer's bargain request;
- *answer*: Answer customer's question.

4.5 Choice of Learning Method

The robot is expected to gain experience while serving actual customers in a store. Since RL involves repeated trial and error, the robot clerk's sales performance during learning will typically be lower. To minimize lost profits, it is desirable to use a learning method that can acquire the appropriate negotiation strategies as quickly as possible.

The interviewed experts (Section 3.1) reported that based on their experience in electronics retail stores, there are on average about 10 negotiations per day. They also estimated that it takes 1–2 months to go from being a novice in negotiation to becoming an expert. Therefore, we expected that a novice robot clerk should also be able to learn to negotiate well in about 1 or 2 months, or approximately 200–500 negotiation interactions with customers. In addition to the achieved profit, this was our main criterion for choosing the learning method to use.

4.5.1 Candidate Methods.

Q-Learning. One method we tested was simple tabular Q-learning [44]. The basic principle of Q-learning is that for each combination of state s and action a it tries to estimate the value of $Q(s, a)$, which is a measure of the expected total reward R . After learning, for any specific state, the agent will choose the action giving the largest Q value. The Q values are kept in a table.

During training, the robot followed an epsilon-greedy action selection strategy with epsilon set to 0.2 to ensure sufficient exploration. We used a reward discount rate of 0.95 and a learning rate of 0.5.

Deep Q-Network (DQN). We also tested a DQN method [31], which has been successfully applied in previous research on (online) negotiation [35]. Instead of using a table for storing the Q values

for each state, in DQN they are approximated using a neural network (named a Q-network). We evaluated a number of network architectures and finally used a fully connected network, with two hidden layers containing 256 neurons and ReLU (rectified linear unit) activations, and a Softmax output layer.

Experience Replay (ER). ER [26, 31] is a mechanism for improving the efficiency of RL. It consists of storing previous experiences (combinations of states, chosen actions, and received rewards) in a buffer, and later randomly sampling part of the experiences and re-using them for learning.

We created both Q-learning agents and DQN agents with ER. In addition to the standard learning during the serving of customers, for every five customers served, 960 experiences were randomly extracted (possibly with repetitions) from the buffer and used as the target for learning.

4.5.2 Evaluation of Negotiation Performance in Simulation. To evaluate the described learning agents, we prepared a simulation of store negotiations. Analogously to the data collected in the role-play negotiation experiments in Section 3.2, we created customer agents that reproduced the behavior of several types of customers: “customer with a small budget who bargains aggressively” (two different implementations), “customer with a small budget who is reluctant to bargain” (two different implementations), and “customer with little intention of buying” (one implementation).

Each customer agent was implemented using IF-THEN rules. For example, in one implementation, the “customer with a small budget who bargains aggressively” initially takes the action *haggle*. If they get a *discount* or *appeal*, they will randomly choose either the action *haggle* or *think*; otherwise, they will choose the action *complain* or *quit*. If they are refused further discounts, they will purchase. They will also purchase when there are five levels of discounting. Randomness was added to the action selection and the conversation intervals to allow for variation in the data.

We ran simulated negotiations with all tested learning agents and examined how their performance changed over time. For each negotiation interaction, the customer type was chosen at random and the learning agent interacted with them. For every 20 served customers, we ran an evaluation with the agents using the strategy learned until that point. In the evaluation, they negotiated with 1,000 customers (200 for each implemented version), whose order was randomized and kept fixed through all evaluations. The average profit from the evaluation was used to measure the performance. For each agent, the evaluation was repeated twenty times and the average performance was saved.

As simple references for the evaluation results, we considered several fixed-action strategies. A strategy that only uses the *appeal* action, evaluated on the same 1,000 customers as above, achieves an average profit of approximately 86. By contrast, a strategy that always offers a *small_discount* yields a lower average profit of about 56, and a strategy that always offers a *large_discount* results in a negative average profit of around -31 due to excessive revenue loss. For comparison, the highest achievable average profit in this environment is approximately 180 monetary units.

4.5.3 Evaluation Results. A comparison of the learning agents’ performance is shown in Figure 1. The fixed-action baselines (*appeal*, *small_discount*, and *large_discount*) are indicated as horizontal lines for reference.

These results show that both Q-learning agents were able to learn successful negotiation strategies approaching the highest achievable profit. DQN agents were considerably slower to learn and achieved lower performances. For the used DQN neural network architecture, the learning converged after about 2,000 customers and scored about 120 (results were similar for other network architectures). In contrast, the Q-learning agents’ learning converged after about 200 customers. For both Q and DQN agents, the use of ER made the learning rate higher and improved the final

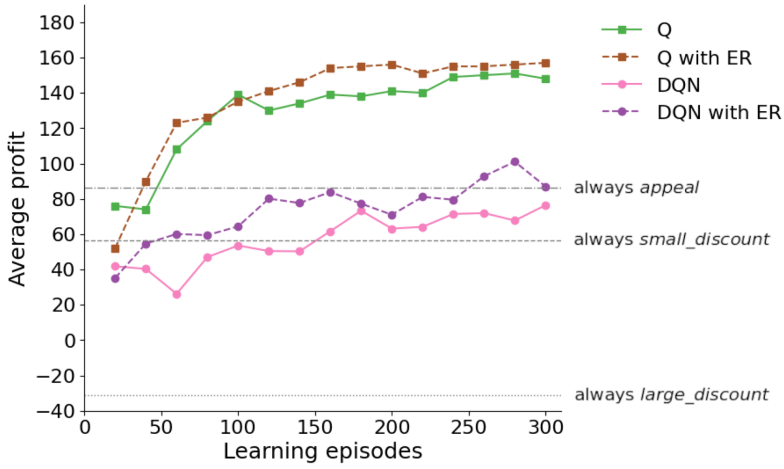


Fig. 1. Comparison of test scores of DQN and Q-learning agents.

