Field Experiments on the Effects of Multiple-Robot Expressions for Robot Influence in Recommendation Situations

Yota Hatano, Jun Baba, Junya Nakanishi, Yuichiro Yoshikawa, *Member, IEEE*, and Hiroshi Ishiguro, *Member, IEEE*

***Abstract*—Teleoperated social robots are becoming increasingly prevalent in society. To address the issue of their lack of influence, researchers have explored the use of multiple robots. Previous studies comprehensively investigated interactions with multiple robots, conducted web surveys, and revealed attributes that had positive, negative, and neutral effects on robot influence. In this study, we formulated hypotheses regarding multiple-robot expressions in the recommendation scenarios based on previous research findings and tested them through field experiments. We suggested that multiple-robot expressions with the attribute known as "Single-Sympathy," which represents that robots have a single role conveying a single intention, enhance robot influence, and there may be positive effect attributes specific to the field experiment. Additionally, our results indicate that other than multiple-robot expressions, such as drawing the attention of passersby, are also important, and non-operating robots possibly affect the sales of recommended products, indicating the importance of designing interactions throughout the entire recommendation situation. Our study provides new insights into the designing of multiple-robot interactions.**

***Index Terms*—Service robotics, social HRI.**

I. Introduction

TELEOPERATED social robots are increasingly being utilized, for example, in cafes, aiding employees in providing customer services remotely [1]. Providing customer services remotely has several advantages, including enhanced operator motivation and access to a broader talent pool for employers [2], [3].

However, teleoperated social robots do not have the same level of "influence" as human staff when making recommendation tasks [4]. A recommendation task refers to suggesting specific products or customer services to customers to encourage them to change their attitudes and behaviors toward the products and services. Previous studies have reported that customers tend not to heed a robot's recommendation conversation and focus more on the robot itself than on the contents of its recommendation conversation [5], [6]. Robots are not able to sufficiently promote products and make them appealing to customers, showing less influence in recommendation tasks. Customer service, including recommendation tasks, is an important task that can affect sales; hence, methods and technologies to enhance the influence of robots while making recommendations are needed.

Previous studies on the use of teleoperated social robots in recommendation tasks typically featured a single operator teleoperating a single robot. As explained earlier, a single robot does not have sufficient influence. However, several studies have compared the effectiveness of single and multiple robots and reported that multiple robots could enhance their influence on users. Shiomi et al. found that users' motor skills improved more when praised by two robots, compared to receiving praise from just one robot [8]. Marcos et al. reported that the number of hand sanitizer dispenser users increased more when multiple robots were used as opposed to when a single robot was used [10]. Therefore, we can consider using multiple robots can also increase the influence of recommendation tasks.

The following research question emerges: How do multiple robots collaborate to make recommendations? Shiomi et al. reported that multiple robots recite the same words to exert peer pressure on the interlocutors [9]. Marcos implemented sequential persuasion in which robots that were positioned sparsely persuaded passersby in order [10]. We wondered whether there were other effective or ineffective methods of collaboration. In previous research on persuasion (one-on-one conversation), the approach of the persuader was established to influence the success rate of persuasion significantly [7]. For example, the bandwagon effect [13], door-in-the-face technique [14], foot-in-the-door technique [15], low-ball procedure [16], lure effect [17], and that's-not-all technique [18] have been cited as influential interaction strategies. The bandwagon effect refers to the phenomenon wherein customers prefer what many other people choose and telling customers that other customers also use the service makes them more likely to use it. Similarly, in a recommendation task using multiple robots, how robots interact with customers is considered a factor in the robot's influence. We refer to these interaction patterns involving multiple robots

Manuscript received 30 September 2023; accepted 31 January 2024. Date of publication 26 February 2024; date of current version 5 March 2024. This letter was recommended for publication by Associate Editor S. Schneider and Editor A. Peer upon evaluation of the reviewers' comments. This work was supported by JST Moonshot R&D under Grant JPMJMS2011. *(Corresponding author: Yota Hatano.)*

Yota Hatano, Junya Nakanishi, Yuichiro Yoshikawa, and Hiroshi Ishiguro are with the Graduate School of Engineering Science, Osaka University, Toyonaka 560-0043, Japan (e-mail: hatano.youta@irl.sys.es.osaka-u.ac.jp).

Jun Baba is with AI Lab, CyberAgent Inc., Tokyo 150-0042, Japan.

