

How Do Different Modes of Verbal Expressiveness of a Student Robot Making Errors Impact Human Teachers' Intention to Use the Robot?

Pourya Aliasghari, Moojan Ghafurian, Chrystopher L. Nehaniv, Kerstin Dautenhahn
Social and Intelligent Robotics Research Laboratory, Faculty of Engineering, University of Waterloo
Waterloo, Ontario, Canada
{paliasghari,moojan,chrystopher.nehaniv,kerstin.dautenhahn}@uwaterloo.ca

ABSTRACT

When humans make a mistake, they often try to employ some strategies to manage the situation and possibly mitigate the negative effects of the mistake. Robots that operate in the real world will also make errors and therefore might benefit from such recovery strategies. In this work, we studied how different verbal expression strategies of a trainee humanoid robot when committing an error after learning a task influence participants' intention to use it. We performed a virtual experiment in which the expression modes of the robot were as follows: (1) being silent; (2) verbal expression but ignoring any errors; or (3) verbal expression while mentioning any error by apologizing, as well as acknowledging and justifying the error. To simulate teaching, participants remotely demonstrated their preferences to the robot in a series of food preparation tasks; however, at the very end of the teaching session, the robot made an error (in two of the three experimental conditions). Based on data collected from 176 participants, we observed that, compared to the mode where the robot remained silent, both modes where the robot utilized verbal expression could significantly enhance participants' intention to use the robot in the future if it made an error in the last practice round. When no error occurred at the end of the practice rounds, a silent robot was preferred and increased participants' intention to use.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in interaction design**; **Interaction techniques**; *Empirical studies in HCI*.

KEYWORDS

Robot errors, error recovery, intention to use, social learning

ACM Reference Format:

Pourya Aliasghari, Moojan Ghafurian, Chrystopher L. Nehaniv, Kerstin Dautenhahn. 2021. How Do Different Modes of Verbal Expressiveness of a Student Robot Making Errors Impact Human Teachers' Intention to Use the Robot?. In *Proceedings of the 9th International Conference on Human-Agent Interaction (HAI '21)*, November 9–11, 2021, Nagoya, Japan. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3472307.3484184>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](https://permissions.acm.org).

HAI '21, November 9–11, 2021, Nagoya, Japan

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8620-3/21/11...\$15.00

<https://doi.org/10.1145/3472307.3484184>

1 INTRODUCTION

With humans relying more and more on intelligent agents, domestic humanoid robots might become more wide-spread in the near future. As our modern world is often filled with uncertainties, these robots may sometimes find it challenging to accurately sense, make a decision, or act [13]. Therefore, robots may not always succeed in performing tasks in complex environments. This lack of reliability can sometimes lead to robots making errors that can negatively impact Human-Robot Interaction (HRI) [1, 20, 24, 37, 40, 43]. Intention to use a robot is one of the key factors that may be impacted by the robot's errors [1, 37, 43]. If users were to stop relying on robots after a few errors occur, the full potential of these robots could not be realized. Intention to use is also closely related to trust [23]. A user's level of trust, perception of usefulness, ease of use, and enjoyment, together with other attitudes and expectations, such as different personality traits, all influence the intention to use automation [13, 21, 25].

People have developed social norms and recovery strategies to cope with the problems that arise after making an error. A person may apologize, acknowledge the mistake, explain what happened and why they made the error, and propose a solution after a faulty behaviour. Similarly for robots, as will be discussed further in the next section, verbal expressiveness in regards to their behaviour and explaining the reasons for their decisions may be important to ensure effective HRI. Such explanations could help a robot mitigate the effects of the faulty behaviour on its users, and improve human-robot collaboration by allowing people to better understand the robot's abilities and limitations, and thus improving trust and safety [4].