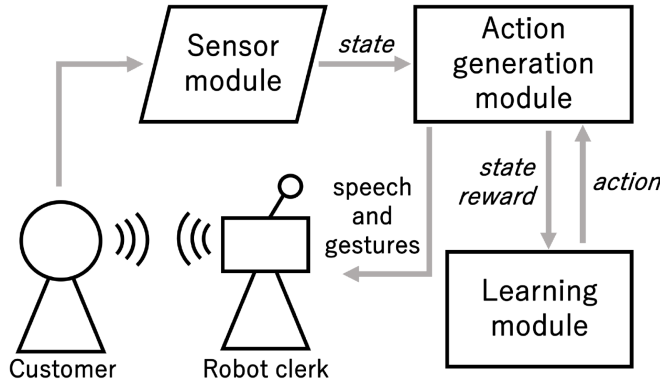


Fig. 2. Structure of negotiation system.

performance. The Q-learning agent with ER achieved the highest profit of about 160, which is why we chose it as the learning method for our negotiation system.

5 Complete Robot Negotiation System

Figure 2 shows an overview of the full sales negotiation robot system implemented in this study. It consists of a humanoid robot clerk, a sensor module that measures the customer's behavior, a learning module that performs the learning and action selection, and an action generation module that coordinates the execution of the robot's actions.

5.1 Robot Used

We used a 120 cm tall Robovie II robot, which has a humanoid upper body on a two-wheeled mobile platform and movable head and arms. The head is equipped with a camera and a speaker for the robot's speech. The robot can gesture using its head and arms. The robot did not change position during the service.

5.2 Sensor Module

The sensor module observed the customer's speech and movement and extracted features to be used for learning (Section 4.3).

The customer's speech was recorded through a wireless headset worn by the person in the customer role. The speech was converted into text in real-time using Google Cloud's Speech-to-Text service.¹ The extracted sentences were then sent to Google Dialogflow,² which was used to perform the utterance intention recognition. The intentions we wanted to recognize are the specific six *customer actions* that were part of the state (see Section 4.3), as well as other questions and utterances not related to negotiations, such as greetings or questions about the product features. We therefore prepared example sentences for each of the intentions and trained Dialogflow with them. With about 20–50 example sentences per intention, we obtained satisfactorily high recognition accuracy.

The speech detection output was also used to estimate the time *interval* between the robot's and customer's utterances. Intervals longer than 15 seconds were returned as "long"; otherwise, they were returned as "short."

The customer's movement was captured using a network of 10 Microsoft Azure Kinect sensors installed on the room's ceiling, which returned the customer's body skeleton consisting of 32 points. This was used to detect the customer's *gaze*, approximated by the direction that their head was facing. The head direction was calculated as the vector perpendicular to the line connecting the left and right ear points in the skeleton. When this direction was within 30° of the direction from the customer to the robot, it was treated as facing toward the robot. The head direction was calculated 10 times per second. If in the time between the robot's previous speech to the end of the customer's speech, the customer gazed at least half of the time toward the robot, the customer's gaze was classified as "robot"; otherwise, it was set to "other."

5.3 Learning Module

This module implemented the learning method described in Section 4.5. Based on the information on the current states and rewards, it performed the learning, decided the next action to be performed, and sent it to the action generation module.

5.4 Action Generation Module

The action generation module mapped the robot actions selected by the learning module into specific speech and gestures of the robot. To prevent an excessive repetition of the robot's behavior, multiple speech choices were prepared for each chosen action. For example, *appeal* was mapped into one of several possible utterances about a service to be offered, such as "Buy today and we'll pay for shipping." The number of candidate utterances for each of the actions was sufficiently large to ensure that there were essentially no repetitions during one interaction episode.

To make the conversation more natural, the module also added a few gestures to the conversation. The robot would point with its arm at the corresponding product when introducing it, and nod its head at the end of a conversation. In addition, while talking with the customer, the robot also made slight repetitive motions with its arms to resemble breathing and make it look more alive.

This module also controlled non-negotiation actions such as greetings and product introductions at the beginning of interactions, purchase processing and greetings at the end of a dialog, and responses to questions with predetermined answers. When such situations were detected based on the data from sensor modules, the action generation module directly chose the utterances and gestures, bypassing the learning module (states were also not forwarded to it and learning was not performed).

¹<https://cloud.google.com/speech-to-text>.

²<https://cloud.google.com/dialogflow>.

Designing and integrating these components into a functioning end-to-end system was nontrivial. The overall performance depended not only on the learning algorithm, but also on reliable estimation of abstract user states and the natural execution of high-level robot actions—challenges that are often overlooked in learning-focused studies.

6 Evaluation Experiment

6.1 Overview

We wanted to evaluate whether a robot that learns to negotiate using the proposed method in one specific store environment truly acquires sales negotiation strategies that work well for that environment. For this purpose, we prepared two store environments having different types of customers. The study was then conducted in two phases: the learning phase and the test phase. In the learning phase, in each of the stores, the robots interacted with customers while learning how to negotiate. In the test phase, the now-trained robots interacted with customers again, and for each store, we compared the average profit of the robot trained in that store with that of the robot trained in the other store.

6.2 Store Settings

The experimental environment consisted of a simple mock-up of an electric appliances store. Customers were in a room that simulated the interior of such a store and engaged in conversation with a robot clerk while looking at the products on display. The products were five types of home appliances, all in the same price range: a laptop computer, a robot vacuum cleaner, a refrigerator, a washing machine, and an air conditioner.

We prepared two different store environments:

- *The “bargain” store*, which advertised their openness to negotiating and discounting prices.
- *The “service” store*, which emphasized their various services (other than discounts), such as after-sales support or points that can be used on future purchases.