Digital Object Identifier 10.1109/LRA.2024.3369946

TABLE I
EXPRESSIONS AND THEIR EXPRESSIVE ATTRIBUTE VALUE

No.	Expression name	Expressive attributes			
		ATa	RP	NRTA	ATi
1	Synchronized utterance	User	Dense	Single-Sympathy	Simultaneous
2	Group conformity	User	Dense	Single-Sympathy	Order
3	Synchronized announcement	User	Sparse	Single-Sympathy	Simultaneous
4	Sequential persuasion	User	Sparse	Single-Sympathy	Order
5	Collective agreement	User	Dense	Multiple-Sympathy	Order
6	Synchronized attention grab with gathered attention	User	Sparse	Multiple-Sympathy	Simultaneous
7	Being left out	User	Dense	Multiple-Refutation	Order
8	Synchronized dissent	User	Sparse	Multiple-Refutation	Simultaneous
9	Contrasting perspectives	User	Dense	Multiple-Sympathy & Refutation	Order
10	Synchronized bilateral presentation	User	Sparse	Multiple-Sympathy & Refutation	Simultaneous
11	Autonomous support	User	Dense	Multiple-Question	Order
12	Responsive companions	User	Dense	Multiple-Reaction	Order
13	Split-robot presentation	User	Dense	Multiple-Neutral	Order
14	Multi-location presentation	User	Sparse	Multiple-None	Simultaneous
15	Impacting unison	Robot	Dense	Single-Sympathy	Simultaneous
16	Group consultation	Robot	Dense	Single-Sympathy	Order
17	Leader reinforcement	Robot	Dense	Multiple-Sympathy	Order
18	Isolation empathy	Robot	Dense	Multiple-Refutation	Order
19	Negation overturn	Robot	Dense	Multiple-Sympathy & Refutation	Order
20	Intriguing unconfidence	Robot	Dense	Multiple-Question	Order
21	Reaction between robots	Robot	Dense	Multiple-Reaction	Order
22	Topic catalysts	Robot	Dense	Multiple-Neutral	Order

as "multiple-robot expressions." Although previous studies provided several examples of multiple-robot expressions [9], [10], [19], [20], differences in the influence of robots on various multiple-robot expressions have not been systematically investigated.

Our previous study identified 22 multiple-robot expressions and four expressive attributes that characterized them [21]. In addition, we conducted a web survey using videos of multiple-robot expressions with three robots. We revealed the attributes that enhanced the robot influence in the recommendation task. However, we only evaluated one-way interaction with robots; there is no research on real situations where two-way interaction with users occurs. Therefore, based on previous research, we conducted a field experiment to assess the effects of multiple robot expressions and their expressive attributes on sales promotion. In the experiments conducted in this study, we adopted the Wizard-of-Oz method. However, we believe that the insights of multiple-robot expressions apply to both teleoperated robots and autonomous robots. The critical contribution of this study is investigating multiple-robot expressions and finding attributes that affect robot influence in real situations involving interactions with users.

II. Multiple-Robot Expressions and Expressive Attributes

Multiple-robot expressions indicate the interaction patterns involving multiple robots. Our previous study [21] organized 22 multiple-robot expressions and four expressive attributes that characterized them (Table I). The details of the expressive attributes are as follows:

Action Target (ATa): ATa represents whether the robot's speech or motion is directed towards the user or another robot nearby. It determines whether robots collaborate to communicate with the user or interact with each other while allowing the user to overhear.

Number of Roles and Type of Attitude (NRTA): NRTA represents the number of roles robots assume and the attitudes of surrounding robots towards the teleoperated social robot's expressed intentions. If there's only one role, the attitude is typically sympathy. If multiple roles are present, six attitude types are possible: neutral, sympathy, refutation, sympathy and refutation, question, and reaction.

Robot Placement (RP): RP represents the arrangement of multiple robots relative to each other. It has two values. One value is "dense placement," where robots gather closely in front of a user group, enhancing their social presence by clustering them together. The other value is "sparse placement," where robots are discreetly positioned at fixed intervals from each other.

Action Timing (ATi): ATi indicates whether all robots speak or move simultaneously or in sequential order. Simultaneous speech amplifies the sound and extends its range, making it easier for users to notice. Repeating information sequentially can also create a lasting impression and help users remember the content.

In our previous study, we conducted a web survey to investigate the attributes that affect robot influence when compared with a single robot. We conducted a web survey to determine how much each multiple-robot expression changed the questionnaire score compared with a single robot, which was used as the robot influence score. We used the robot influence score as the objective variable. ATa had 2, NRTA had 7, RP had 2, and ATi had 2 attribute values, and one hot vector of these categorical variables was used as the explanatory variable. To avoid multicollinearity, we reduced one dimension from the vector of each attribute value. To check for interactions between variables, interaction terms were created from all pairs of categorical variables (21 pairs), converted to one hot vector, and added to explanatory variables. The final dimension of explanatory variables was 30. Using the explanatory variables, we conducted multiple regression analysis using the forward-backward stepwise selection

TABLE II
RESULTS OF MULTIPLE REGRESSION ANALYSIS (ROBOT INFLUENCE SCORE)

No.	Expressive attributes combination	Estimate	Std.Error	t value	Pr(>|t |)
1	”NRTA: Multiple-Sympathy & Refutation” · ”RP: Dense”	1.425718	0.234751	6.073321	1.725271e-09
2	”NRTA: Multiple-Sympathy” · ”RP: Dense”	0.951288	0.236275	4.026198	6.058646e-05
3	”ATi: Order”	0.534467	0.114487	4.668376	3.411361e-06
4	”NRTA: Multiple-Refutation” · ”RP: Dense”	0.351749	0.228892	1.536748	1.246436e-01
5	”NRTA: Single-Sympathy”	0.296463	0.109749	2.701276	7.013954e-03
6	”NRTA: Multiple-Question”	-0.391393	0.144371	-2.711028	6.812317e-03
7	Intercept	-0.400573	0.119138	-3.362256	7.997838e-04
8	”NRTA: Multiple-Refutation”	-0.871356	0.199138	-4.375644	1.326300e-05
9	”NRTA: Multiple-Sympathy”	-0.889427	0.207896	-4.278237	2.048254e-05
10	”NRTA: Multiple-Sympathy & Refutation”	-1.450388	0.205193	-7.068410	2.784431e-12