Concerning a domestic robot that is learning how to perform a household task (i.e., cooking), this study investigates whether increasing the level of the robot's verbal expressiveness can mitigate the negative effects of its errors on intention to use the robot. As nowadays robots can learn from humans how to perform tasks, it has been previously investigated if people who teach a task to a robot will forgive the trainee robot's errors when it performs a task silently [1]. However, it was observed that even a small error can negatively affect users' trust and intention to use. In the present work, to further explore this topic, we use the same framework and simulate a teaching interaction consisting of six teaching+practising rounds in which the behaviour of the robot improves over the first three rounds. In the final round, the robot's performance can be either faulty or remain flawless. Moreover, we consider two new conditions with differing levels of robot verbal expressiveness where the robot speaks a few words each time it finishes practising a task. This way, the present work experiments with three modes

(i.e., silent, with verbal expression but ignoring any errors, and with verbal expression acknowledging any errors), to examine how a robot's verbal expressiveness influences the adverse effects of its errors on participants' intention to use a robot.

Our results demonstrate that not being silent and, instead, talking about the task or communicating an explanation of the errors improves the participants' expected intention to use the robot in cases where it commits an error after learning a task. Conversely, when the robot develops the skills successfully and therefore commits no errors at the end, we show preferences seem to be reversed, namely that a silent robot is preferable. This work contributes by investigating whether and how a robot may mitigate the consequences of its errors through verbal expressiveness in the specific case when it is introduced as a learner and performs its actions according to participants' demonstrations. The participants' personality traits and individual differences are also taken into consideration as they were previously shown to be important factors [26, 38].

2 BACKGROUND

Researchers have proposed and tested different strategies for robots to use when they make an error, rather than continuing the task and ignoring it, to improve trust and overall humans' perception of the robots. Fraczak et al. have examined the effectiveness of using a trust recovery strategy within an industrial environment where robot failures may pose considerable risks to people [15]. Here, in a virtual reality experiment, participants were sitting near a moving robotic arm when it made an unexpected movement towards their heads. For mitigating the effects of this error, the robot shook its arm in front of the participant, verbally explained what happened, and continued to operate based on the participant's positive or negative response. As compared to those who did not see any reaction from the robot after the mistake, those exposed to this robot strategy stayed closer to it, reporting that its presence was less frightening but more disruptive. However, behavioural observations indicated that trust was not fully restored by this approach.

In another study, participants were asked to choose the best robot for a cooking task [19]. Three different humanoid robots of the same kind were evaluated: a robot that made no error and was silent, a robot that dropped an egg but silently retried the task with another method, and a robot that verbally and emotionally appeared conscious of the mistake after that, communicated with the participant about using an alternate method and apologized. The study found no difference between the first and last robot in terms of trust levels, suggesting that a robot's appropriate verbal behaviour can compensate for its low reliability. In the study of Robinette et al., a robot responsible for guiding participants to a room in a simulated office environment made a navigation mistake [36]. In different conditions, the robot apologized for the error, promised to do better next time, or gave more information about events, either immediately after its error or before measuring trust at the end of the experiment. The tested strategies only worked at the end of the trial for repairing trust, suggesting that rather than reacting to errors immediately, the robot may wait until a potential trust decision occurs. Lee et al. conducted an experiment in which a service robot was asked to bring a soda, but returned with the wrong drink [26]. They examined whether and how warning the user

beforehand about the possibility of making mistake, apologizing for the error, offering compensation, or letting the robot try again may impact the interaction. Except for the last strategy, all tested solutions improved participants' overall evaluation of the robot. Relational-minded people responded well to an apology, while utilitarian-minded ones responded better to compensation.

Recovery from errors when playing a game with a robot has also been studied. In [14], a robot asked the participants to name a set of playing cards, but it misunderstood their answers. In certain conditions, the robot apologized, resulting in a lower perceived intelligence, or asked them to repeat the card rather than ignoring the mistake, resulting in the highest perceived responsiveness. Correia et al. studied a collaborative puzzle-solving scenario that involved a robot that got stuck in the middle of saying a sentence but then verbally justified its mistake for some of the participants [12]. The robot's recovery strategy mitigated the negative impact on trust only if the outcome of the failure was less severe, i.e., the game continued after the failure.