To achieve different sales negotiation conditions in each store, we manipulated the type of participants who played the customers visiting each store, as explained below. As an additional prompt to participants, we placed promotional posters in the store environment to advertise that store’s specific offers (i.e., discounts or services).

The choice of store types was informed by interviews with experienced clerks (Section 3), who described clear differences in customer expectations across stores. Some environments encouraged negotiation and price sensitivity, while others emphasized service quality and long-term value. Our store settings reflect these real-world patterns and were designed to elicit natural participant behavior without scripted role-play.

6.3 Procedure

The complete flow of the evaluation experiment is shown in Figure 3. The procedure was approved by our institution’s ethics board.

6.3.1 Participants Choosing the Store. In both the learning and test phase experiments, before the experiment, we first introduced the two stores to the participants using mock printouts of online reviews about the stores. The online reviews contained customer comments in which they stressed their satisfaction with the store’s specific characteristics (such as “I was able to get a great discount!” or “Make sure you ask for their extra services!”), plus a few negative reviews (e.g., “They would not offer me a bargain no matter how much I asked.”). We then asked the participants to

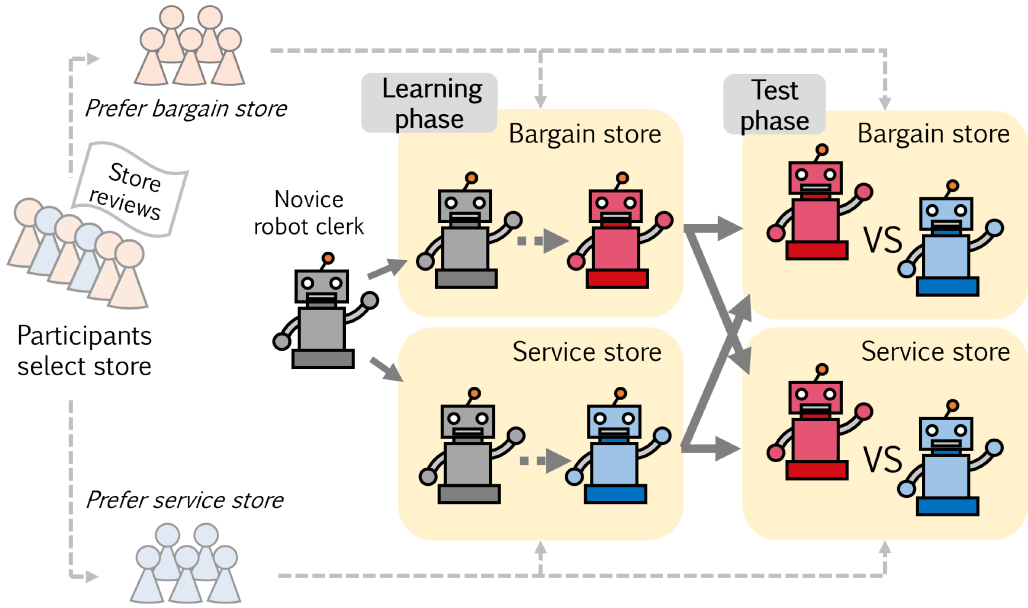


Fig. 3. Flow of the evaluation experiment.

choose which of the stores they would prefer to visit. This way the participants naturally split into two groups with different propensities for bargaining.

6.3.2 Learning Phase. After choosing the preferred store, we explained to the participants that they should imagine visiting a store to purchase one of the five products on display. We told them that they needed a new product because the one they were using had broken, but we did not tell them whether or not to purchase that product; we left that decision up to them. The idea was to ensure that their purchasing decision would depend on the interaction with the robot and not on external framing. The customer interacted with the robot until they decided to either purchase the product or quit the negotiation.

First, we let the participants try the negotiation interaction with the robot clerk (which acted based on predefined rules) several times to become accustomed to it. After that, they had negotiation interactions with either the learning robot in the bargain store or the one in the service store, depending on their store choice. The negotiation interactions were repeated for 2 hours (about 25 times on average per participant), rotating the order of products that they were buying.

During these interactions, the robot was running the RL algorithm and gradually updating its negotiation strategy. At both stores, the robot learning was stopped after a total of 200 episodes of negotiation interactions, which was achieved after the first 17 experiment participants, nine of whom chose the bargain store and eight the service store. The state of the learning robot was then saved. Hereafter, we refer to the robot that was trained in the bargain store environment as “bargain robot” and the one in the service store environment as “service robot.”

6.3.3 Test Phase. Similar to the learning phase, in the test phase all participants first went through the choice of stores, after which they were explained the task and given a chance to try interactions with the robot.

Each participant then had 10 negotiating interactions. They interacted with the bargain robot in five of these and with the service robot in the other five. The order in which each robot served was



Fig. 4. Example scene of an interaction between the negotiation robot and a participant playing the customer during the experiment.

counterbalanced between participants. The participants were not informed about the change in robot type. For both the bargain and service robot each participant had to purchase all five products once, with the order of the products counterbalanced to prevent any ordering effect. An example of a negotiation scene between a participant and the robot is shown in Figure 4.

The experiment was run with a total of 40 participants, 20 of whom selected the bargain store and 20 the service store. This resulted in 200 interaction episodes in each store, out of which 100 were with the bargain robot and 100 with the service robot.

6.4 Measurements

For each store separately, we evaluated the profit of the sales during the test phase of the bargain and service robot. Profit was defined as the total profit from sales of all five appliances, and for each participant it was evaluated separately for the bargain and service robot conditions. These profits were then used to evaluate whether there was a difference in the performance of the robots at each store.

In addition, we compared the Q-tables of the bargain and service robots to examine how much difference there was in their action choices, as well as how much the use of nonverbal information affected the action choice.