“Estimate” means the coefficient of each explanatory variable. “Std. Error” means standard error. The “t-value” means test quantity for the t-test. “Pr” is the p-value replaced by the t-value. R.Squared = .1927, adj.R.Squared = .1861.

method based on the Akaike Information Criterion. Table II presents the results of this analysis, with “Intercept” referring to the intercept value during multiple regression analysis.

According to Nos. 5, 6, 8, 9, and 10 in Table II, in NRTA, “Single-Sympathy” has a positive effect of 0.30, while “Multiple-Question,” “Multiple-Refutation,” “Multiple-Sympathy,” and “Multiple-Sympathy & Refutation” have negative effects of -0.39, -0.87, -0.89, and -1.45, respectively. However, according to Nos. 1, 2, and 4, “Multiple-Sympathy & Refutation,” “Multiple-Sympathy,” and “Multiple-Refutation” have a positive effect when combined with “RP: Dense.” The data indicate that when multiple-robot expressions are configured for “Dense” and “Multiple-Sympathy,” there is a positive effect of 0.06. In contrast, the conditions of “Multiple-Sympathy & Refutation” and “Multiple-Refutation” show negative effects of -0.02 and -0.52, respectively. No. 3 in Table II showed a positive effect of 0.53 when ATi is “Order.” In addition, RP and ATa did not affect the robot influence because they did not appear in the multiple regression analysis results.

In conclusion, referring only to the relatively large coefficients ($>$0.1), the results suggest that “Single-Sympathy” and “Order” have a positive influence, while “Multiple-Question” and “Multiple-Refutation” have a negative influence. However, in the web survey used for the above analysis, videos of one-way speech by robots were recorded for approximately 10 seconds. We believe that its applicability to real-world recommendation situations, where interactions with users can occur has not been sufficiently investigated. Therefore, we investigated whether expressive attributes affect the robot’s influence even in user interactions in a field experiment based on the results of the previous study.

III. FIELD EXPERIMENT

We conducted a field experiment using three teleoperated social robots to explore the effect of each multiple-robot expression on the robot influence. In this study, the experimental environment limited the attributes that could be implemented. Three robots had to be set up on a one-meter-wide shelf of the recommended products. Therefore, only “RP: Dense” could be implemented. We employed the Wizard-of-Oz method to conduct these experiments. Each robot was teleoperated by an operator from a remote location, and we assumed it was difficult to perfectly match the timing of speech and action between remote humans; hence, only the “ATi: Order” could be implemented. The results of a multiple regression analysis in our previous study showed that ATa has a non-significant effect on robot influence. Therefore, we investigated the effect of differences in NRTA on the robot influence.

A. Hypothesis

First, we used multiple regression analysis from previous studies to identify NRTA that had positive, neutral, or negative effects on the robot’s influence. We conducted clustering by producing coefficients from the results of multiple regression for each attribute value in the NRTA. The Elbow method was employed to determine the optimal number of clusters. This method identifies the most suitable cluster number by finding the point where the Sum of Squared Errors (SSE) within clusters begins to decrease sharply. By applying seven attribute values of NRTA, it was concluded that dividing the data into three clusters was the most appropriate. This conclusion was based on the observation that the change in SSE became less significant beyond the three clusters. With the number of clusters set to three, we conducted clustering using the K-Means method. Attribute values were divided into the following three clusters and the names were assigned to each cluster based on the effect of each attribute value on the robot influence, as follows: those with Single-Sympathy were called positive effect NRTA category, those with Multiple-Refutation and Multiple-Question were called negative effect NRTA category, and those with the rest of NRTA were called non-effect NRTA category.

For each category, the effects of other attributes on the influence of the robot were also considered, based on the results of the multiple regression analysis. Referring to Table II, the values that can be added from the coefficients of attributes that expressions have and intercept are called “predicted influence.” We then calculated the predicted influence of each category. Given the experimental constraints, the coefficients of the multiple regression analysis indicate that the predicted influence of positive effect NRTA category was 0.43, and positive effect NRTA category is expected to have relatively more influence than a single robot. Similarly, the predicted influences of negative effect NRTA category were -0.25 and -0.38, and negative effect NRTA

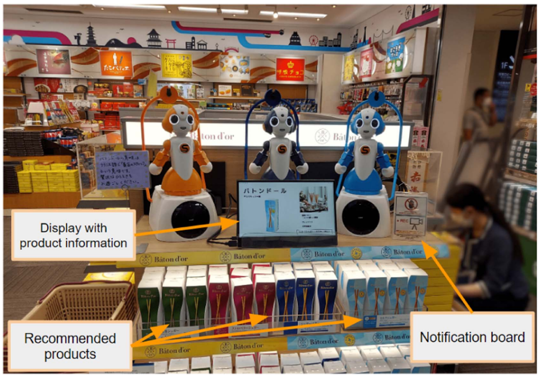

Fig. 1. Three robots recommend snacks at a souvenir shop.

category is expected to have a lower influence than a single robot. In this experiment, we sampled Multiple-Sympathy and Multiple-Neutral as non-effect NRTA category. The predicted influences of non-effect NRTA category were 0.20 and 0.13, respectively, and non-effect NRTA category is expected to have more influence than a single robot. Thus, the following three hypotheses were formulated.