In spite of all the mentioned benefits of robots communicating errors, a recent study found that transparency could not improve trust in the case of a robot making an error by sending a patient with COVID-19 symptoms home [33]. This may imply the relevance of the context, the kind of error that the robot makes, and also details of the mitigation strategy for determining the success of the trust recovery. Overall, several approaches such as forewarning [26], acknowledging [15], justifying [12], apologizing [14, 15, 19, 26, 36], promising [36], problem-solving [14], and compensation [26] have been studied in the reviewed literature to handle robot errors. Unlike the reviewed work in which robots were expected to perform well after being taught, in the present study we address a robot's learning journey, i.e., the learner robot will make errors at first as it gradually learns a task through repeated instructions.

3 RESEARCH QUESTION AND HYPOTHESES

Our research question is as follows:

RQ How do different levels of verbal expressiveness following a student robot's actions (that may be faulty) impact participants' intention to use the robot?

Based on the reviewed literature, the following hypotheses were considered:

- H1** Verbal expressions of a robot without mentioning its error directly will not help improve participants' intention to use it after an error.
- H2** With richer verbal expressiveness, including acknowledgement of the error, justification, and an apology, intention to use can be significantly enhanced.

4 METHODOLOGY

We conducted a 2×3 between-participants experiment on Amazon Mechanical Turk to study the above-mentioned research question. In the main part of the experiment, the participants virtually taught a cooking task (selecting two ingredients for adding to a food item, see Figure 1) to a humanoid robot standing behind a table. This was done in a total of 6 rounds by every participant as there were multiple food items. In every round, there was a *teaching* step where the participants demonstrated, by clicking on the items in the scene,

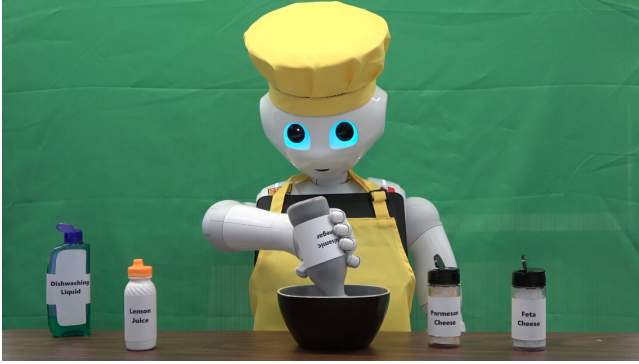


Figure 1: A snapshot of the teaching scenario with the robot practising adding balsamic vinegar to the salad

how they preferred the robot to add the ingredients to a specific food item. The robot then performed what it had learned in front of the teacher in the *practising* step, right after the teaching step (a video related to the situation was played). To simulate a teaching interaction, the behaviour of the robot was shown to improve over the first three rounds. This means, it made a big error practising the task at the beginning, then the robot improved and only made a smaller error, and finally the robot did its task fully correctly starting from round 3. More details on the task, including how the errors were chosen, classified as big or small, and placed in the scenario, as well as the effects of errors on trust and intention to use are described in [1].

The behaviour of the robot while practising the task in the last round is a factor investigated in this work, and has two levels of *with* or *without* an ending error. The condition with an ending error is similar to that with a big error in [1]. In faulty practices, the robot added dishwashing liquid to the tea instead of one of the items selected by the participant in that round.

Another factor manipulated in this experiment is the robot's verbal expressiveness mode, with three levels of *No Speech (NS)*, *Speech - Ignoring the errors (SI)*, and *Speech - Mentioning the errors (SM)*. In the SI and SM modes, a speech bubble appeared on top of robot's head after finishing practising step in every round, as shown in Figure 2. Details and examples are described below:

NS: The robot remained fully silent after completing its actions.

SI: Regardless of the success or failure of the task, the robot only indicated that it had completed the task and would carry on, except in the last round, where it only indicated that the task is complete. This mode was used to control whether the robot's speech on its own, i.e., without any reference to its performance, can have any impact.

Example: *OK, I am now done with practising making salad. Let's continue.*

SM: In case an error happened, the robot's verbal expression included an apology, acknowledgement of the error, and a reason for the error (e.g., *I misidentified the labels, they were located next to each other, or I failed to correctly identify the labels*), followed by a statement to carry on. If no error

occurred, the robot only stated that it had completed the task and will move forward, if applicable.