6.5 Participants

We hired the participants from the general public through online advertising. The only condition for hiring was that they had to have made at least one large purchase at an electric appliance store. The ages of the 17 participants during the learning phase ranged from 23 to 66 years old ($M = 43.4$, $SD = 10.6$), with five of them being male and 12 female. The 40 participants during the testing phase were 14 male and 26 female, with ages from 23 to 67 years old ($M = 43.2$, $SD = 14.2$). Participants in the learning and testing phases were distinct with no overlap between them. All participants were paid for their participation.

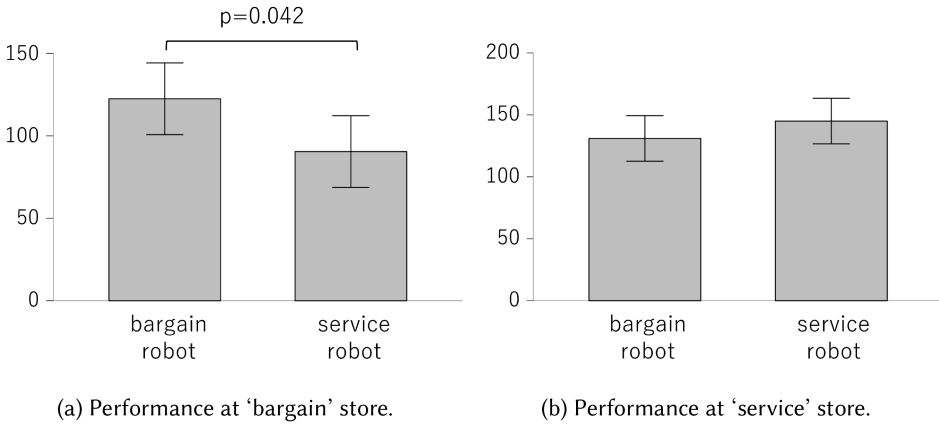


Fig. 5. Comparison of average profits per robot clerk at each store (error bars indicate standard errors).

6.6 Predictions

We hypothesized that each robot would acquire a strategy that is optimized for the store where it learned. We thus made the following predictions.

- Prediction 1: When serving customers who come to the bargain store, the average profit of the bargain robot will be higher than that of the service robot.
- Prediction 2: When serving customers who come to the service store, the average profit of the service robot will be higher than that of the bargain robot.

7 Results

7.1 Performance during Testing

A comparison of the average profits that each robot clerk achieved during the test phase at each store is shown in Figure 5. Shapiro-Wilk tests did not show deviations from normality (for both stores $p > .3$), so Student's t -tests were used for the within-subject comparisons.

In the bargain store, the bargain clerk made a significantly higher average profit ($M = 122.5$, $SD = 51.3$) than the service clerk ($M = 90.5$, $SD = 45.4$), $t(19) = 2.177$, $p = .042$, $d = 0.487$, thus supporting our Prediction 1. In the service store, there was no significant difference between the average profit of the bargain robot ($M = 131.0$, $SD = 54.9$) and the service robot ($M = 145.0$, $SD = 49.3$), $t(19) = -1.127$, $p = .274$, $d = -0.252$, so Prediction 2 was not supported.

7.2 Acquired Strategies

To confirm that the bargain and service robots learned distinct strategies, we compared their Q-tables after learning. We checked for how many states the two robots learned to take different actions, and found that such states accounted for 19.8% of the whole state space. Since some states were visited more often than others during the learning, to estimate how much difference there was between the actions the robots would actually take, we weighted the stores according to the frequency of visits. With such weighting, we obtained a difference in 36.1% of actions.

We also analyzed what the robots' specific action choices mean in terms of negotiation strategies. Altogether, the learned strategies appeared to align well with what we intuitively expected an experienced clerk in each of the stores would do. One specific example of a difference in action choice was in the situation where a discount had not yet been offered but the customer asked for one. In such cases, the trained bargain robot most often offered a discount (an exception is noted in

Table 1. Example Interaction from the Bargain Store Condition

Customer said [<i>detected action</i>]	[<i>chosen robot action</i>] Robot said
I'm looking for a robot vacuum cleaner.	Welcome. What kind of product are you looking for? This robot vacuum cleaner is recommended. The price is 100,000 yen.
The price is a bit high. [<i>complain</i>]	[<i>small_discount</i>] If the price is a concern, we can offer a discount. How about 95,000 yen, with a 5,000 yen reduction?
Still too expensive. [<i>complain</i>]	[<i>refuse</i>] I'm sorry, it seems we can't meet your expectations.
Do you offer any other services? [<i>demand</i>]	[<i>appeal</i>] We also offer a points-back service. You'll get points worth 10% of the purchase, so it's a good deal.
Is there a shipping fee? [<i>demand</i>]	[<i>appeal</i>] If you purchase today, we'll cover the shipping cost.
Alright, I'll buy it. [<i>buy</i>]	Thank you for your purchase. Please proceed to the service counter to complete your order.

The table shows the spoken utterances from both customer and robot. The robot responds to price-related complaints with a discount and later uses service-related appeals to address ongoing concerns.

the next section), whereas the choice of the service robot was to offer a service. Conversely, if a customer requested additional services, the service robot would offer them, but the bargain robot would offer a discount instead.

Comparing each robot's overall choice of actions, the bargain robot was more likely to go for the *small discount* behavior, choosing to offer it in 46.2% of the visited states, compared to the service robot, which chose that action only 11.8% of the time. Conversely, the service robot selected to offer a non-discount service (the *appeal* action) 76.9% of the time. These differences in the learned strategies reflect the framing and reward structure of each store, which were informed by interview findings and intended to model common negotiation styles observed in real-world retail settings (Section 3).