(H1) Positive effect NRTA category has more influence than a single robot.
(H2) Non-effect NRTA category has more influence than a single robot.
(H3) Negative effect NRTA category has less influence than a single robot.

B. Experimental Environment

We conducted a field experiment at a souvenir shop located at an airport from 14:00 to 19:00 over 11 days in September 2022. In this experiment, operators teleoperated the social robots to recommend products to customers. The recommended products were snacks that cost approximately $4.4 and were available in four different flavors.

The experimental setup is shown in Fig. 1. Three robots were teleoperated by three operators using the system described in Baba's study [6]. The operators could interact with the users in the same manner as in video calls. The robots autonomously generated gestures in response to the speech of the operators. Additionally, operators can generate poses, such as waving or raising a hand, using command buttons.

The three robots were each teleoperated by an operator. The robot used was "Sota" manufactured by Vstone. A camera and speaker were attached to each robot. The display in front of the robot displayed the recommended product. The recommended products were placed on a shelf below the robot. During the robots' off-hours, we placed a board in front of the robots indicating they were "sleeping" and turned them off. We placed a notification board to inform all pedestrians that this was an experiment and that a camera attached to the robot recorded the experiment. The experiment was conducted on an opt-out basis for participants who chose to be excluded from the video data. This experiment was approved by the facility authorities and the Research Ethics Committee of Osaka University.

C. Recommendation Task

Five part-time collaborators, Japanese female voice actors experienced in teleoperating social robots, were hired to teleoperate the robots. When robots speak, it is important to match the impression of a voice with that of the appearance [26], [27]. In anthropomorphic robots, it is known that the mismatch between appearance and voice can induce a sense of eeriness [28]. In this experiment, we used Sota with a neutral and childlike appearance [29]. Operators were required to produce a high-pitched voice to match its appearance. These operators could provide high-pitched voices that matched the robot's appearance, and one operator used a voice changer to raise the pitch of her voice. Each day, three operators participated in the experiment. Operators assumed the role of a robot, calling out to passersby and recommending products. The operators were provided with a manual containing information on the recommended products. We also considered that the operators needed to practice for controlled experimental conditions because they would try multiple-robot expressions each with a different scenario during the experiment. Therefore, the operators practiced multiple-robot expressions for one and a half hours before the experimental period. Before the experiment began, we provided the operators detailed instructions for the day's expression, as outlined in the following subsection.

D. Interaction Design

Our instructions to the operators for each multiple-robot expression were as follows. We implemented these expressions following the method in the previous study [21].

Group consultation(GCs): Operators turn the robots to look at each other. When users have no response to the center robot's utterances, operators show users that robots consult with each other, such as "I hope he/she buy!," "Yeah, I hope so."

Group conformity(GCf): Operators always keep the robot facing the users and relay the same content without pausing. We instructed the operators to exert a certain level of pressure on the users even at the risk of them appearing brazen.

Leader reinforcement(LR): Operators turn the center robot to the other robots on both sides and speak. The left and right robots aimed to enhance the authority and influence of the central robot's statements through agreements and positive reactions.

Isolation empathy(IE): Operators turn the center robot to the other robots on both sides and speak. We instructed the operators that the robots on both sides of the center robot should turn to the center robot and deny the center robot's statement, such as "You are too brazen." The denied center robot briefly looks silently at the robots on either side for about one second. This action is designed to induce empathy in the users before the robot continues with its explanation.

Being left out(BLO): This expression is fundamentally similar to IE but differs in that both robots consistently face the users.

TABLE III
SCHEDULE OF THE FIELD EXPERIMENT

Day	Expression	NRTA	Category
1 (Tue)	Split-robot presentation (SP)	Multiple-Neutral	Non-effect
2 (Wed)	Group conformity (GCf)	Single-Sympathy	Positive
3 (Thu)	Group consultation (GCs)	Single-Sympathy	Positive
4 (Fri)	Leader reinforcement (LR)	Multiple-Sympathy	Non-effect
5 (Sat) 6 (Sun) 7 (Mon)	Free expression trial	-	-
8 (Tue)	Isolation empathy (IE)	Multiple-Refutation	Negative
9 (Wed)	Being left out (BLO)	Multiple-Refutation	Negative
10 (Thu)	Autonomous support (AS)	Multiple-Question	Negative
11 (Fri)	Single Robot(SR)	-	Baseline

The dialogues of the operators are also slightly different, for example, “This robot is so brazen, isn’t it?”