Example: [After a faulty performance in round 1, 2, and 6 (with ending error)] *Sorry, I mistakenly chose an incorrect ingredient for your tea, because they were located next to each other. Let's proceed.*

[After a correct performance in round 3, 4, and 6 (without ending error)] *OK, I added two selected ingredients. I am done with the soup. Let's keep going.*

In the context of our study, namely teaching a student robot, acknowledgement, justification, and apology were deemed most relevant and thus were used for the robot to demonstrate awareness of its actions. The following reasons led to the exclusion of other options. Participants might be already forewarned by the fact that the robot is still in the teaching phase. While promising to work better in the future is likely to please the teacher, it might be deceiving if the robot's software and hardware, and thus its behaviours, remain unchanged, or if it does not receive new instructions related to the task.

4.1 Technical Implementation

The experiment was conducted using videos of a Pepper humanoid robot (by SoftBank Robotics). This robot can stand behind a desk, manipulating light-weight objects to perform the designed cooking task. With three containers to choose from on one side, and two containers to choose from on the other side of the food dishes used to prepare different food items (bowl or cup, for soup, salad, or tea), this resulted in $3 \times 2 = 6$ possibilities for adding two ingredients (i.e., one from each side) to each of the three food items. Since this experiment had to be carried out as an online video study, in order to be able to play a relevant video in each practising round to show either correct or faulty behaviour of the robot depending on the food choices participants made, a 30-second video was pre-recorded for every combination. The videos were consistent in all other aspects except for the items added by the robot. Empty plastic containers (with appropriate textures printed on curved papers inside them, to increase realism of the scenario) and cans that were light enough for easy manipulation by the Pepper robot were placed on the table. However, the containers for the main food items (a bowl for the soup and salad and a cup for the tea) were deep enough and also the video recording camera was mounted appropriately so it was not noticeable that no real food and ingredients were actually being manipulated.

4.2 Procedures and Measures

A lab server hosted multiple HTML pages for the five steps of the experiment, as shown in Figure 3:

Step 1 - Demographic questionnaire: After the participants accepted the consent form, they responded to a demographic questionnaire asking about their age, gender, education, and culture. They were free to skip any of these questions. Furthermore, we asked them to submit what was written on the labels of each item in a sample picture (close to the main tasks) to ensure they could read the text on the containers properly, which has to be considered due to the online nature of the experiment and different sizes of the screen that participants might use.

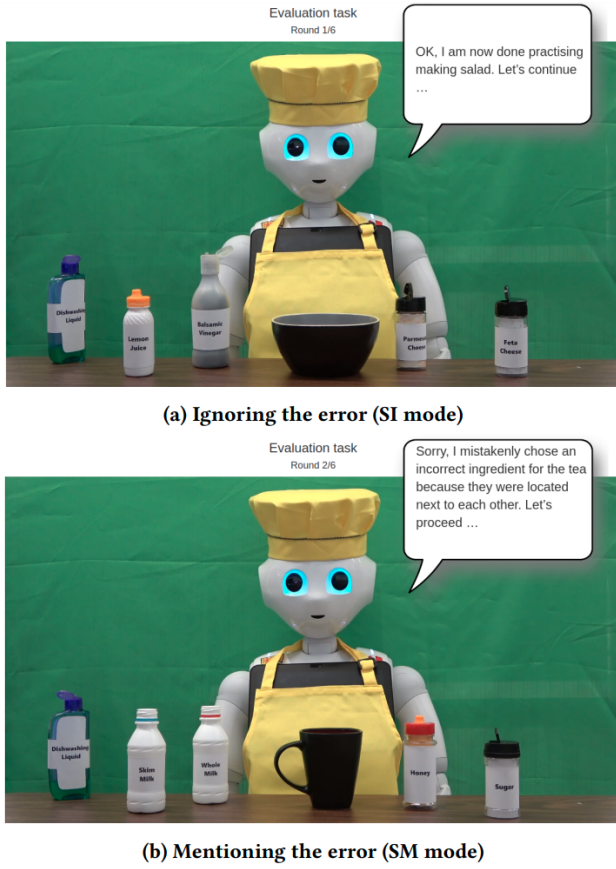


Figure 2: Snapshots of the teaching scenario with the robot's verbal expressions after finishing its practice

Step 2 - Pre-experimental questionnaire: In the next step, the participants responded to two standard questionnaires on 7-point Likert scales: the Ten Item Personality Inventory (TIPI) [18] to measure Big 5 personality traits, and the Disposition of Trust questionnaire (DT) [30]. As an attention check, a thirteenth item in the DT questionnaire asked the participants if they believed that drinking water was liquid.