To illustrate these patterns concretely, Tables 1 and 2 show example interactions from each store condition. These highlight how the bargain robot tended to offer discounts early in the exchange, while the service robot consistently redirected requests toward added value or product features.

7.3 Use of Nonverbal Information

To measure whether the nonverbal features Gaze and Interval affected the choice of actions, we examined how the robots' action choices changed if we altered one of the nonverbal features and kept the other three features equal (while also weighting the states to account for the frequency of visits). Different gaze directions led to different behavior choices in 37.9% of the cases for the bargain robot and in 32.3% of the cases for the service robot. The difference in the interval until the customer responded resulted in 31.9% and 39.3% different action choices for the bargain and service robot, respectively. This suggests that the nonverbal features were indeed an important factor for the resulting robot behaviors.

Table 2. Example Interaction from the Service Store Condition

Customer said [<i>detected action</i>]	[<i>chosen robot action</i>] Robot said
I'd like a washing machine.	Welcome. What kind of product are you looking for? This washing machine is recommended. The price is 100,000 yen.
What kind of functions does it have? [<i>question</i>]	[<i>answer</i>] It's equipped with various wash programs that adapt to fabric type and soil level.
Does it have a dryer? [<i>question</i>]	[<i>answer</i>] This model can wash up to 10 kg of laundry in one cycle.
I want one with a drying function. [<i>question</i>]	[<i>answer</i>] This model supports both washing and drying. It's ideal for rainy days or small spaces.
Please give me a 5,000 yen discount. [<i>haggle</i>]	[<i>appeal</i>] Sorry, but we offer a points-back service: you'll earn 10% of the price in points.
Can you deliver it? [<i>demand</i>]	[<i>appeal</i>] If you purchase today, we'll cover the shipping cost.
How soon can it be delivered? [<i>question</i>]	[<i>answer</i>] It can be delivered within 3 days, or you can take it home today.
Then I'll buy this washing machine. [<i>buy</i>]	Thank you for your purchase. Please proceed to the service counter to complete your order.

The table shows the spoken utterances from both customer and robot. The robot emphasizes feature explanations and service-based appeals, while consistently avoiding discounting.

Consider this illustrative example: when a customer, having received no prior discount, initiates a bargaining request, the bargain robot would in most cases offer a discount. Nonetheless, if there was also a long pause in the conversation and the customer was not looking at the clerk, the robot learned to instead opt to offer a service. Since such customer behavior could be a sign that they are not interested in the purchase, it is likely that such an action choice was indeed more appropriate. Note how these types of rules are quite nuanced and might be challenging to articulate and manually program into the robot.

8 Discussion

8.1 Main Findings

The results show that using RL enables robots to successfully acquire a nuanced skill like sales negotiation and achieve strategies that match the specific store in which they are deployed, just by negotiating with customers. Moreover, the analysis of the trained agent's strategies also suggests that taking the human's subtle nonverbal behaviors into account can be important for achieving proficient interaction skills.

In addition to the learning process itself, this work demonstrates the feasibility of designing a complete interaction system that combines real-time perception, abstract state estimation, and socially appropriate action generation, all of which are critical for robust deployment.

While the slowness of learning systems is often an issue, we were able to achieve convergence to an effective navigation strategy in a reasonably small number of steps, making the robot learning of negotiation while on the job realistic. Additionally, during our pilot tests we confirmed that by starting from a robot that was partially pretrained in negotiation, the learning could be sped up

even further (this seemed to hold even if pretraining was done in a different store). Pretraining would reduce the period of suboptimal behavior and help mitigate potential financial costs in real deployments. Although actual commercial deployments would likely rely on pretraining to ensure acceptable performance from the outset, training from scratch allowed us to more clearly analyze the learning dynamics within our experimental setting. While our focus is on negotiation in a retail context, the findings may also apply to other customer-facing service domains—such as hospitality, reception, or delivery—where robots must interpret social cues, respond appropriately, and adapt their strategies over time.

Beyond learning alone, future systems may benefit from incorporating top-down guidance from store owners, such as preferred strategies or constraints, through mechanisms like reward shaping or policy regularization. This would allow for greater alignment with store values and policies.

Although the learning approach itself is not specific to robots, physical embodiment played a meaningful role in the interaction. Participants engaged with the robot face-to-face in a shared space, and their responses—such as gaze shifts, hesitation, or physical orientation—were shaped by this co-present setting. These subtle cues were not only detectable by the system, but also shaped the negotiation dynamic in ways that may not occur in screen-based or virtual settings. As such, embodiment contributed to the realism and ecological validity of both the interaction and the learning process.

Finally, while recent advances in large language models offer powerful tools for generating fluent and contextually appropriate language, these systems are not yet designed to optimize task-specific outcomes through real-time feedback. Our learning-based approach complements this direction by focusing on behavioral adaptation and performance-driven strategy acquisition in embodied, goal-oriented interactions.

8.2 Why the Bargain Robot Performed So Well?

Even though the average profit of the bargain robot was significantly higher than that of the service robot in the bargain store, there was no significant difference between the performance of the robots in the service store. We speculate that the reason may be the imbalance between the “giving discount” action of the bargain robot and the “offering service” action acquired by the service robot. Intuitively, even if service store customers prefer getting services, they are not likely to react negatively when offered a discount. Conversely, if a customer is hoping for a price discount but is only offered services, they might not be satisfied.

This difference may be explained by the well-documented phenomenon of delay discounting [30, 46], which describes the tendency of people to prefer rewards delivered soon to the ones delivered at a later stage. A discount offer provides an immediate reward, which is why it elicits a positive response from customers. Therefore, the bargain robot’s strategy to offer discounts is typically well received. In contrast, the value of services is typically not immediately available, and as such delayed rewards services may seem less attractive to customers.