Autonomous support(AS): Operators ensure that the robots are always facing the users. After the center robot speaks to the users, when there is a pause in the conversation, the robots on both sides follow up to users, such as “Did you understand that?”

Split-robot presentation(SP): This expression is a simple form of non-collaboration between robots. Operators always keep the robots looking at the users, and speaking each sentence separately. Each operator is assigned a specific role, and operators are instructed not to assume each other’s roles.

To maintain consistency in the interaction flow, the operators followed a unified sequence: inviting customers, chatting, encouraging them to pick up products, describing the product, and encouraging them to purchase. Initially, the robots greeted the passersby with “Welcome” and asked them questions unrelated to the recommended product, such as “Are you going on a trip?” When a response was received from the user, the conversation continued for a while before moving on to the next sequence. The robot encouraged the customer by saying, “Pick up the product once,” and then provided details about the product. Finally, the robot said, “Please buy it,” to encourage the customer to make a purchase. In the web survey in our previous study, the participants watched videos showing a scene where they were recommended the products. Therefore, to replicate the same scenario, multiple-robot expressions began when the robots encouraged customers to pick up products.

We also implemented a recommendation condition using only a single robot as a baseline. Additionally, for reference, we also implemented a condition of “Free expression trial,” in which operators recommend products without constraints of expressions. To exclude the influence of the day of the week, we checked the sales data of the recommended products for two weeks without robots and found that sales tended to be higher from Saturday to Monday. Therefore, we conducted the conditions of multiple-robot expressions from Tuesday to Friday, and “Free expression trial” from Saturday to Monday. A single robot is used as a baseline on the final day of the experiment. The dates and categories for each multiple-robot expression are listed in Table III.

E. Evaluation and Analysis

To compare the effectiveness of each multiple-robot expression, we annotated the video recorded by the cameras attached to the robots during the experiment and analyzed user behavior. Two part-time workers with previous experience in similar video annotation tasks were hired for the annotation. One person annotated all the videos, the other one annotated one hour of video each day, and Cohen’s Kappas was. 785. The annotated items and options were as follows:

Watch: Whether the customer watched the multiple-robot expression that was conducted for each day.

Look: Whether the customer looked at the product before picking it up.

Talk: Whether the user responded to the robot’s utterance and spoke to the robot after watching the multiple-robot expressions before picking up the product.

Purchase: Whether the customer took the product with them.

A two-sample z-test for proportions between each category and baseline was conducted to test hypotheses. The test was employed to evaluate the null hypothesis that the purchase rates in both groups were equal. To compare the purchase rates influenced only by multiple-robot expressions, the purchase rate was defined as the number of people who purchased the recommended products divided by the number of people who looked at the products.

IV. RESULTS

Table IV shows the number of annotated actions, including look, talk, and purchase, to observe the pure differences in expressions. Watch expression counts the number of times the direction was properly executed as instructed. This value did not differ significantly across categories; it is higher in the SR condition than in the other expressions because SR condition is the most loosely constrained by the instructions, and its value is almost the same as the number of stops.

According to Table IV, the purchase rates for positive effect NRTA category and baseline were 0.533 and 0.256, respectively. Based on the results of the two-sample z-test for proportions, the null hypothesis was rejected at a significance level of 0.05 ($z = 1.972, p = .049$, *Cohen’s* $h = 0.577$). This suggests a statistically significant difference between positive effect NRTA category and the baseline. According to Table IV, the purchase rates for non-effect NRTA category and baseline were 0.263 and 0.256 respectively. Based on the results of the two-sample z-test for proportions, the null hypothesis was not rejected, indicating that there was no statistically significant difference in the purchase rates between baseline and non-effect NRTA category ($z = 0.061, p = .951$, *Cohen’s* $h = 0.017$). According to Table IV, the purchase rates for negative effect NRTA category and baseline were 0.346 and 0.256 respectively. The two sample z-test for proportions showed no statistically significant difference between negative effect NRTA category and the baseline ($z = 1.026$, $p = .305$, *Cohen’s* $h = 0.196$). Therefore, the null hypothesis

TABLE IV
NUMBER OF CUSTOMER ACTIONS AT EACH MULTIPLE-ROBOT EXPRESSION

Expression category	Expression	Watch expression	Look	Talk	Pur-chase	Purchase rate(%)	Sum of look	Sum of talk	Sum of purchase	Purchase rate(%)
Positive effect NRTA	GCf	55	6	2	3	50.0	15	2	8	53.3
	GCs	17	9	0	5	55.6				
Non-effect NRTA	SP	40	6	2	1	16.7	19	2	5	26.3
	LR	39	13	0	4	30.8				
Negative effect NRTA	IE	39	22	10	10	45.5	81	42	28	34.6
	BLO	44	26	13	5	19.2				
	AS	41	33	19	13	39.4				
Baseline	SR	124	43	12	11	25.6	43	12	11	25.6

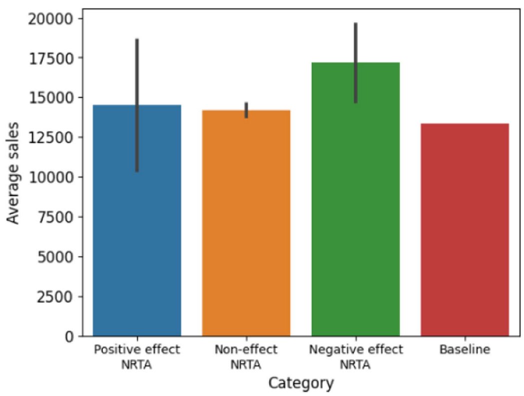

Fig. 2. Average sales of recommended products for each category. (Error bars indicate standard errors.).