8.3 Limitations

The experiments were conducted as purchases of fictitious products in a mock store environment, and this may have affected the behavior of the participants. For example, it is possible that they made purchasing decisions more easily than they would have in an actual shopping situation. We did consider using real monetary stakes to increase realism, but decided against it due to ethical constraints, as well as concerns about fairness and coercion.

Additionally, since participants repeated the negotiation interaction many times, their level of engagement may have decreased over time, potentially affecting the realism of their responses.

Although we varied the product types and randomized the order of conditions to reduce monotony, this remains a limitation of our experimental design.

The results may also have been influenced by the particular region or culture in which the experiment was conducted. As experienced clerks also noted in our interviews, such factors can strongly influence the negotiation behavior of customers, even within the same country.

Our study only focused on sales negotiations, which take place after a customer has already decided on the product they are going to buy. However, this is just a part of the sales process. As indicated by the interviewed clerks, many customer characteristics can also be learned during the product selection stage, which can be useful for negotiation. Thus there is room for learning strategies for improving sales outside of negotiations, such as during greetings and recommendations.

Finally, while we analyzed how nonverbal features influenced the robot's behavior, we did not conduct an ablation study to isolate their effect on task performance. Such an analysis would be valuable to assess whether including nonverbal cues measurably improves negotiation outcomes and remains an important direction for future work.

8.4 Ethical Considerations

The use of persuasive strategies in robot-mediated sales raises important ethical questions. While our system does not use deception or hidden incentives, optimizing negotiation success via RL could lead to subtle behaviors that influence users in ways they may not fully recognize. In future deployments, it will be important to ensure transparency about the robot's role, provide users with clear opt-out mechanisms, and carefully consider how persuasion is framed. Special attention should also be given to interactions with vulnerable users, where social pressure or asymmetries in information could impact decision-making.

Looking ahead, an important extension of this work will be in-person evaluations in actual store environments. Such deployments would allow us to assess not only the effectiveness of learned strategies, but also their impact on trust, customer experience, and ethical acceptability in real-world settings. They would also introduce new challenges, such as adapting while minimizing disruption, aligning robot behavior with store policies, and ensuring fairness across diverse users. These considerations will be central to responsibly scaling socially adaptive service robots beyond laboratory contexts.

9 Conclusions

This work focused on the question of how robot clerks can acquire the skills to effectively conduct sales negotiations in stores. Due to the difficulty of elucidating the exact strategies that should be used when doing such a nuanced task as negotiation, we proposed the learning of the task "on the job," using RL based on direct experiences of sales negotiations with customers. This allows the robot to not only gradually improve its performance, but also to be able to adapt to the slight differences that typically exist between stores.

To gain from the fact that a social robot clerk is not simply a text agent, but also physically situated in space, we focused on the use of both verbal and nonverbal information for sales negotiation, basing the choice on what we found about how human clerks conduct negotiations. Moreover, we evaluated different learning methods for their learning speed and settled on using Q-learning with ER, which was able to reach good negotiation performance within a number of steps comparable to those taken by human clerks.

We then evaluated the proposed negotiation learning robot system in two simulated store settings with different clientele. By first letting the robot learn separately in each store, and then comparing the performance of both trained robot clerks in both of the stores, we proved that it is possible to obtain effective customer service strategies adapted to each store environment.