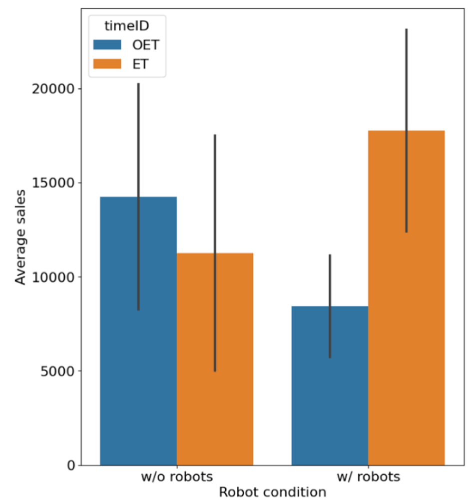

Fig. 3. Comparison of sales between the time the robot was not in operation and when it was in operation ET indicates the experimental time when robots teleoperated from 14:00-19:00, and OET indicates the outside experimental time (Error bars indicate standard errors.).

was not rejected, indicating that there was no statistically significant difference in the purchase rates between the baseline and negative effect NRTA category.

Fig. 2 shows the average sales of the recommended products when the robots in operate for each category. An analysis of variance (ANOVA) test indicates that there were no statistically significant differences in the sales of recommended products among the categories ($F(3, 4) = 0.843$, $p = .537$, $\eta^2 = 0.388$).

We made another interesting discovery: the sales of recommended products dropped outside the robot's operating hours during the experiment. "w/ robots" refers to the data for the eleven days of the experimental periods with the robots, while "w/o robots" refers to the data for eleven days of beginning on the same day of the week in July when robots were not installed (Fig. 3). "Experimental time(ET)" refers to the time when the operators teleoperated social robots from 14:00-19:00, while "Outside experimental time(OET)" refers to other times the store was open. For the "w/o robots," the average sale during OET was 14,243 yen, while during ET, the average sale was 11,241 yen. On the other hand, for the "w/ robots," the average sale of OET was 8,427 yen and that of ET was 17,756 yen. We conducted a comprehensive statistical analysis to investigate the differences in average sales between ET and OET under the conditions "w/o robots" and "w/ robots." For the "w/o robots" condition, the Levene's test for equal variances yielded that the variances between ET and OET are equal ($p = .949$). A subsequent two-sample t-test revealed that there was no significant difference in the average sales between ET and OET under the "w/o robots" condition ($t(20, 0.05) = -1.153$, $p = .262$, *Cohen's* $d = 2.198$). Similarly, for the "w/ robots" condition, the Levene's test indicated equal variances between ET and OET ($p = .146$). However, the subsequent t-test revealed that there was a statistically significant difference in average sales between ET and OET ($t(20, 0.05) = 5.156, p < .01$, *Cohen's* $d = -0.491$).

V. DISCUSSION

A. Reproducibility of Web Surveys

The results of the statistical test showed a significant difference in purchase rates between the positive NRTA effect category and the baseline. Therefore, (H1) is supported, indicating that multiple-robot expressions of positive effect NRTA category may have a high effect on robot influence and purchase rates.

The results of the statistical test revealed no significant difference in purchase rates between the non-effect NRTA category and the baseline. This suggests that (H2) was not supported. One possible reason for this was the small number of participants. A power analysis was conducted to evaluate the robustness of the statistical tests. Power analysis revealed a low statistical power of 0.056, suggesting that the result may not have been adequately powered to detect significant differences in purchase rates. The lack of a statistically significant difference could be attributed to the low statistical power. Only 19 people looked at the recommended products in non-effect NRTA category. This raises concerns regarding the adequacy of the number of participants for detecting actual differences. Future studies with a larger number of participants are required to provide more conclusive results.

Based on the results of the statistical test, no significant difference in purchase rates was observed between the negative NRTA effect category and the baseline. Therefore, it can be concluded that (H3) was not supported. One possible reason for not supporting (H3) is that the experimental situation differed from that of the web survey in the previous study. Specifically, users may be involved with multiple-robot expressions. Peng et al. reported that users are more accepting of services when a robot prompts the user for input than when the robot provides a one-way service [23]. In the IE and BLO conditions, the robot looked silently at the other robots for approximately one second after another robot spoke, effectively excluding that robot from the interaction. These expressions have room for user involvement, such as some reactions, because they invite the user's sympathy during silent moments. AS also has the same assumption because it is an expression that raises questions for the users such as "Did you understand that?" Table IV shows that the number of watch expression did not differ by category. However, the number of talk was larger in the negative effect NRTA category, suggesting that user involvement could have occurred. As mentioned above, in the actual field of recommendation situation, we noticed that there could be the attribute "User Involvement (UI)," which did not appear in the video evaluation. UI is an expressive attribute that motivates users to engage in dialogue and has two attribute values, involved or not involved. UI did not appear in the web survey because the user watched only a video of the robot speaking one-sidedly. Another possible reason for this is the lack of a complete experimental control. In this experiment, we used the Wizard-of-Oz method, which aimed to enable flexible adaptation in a field experiment where users do not engage in specific actions. However, it is possible that factors other than multiple-robot expressions, such as operator skills, may have affected the purchase rate. In this experiment, we conducted a free expression trial period on days 5, 6, and 7. In the trial, operators recommended products without the restriction of multiple-robot expressions to explore the possible expressions we have undiscovered. Before the trial, four users talked to the robots and, after the trial, 42 users talked to the robots. This indicates that the chat skills of the operators may have improved throughout the trial period, suggesting a lack of complete experimental control.