References

- [1] Neziha Akalin and Amy Loutfi. 2021. Reinforcement learning approaches in social robotics. *Sensors* 21, 4 (2021), 1292. DOI: <https://doi.org/10.3390/s21041292>
- [2] Brenna D. Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. 2009. A survey of robot learning from demonstration. *Robotics and Autonomous Systems* 57, 5 (2009), 469–483. DOI: <https://doi.org/10.1016/j.robot.2008.10.024>
- [3] Reyhan Aydoğan, Onur Keskin, and Umüt Çakan. 2021. Let's negotiate with Jennifer! Towards a speech-based human-robot negotiation. In *Advances in Automated Negotiations*. Takayuki Ito, Minjie Zhang, Reyhan Aydoğan (Eds.), Springer, Singapore, 3–16. DOI: https://doi.org/10.1007/978-981-15-5869-6_1
- [4] Reyhan Aydoğan, Onur Keskin, and Umüt Çakan. 2021. Would you imagine yourself negotiating with a robot, Jennifer? Why not? *IEEE Transactions on Human-Machine Systems* 52, 1 (2021), 41–51. DOI: <https://doi.org/10.1109/THMS.2021.3121664>
- [5] Dražen Brčić, Tetsushi Ikeda, and Takayuki Kanda. 2017. Do you need help? A robot providing information to people who behave atypically. *IEEE Transactions on Robotics* 33, 2 (2017), 500–506. DOI: <https://doi.org/10.1109/TRO.2016.2645206>
- [6] Kushal Chawla, Rene Clever, Jaysa Ramirez, Gale M. Lucas, and Jonathan Gratch. 2024. Towards emotion-aware agents for improved user satisfaction and partner perception in negotiation dialogues. *IEEE Transactions on Affective Computing* 15, 2 (2024), 433–444. DOI: <https://doi.org/10.1109/TAFEC.2023.3238007>
- [7] Sonia Chernova and Andrea L. Thomaz. 2014. Robot learning from human teachers. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 8, 3 (2014), 1–121. DOI: <https://doi.org/10.2200/S00568ED1V01Y201402AIM028>
- [8] Sonia Chernova, Manuela Veloso, and Reid Simmons. 2017. HRI challenges in building a robot waiter and more: Lessons learned from the CoBot project. In *2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM, 378–386. DOI: <https://doi.org/10.1145/2909824.3020220>
- [9] Vignesh Chidambaram, Yuhua Chiang, and Bilge Mutlu. 2012. Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues. In *7th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 293–300. DOI: <https://doi.org/10.1145/2157689.2157798>
- [10] Sawyer Collins and Selma Šabanović. 2021. “What does your robot do?” A tabletop role-playing game to support robot design. In *2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 987–994. DOI: <https://doi.org/10.1109/RO-MAN50785.2021.9515507>
- [11] Francesco Del Duchetto, Paul Baxter, and Marc Hanheide. 2019. Lindsey the tour guide robot-usage patterns in a museum long-term deployment. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1–8. DOI: <https://doi.org/10.1109/RO-MAN46459.2019.8956329>
- [12] David DeVault, Johnathan Mell, and Jonathan Gratch. 2015. Toward natural turn-taking in a virtual human negotiation agent. In *2015 AAAI Spring Symposium*.
- [13] 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 Transactions on Human-Robot Interaction* 8, 3 (2019), 1–24. DOI: <https://doi.org/10.1145/3326462>
- [14] Kallirroi Georgila and David R. Traum. 2011. Reinforcement learning of argumentation dialogue policies in negotiation. In *12th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, 2073–2076. DOI: <https://doi.org/10.21437/Interspeech.2011-544>
- [15] Rachel Gockley, Allison Bruce, Jodi Forlizzi, Marek Michalowski, Anne Mundell, Stephanie Rosenthal, Brennan Sellner, Reid Simmons, Kevin Snipes, Alan C. Schultz, et al. 2005. Designing robots for long-term social interaction. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 1338–1343. DOI: <https://doi.org/10.1109/IROS.2005.1545303>
- [16] Jonathan Gratch, David DeVault, Gale M. Lucas, and Stacy Marsella. 2015. Negotiation as a challenge problem for virtual humans. In *International Conference on Intelligent Virtual Agents*. Springer, 201–215. DOI: https://doi.org/10.1007/978-3-319-21996-7_21
- [17] Jaap Ham, Andreas Spahn, and Sabine Roeser. 2015. Combining robotic persuasive strategies: The persuasive power of a storytelling robot that uses gazing and gestures. In *24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 439–445. DOI: <https://doi.org/10.1109/ROMAN.2015.7333666>
- [18] He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. Decoupling strategy and generation in negotiation dialogues. In *2018 Conference on Empirical Methods in Natural Language Processing*, 2333–2343. DOI: <https://doi.org/10.18653/v1/D18-1256>
- [19] Takuya Hiraoka, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura. 2016. Learning cooperative persuasive dialogue policies using framing. *Speech Communication* 84 (2016), 83–96. DOI: <https://doi.org/10.1016/j.specom.2016.09.002>
- [20] Shiu-Li Huang and Fu-Ren Lin. 2008. Using temporal-difference learning for multi-agent bargaining. *Electronic Commerce Research and Applications* 7, 4 (2008), 432–442. DOI: <https://doi.org/10.1016/j.elerap.2007.04.001>