B. Sales for Each Category

Considering that there was a difference in the purchase rate depending on multiple-robot expressions, we can consider using not only highly influential expressions related to increasing sales. Furthermore, while multiple-robot expressions have a positive effect in recommendation scenarios, they can have a negative effect on stopping and drawing attention to the recommended product. On the other hand, several studies have been conducted on using robots to stop passersby. Okafuji et al. reported that robots performed better than humans in stopping passersby [24], and Amada et al. reported that pseudo-crowds using multiple robots attracted passersby [25]. Alternatively, there are methods to stop passersby using robots. Combined with the results of this field experiment, it may be feasible to have more influential interactions throughout the recommendation situation, including how to make customers stop in front of a recommended product or how to draw their attention to a recommended product, and it is necessary to further research on this point.

C. Sales Outside Robot Operating Hours

The results indicate that sales when OET considerably drop during the "w/ robots" condition. One of the possible reasons is that leaving the robots with a piece of paper indicating "sleeping" and turning off the robots gives customers the impression that the product is not sold here at this time. The effect of leaving the robot unattended on customers has not yet been investigated, and further research is needed on this point.

VI. Limitation

Several limitations need consideration in this experiment. The statistical power of the tests conducted was low, which can be attributed to the small number of participants. A long-term experiment with more participants might yield different results.

The recommended products were inexpensive. The results could vary depending on the products recommended, suggesting the need for further experiments with high-value products in the future.

The robots used in the experiment were small. These robots have been noted for their inherently low pressure [9], which could have limited the effects of the expressions. Since the design of the expressions proposed in this study is not limited to specific robots, future experiments could involve larger-sized robots.

Since the experiment adopted a Wizard of Oz method, the operator's talk skill might have influenced the results. Experimenting with an autonomous system could lead to more precise outcomes.

VII. Conclusion

In this study, we conducted a field experiment to investigate the effects of multiple-robot expressions and expressive attributes based on the previous findings. We found that using expressions with "Single-Sympathy," which represents that robots have a single role conveying a single intention, is effective for robot influence when Robot Placement is Dense and Action Timing is Order. This suggests that the results of the web survey in the previous study are partially valuable. On the other hand, we also found that when performing the recommendation task in the actual field, it is necessary to consider the field-specific attribute of user involvement. In addition, the results of this study suggest that there may be influences on the situations other than recommending products to customers, such as drawing the attention of passersby or the time of non-working robots. Further research throughout the recommendation situation is required to clarify this. The findings of this study are broadly applicable to the design of multiple robot interactions for robot influence.

References

[1] K. Takeuchi, Y. Yamazaki, and K. Yoshifuji, "Avatar work: Telework for disabled people unable to go outside by using avatar robots," in *Proc. ACM/IEEE Int. Conf. Hum.-Robot Interact.*, 2020.

[2] R. Ferreira et al., "Decision factors for remote work adoption: Advantages, disadvantages, driving forces and challenges," *J. Open Innov.: Technol., Market, Complexity*, vol. 7, no. 1, 2021, Art. no. 70.

[3] B. Frankiewicz and T. Chamorro-Premuzic, "The post-pandemic rules of talent management," *Harvard Bus. Rev.*, Oct. 13, 2020.

[4] J. Baba et al., "Teleoperated robot acting autonomous for better customer satisfaction," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, 2020, pp. 1–8.

[5] M. Tonkin et al., "Would you like to sample? Robot engagement in a shopping centre," in *Proc. 26th IEEE Int. Symp. Robot Hum. Interactive Commun.*, 2017, pp. 42–49.

[6] J. Baba, S. Song, J. Nakanishi, Y. Yoshikawa, and H. Ishiguro, "Local vs. Avatar robot: Performance and perceived workload of service encounters in public space," *Front. Robot. AI*, vol. 8, 2021, Art. no. 778753.

[7] S. DellaVigna and M. Gentzkow, "Persuasion: Empirical evidence," *Annu. Rev. Econ.*, vol. 2, no. 1, pp. 643–669, 2010.

[8] M. Shiomi et al., "Two is better than one: Social rewards from two agents enhance offline improvements in motor skills more than single agent," *PLoS One*, vol. 15, no. 11, 2020, Art. no. e0240622.

[9] M. Shiomi and N. Hagita, "Do synchronized multiple robots exert peer pressure?," in *Proc. Fourth Int. Conf. Hum. Agent Interaction*, 2016, pp. 27–33.

[10] M. T. Inky, K. Ogawa, Y. Yoshikawa, and H. Ishiguro, "Using multiple robots to increase suggestion persuasiveness in public space," *Appl. Sci.*, vol. 11, no. 13, 2021, Art. no. 6080.

[11] R. B. Kantharaju, D. D. Franco, A. Pease, and C. Pelachaud, "Is two better than one? Effects of multiple agents on user persuasion," in *Proc. 18th Int. Conf. Intell. Virtual Agents*, 2018, pp. 255–262.

[12] S. V. Suzuki and S. Yamada, "Persuasion through overheard communication by life-like agents," in *Proc. IEEE/WIC/ACM Int. Conf. Intell. Agent Technol.*, 2004.

[13] H. Leibenstein, "Bandwagon, snob, and Veblen effects in the theory of consumers' demand," *Quart. J. Econ.*, vol. 64, no. 2, pp. 183–207, 1950.

[14] R. B. Cialdini, J. E. Vincent, S. K. Lewis, J. Catalan, D. Wheeler, and B. L. Darby, "Reciprocal concessions procedure for inducing compliance: The door-in-the-face technique," *J. Pers. Social Psychol.*, vol. 31, no. 2, 1975, Art. no. 206.

[15] J. L. Freedman and S. C. Fraser, "Compliance without pressure: The foot-in-the-door technique," *J. Pers. Social Psychol.*, vol. 4, no. 2, 1966, Art. no. 195.

[16] R. B. Cialdini, J. T. Cacioppo, R. Bassett, and J. A. Miller, "Low-ball procedure for producing compliance: Commitment then cost," *J. Pers. Social Psychol.*, vol. 36, no. 5, 1978, Art. no. 463.

[17] R. V. Joule, F. Gouilloux, and F. Weber, "The lure: A new compliance procedure," *J. Social Psychol.*, vol. 129, no. 6, pp. 741–749, 1989.

[18] J. M. Burger, "Increasing compliance by improving the deal: The that's-not-all technique," *J. Pers. Social Psychol.*, vol. 51, no. 2, 1986, Art. no. 277.

[19] T. Arimoto, Y. Yoshikawa, and H. Ishiguro, "Nodding responses by collective proxy robots for enhancing social telepresence," in *Proc. Second Int. Conf. Hum.-Agent Interaction*, 2014, pp. 97–102.

[20] K. Seta, M. Yokoyama, S. Yoshida, T. Narumi, T. Tanikawa, and M. Hirose, "Divided presence: Improving group decision-making via pseudo-population increase," in *Proc. 6th Int. Conf. Hum.-Agent Interaction*, 2018, pp. 260–268.

[21] Y. Hatano, J. Baba, J. Nakanishi, Y. Yoshikawa, and H. Ishiguro, "Exploring the recommendation expressions of multiple robots towards single-operator-multiple-robots teleoperation," in *Proc. Int. Conf. Hum.-Comput. Interaction*, 2023, pp. 46–60.

[22] E. Walster and L. Festinger, "The effectiveness of "overheard" persuasive communications," *J. Abnorm. Social Psychol.*, vol. 65, no. 6, pp. 395–402, 1962.

[23] Z. Peng, Y. Kwon, J. Lu, Z. Wu, and X. Ma, "Design and evaluation of service robot's proactivity in decision-making support process," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–13.

[24] Y. Okafuji et al., "Behavioral assessment of a humanoid robot when attracting pedestrians in a mall," *Int. J. Social Robot.*, vol. 14, no. 7, pp. 1731–1747, 2022.

[25] J. Amada, Y. Okafuji, K. Matsumura, J. Baba, and J. Nakanishi, "Investigating the crowd-drawing effect, on passersby, of pseudo-crowds using multiple robots," *Adv. Robot.*, vol. 37, no. 6, pp. 423–432, 2023.

[26] K. Seaborn, N. P. Miyake, P. Pennefather, and M. Otake-Matsuura, "Voice in human–agent interaction: A survey," *ACM Comput. Surv.*, vol. 54, no. 4, pp. 1–43, 2021.

[27] C. McGinn and I. Torre, "Can you tell the robot by the voice? An exploratory study on the role of voice in the perception of robots," in *Proc. ACM/IEEE 14th Int. Conf. Hum.-Robot Interaction*, 2019, pp. 211–221.

[28] W. J. Mitchell, K. A. Szerszen Sr., A. S. Lu, P. W. Schermerhorn, M. Scheutz, and K. F. MacDorman, "A mismatch in the human realism of face and voice produces an uncanny valley," *i-Perception*, vol. 2, no. 1, pp. 10–12, 2011.

[29] G. Perugia, S. Guidi, M. Bicchi, and O. Parlangeli, "The shape of our bias: Perceived age and gender in the humanoid robots of the ABOT database," in *Proc. ACM/IEEE 17th Int. Conf. Hum.-Robot Interaction*, 2022, pp. 110–119.