- [21] Pranav Jindal and Peter Newberry. 2018. To bargain or not to bargain: The role of fixed costs in price negotiations. *Journal of Marketing Research* 55, 6 (2018), 832–851. DOI: <https://doi.org/10.1177/0022243718818451>
- [22] W. Bradley Knox, Samuel Spaulding, and Cynthia Breazeal. 2014. Learning social interaction from the wizard: A proposal. In *3rd Workshop on Machine Learning for Interactive Systems (MLIS '14) at the 28th AAAI Conference on Artificial Intelligence*.
- [23] Daegyu Lee, Gyuree Kang, Boseong Kim, and D. Hyunchul Shim. 2021. Assistive delivery robot application for real-world postal services. *IEEE Access* 9 (2021), 141981–141998. DOI: <https://doi.org/10.1109/ACCESS.2021.3120618>
- [24] Dongheui Lee, Christian Ott, and Yoshihiko Nakamura. 2010. Mimetic communication model with compliant physical contact in human–humanoid interaction. *The International Journal of Robotics Research* 29, 13 (2010), 1684–1704. DOI: <https://doi.org/10.1177/0278364910364164>
- [25] Mike Lewis, Denis Yarats, Yann Dauphin, Devi Parikh, and Dhruv Batra. 2017. Deal or no deal? End-to-end learning of negotiation dialogues. In *2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2443–2453. DOI: <https://doi.org/10.18653/v1/D17-1259>
- [26] Long-Ji Lin. 1992. Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine Learning* 8, 3 (1992), 293–321. DOI: <https://doi.org/10.1007/BF00992699>
- [27] Phoebe Liu, Dylan F. Glas, Takayuki Kanda, and Hiroshi Ishiguro. 2016. Data-driven HRI: Learning social behaviors by example from human–human interaction. *IEEE Transactions on Robotics* 32, 4 (2016), 988–1008. DOI: <https://doi.org/10.1109/TRO.2016.2588880>
- [28] Wing-Yue Geoffrey Louie and Goldie Nejat. 2020. A social robot learning to facilitate an assistive group-based activity from non-expert caregivers. *International Journal of Social Robotics* 12, 5 (2020), 1159–1176. DOI: <https://doi.org/10.1007/s12369-020-00621-4>
- [29] Joanna Mania, Fieke Miedema, Rose Browne, Joost Broekens, and Catharine Oertel. 2020. Towards understanding the effect of voice on human-agent negotiation. In *20th ACM International Conference on Intelligent Virtual Agents*, 1–8. DOI: <https://doi.org/10.1145/3383652.3423896>
- [30] Adriana da Matta, Fábio Leyser Gonçalves, and Lisiane Bizarro. 2012. Delay discounting: Concepts and measures. *Psychology & Neuroscience* 5 (2012), 135–146. DOI: <https://doi.org/10.3922/j.psns.2012.2.03>
- [31] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. 2013. Playing Atari with deep reinforcement learning. arXiv:1312.5602. Retrieved from <https://arxiv.org/abs/1312.5602>
- [32] Amal Nanavati, Malcolm Doering, Dražen Bršćić, and Takayuki Kanda. 2020. Autonomously learning one-to-many social interaction logic from human–human interaction data. In *2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 419–427. DOI: <https://doi.org/10.1145/3319502.3374798>
- [33] Ryuichi Nisimura, Takashi Uchida, Akinobu Lee, Hiroshi Saruwatari, Kiyohiro Shikano, and Yoshio Matsumoto. 2002. ASKA: Receptionist robot with speech dialogue system. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vol. 2, 1314–1319. DOI: <https://doi.org/10.1109/IRDS.2002.1043936>
- [34] Illah R. Nourbakhsh, Clayton Kunz, and Thomas Willeke. 2003. The Mobot museum robot installations: A five year experiment. In *2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vol. 4. IEEE, 3636–3641. DOI: <https://doi.org/10.1109/IROS.2003.1249720>
- [35] Amin Parvaneh, Ehsan Abbasnejad, Qi Wu, Javen Qinfeng Shi, and Anton van den Hengel. 2021. Show, price and negotiate: A negotiator with online value look-ahead. *IEEE Transactions on Multimedia* 24 (2021), 1426–1434. DOI: <https://doi.org/10.1109/TMM.2021.3065169>
- [36] Volha Petukhova, Christopher Stevens, Harmen De Weerd, Niels Taatgen, Fokke Cnossen, and Andrei Malchanau. 2016. Modelling multi-issue bargaining dialogues: Data collection, annotation design and corpus. In *10th International Conference on Language Resources and Evaluation (LREC)*, 3133–3140.
- [37] Ahmed Hussain Qureshi, Yutaka Nakamura, Yuichiro Yoshikawa, and Hiroshi Ishiguro. 2016. Robot gains social intelligence through multimodal deep reinforcement learning. In *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, 745–751. DOI: <https://doi.org/10.1109/HUMANOIDS.2016.7803357>
- [38] Ahmed Hussain Qureshi, Yutaka Nakamura, Yuichiro Yoshikawa, and Hiroshi Ishiguro. 2018. Intrinsically motivated reinforcement learning for human–robot interaction in the real-world. *Neural Networks* 107 (2018), 23–33. DOI: <https://doi.org/10.1016/j.neunet.2018.03.014>
- [39] Luca Raggioli, Fabio Aurelio D’Asaro, and Silvia Rossi. 2023. Deep reinforcement learning for robotic approaching behavior influenced by user activity and disengagement. *International Journal of Social Robotics*. 2023, 1–13. DOI: <https://doi.org/10.1007/s12369-023-01044-7>
- [40] Harish Ravichandar, Athanasios S. Polydoros, Sonia Chernova, and Aude Billard. 2020. Recent advances in robot learning from demonstration. *Annual Review of Control, Robotics, and Autonomous Systems* 3 (2020), 297–330. DOI: <https://doi.org/10.1146/annurev-control-100819-063206>

- [41] Pedro Sequeira, Patrícia Alves-Oliveira, Tiago Ribeiro, Eugenio Di Tullio, Sofia Petisca, Francisco S. Melo, Ginevra Castellano, and Ana Paiva. 2016. Discovering social interaction strategies for robots from restricted-perception Wizard-of-Oz studies. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 197–204. DOI: <https://doi.org/10.1109/HRI.2016.7451752>
- [42] Sandro Shelegia and Joshua Sherman. 2022. Bargaining at retail stores: Evidence from Vienna. *Management Science* 68, 1 (2022), 27–36. DOI: <https://doi.org/10.1287/mnsc.2021.4094>
- [43] Takeshi Shimmura, Ryosuke Ichikari, Takashi Okuma, Hiroyuki Ito, Kei Okada, and Tomomi Nonaka. 2020. Service robot introduction to a restaurant enhances both labor productivity and service quality. *Procedia CIRP* 88 (2020), 589–594. DOI: <https://doi.org/10.1016/j.procir.2020.05.103>
- [44] Richard S. Sutton and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction*. MIT Press.
- [45] Kazunori Terada, Mitsuki Okazoe, and Jonathan Gratch. 2021. Effect of politeness strategies in dialogue on negotiation outcomes. In *21st ACM International Conference on Intelligent Virtual Agents*, 195–202. DOI: <https://doi.org/10.1145/3472306.3478336>
- [46] Aaron D. Tesch and Alan G. Sanfey. 2008. Models and methods in delay discounting. *Annals of the New York Academy of Sciences* 1128, 1 (2008), 90–94. DOI: <https://doi.org/10.1196/annals.1399.010>
- [47] Iis Tussyadiah. 2020. A review of research into automation in tourism: Launching the annals of tourism research curated collection on artificial intelligence and robotics in tourism. *Annals of Tourism Research* 81 (2020), 102883. DOI: <https://doi.org/10.1016/j.annals.2020.102883>
- [48] Katie Winkle, Séverin Lemaignan, Praminda Caleb-Solly, Ute Leonards, Ailie Turton, and Paul Bremner. 2019. Effective persuasion strategies for socially assistive robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 277–285. DOI: <https://doi.org/10.1109/HRI.2019.8673313>
- [49] Yang Xue, Fan Wang, Hao Tian, Min Zhao, Jiangyong Li, Haiqing Pan, and Yueqiang Dong. 2021. Proactive interaction framework for intelligent social receptionist robots. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 3403–3409. DOI: <https://doi.org/10.1109/ICRA48506.2021.9562115>

Received 8 March 2024; revised 17 September 2025; accepted 30 September 2025