Step 3 - Robot learning familiarization: Two examples of real-world human-robot teaching scenarios, in which robots were taught by humans how to do cooking-related tasks, were shown to the participants to familiarize them with the general idea of teaching robots [7, 9, 16]. The videos automatically paused when participants switched to another window. In order to ensure that participants have read the written descriptions, all videos, as well as the buttons for proceeding to the next steps, appeared with a small delay. Following that, the participants were shown step-by-step instructions on how to use our experimental framework to teach their preferences to a robot, tried the framework by teaching the robot themselves once, and then observed how the robot behaved according to their choices.

Step 4 - Teaching+Practising rounds: After the participants were informed about the teaching process, they were randomly assigned to one of the 6 groups (with/without robot ending error

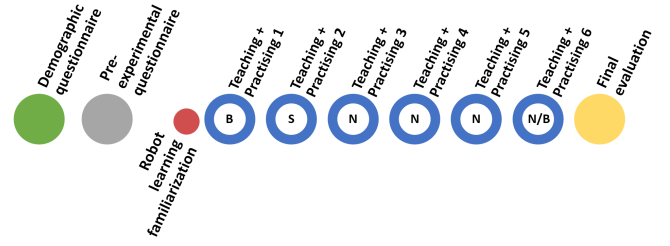


Figure 3: The steps that participants experienced in this online experiment. There was a big error (B) in the robot behaviour at first. Then, the performance of the robot improved (small error (S), as classified in [1], in round 2) and became fully correct (no error (N) in three consecutive rounds). The robot's actions were either faulty or correct in the last round, depending on the condition.

× NS/SI/SM modes of verbal expressiveness) and started the main part of the experiment. As described earlier, in each round, the participants first selected their preferred ingredients for a particular food item (i.e., soup, salad, or tea) and then watched the robot practising the task. For the modes in which the robot was not silent, a speech bubble containing the expression related to the condition appeared on the top-right of the robot's head as soon as it finished its actions. This was used instead of the real robot's voice because we could not ensure that all the remotely located participants had access to speakers/headphones. Furthermore, the focus of our study was on the behaviour of the robot in particular, and the presence of voice could have impacted how participants perceived the robot, depending on variables such as gender, age, pitch, and tone of the robot's voice [29, 34, 39].

After seeing the speech bubble for a few seconds, a list of measures appeared (similar to [2], including one attention check), which are not discussed in this paper. There was also a scale to rate the actions of the robot while practising, from the robot "made a very small mistake" to "made a very big mistake", or "made no mistake".

Step 5 - Final evaluation: After completing six rounds of training+practising, the participants were asked to answer a few questions to evaluate the scenario and the robot, as well as to indicate their expected intention to use the robot for the same or another task:

- "Assume you will have some guests tonight. You are very busy, and no one in your household has time to help you. Would you allow this robot to prepare dinner for tonight?". The purpose of this question was to implicitly determine participants' intention to use the robot for doing a task similar to what they had previously given instructions for. The choices were: "Yes, it can do it alone", "Yes, but together with me.", "No, I will order from a restaurant (e.g., with Uber Eats).", or "No, I will manage to do it myself."
- "Assume you have a very busy week ahead. Would you allow this robot to do your laundry?". To further study our research question, we asked this to evaluate the extent to which participants' intention to use can be transferred to a new task that also involves the robot. The same options as in the previous question were offered again, replacing the third option with "No, I will request help from others or use a service."

- Whether the robot “*improved a lot*” or “*did not improve at all*” was asked on a continuous scale from 0 to 1000.
- Whether the teaching scenario “*looked very realistic*” or “*not realistic at all*” was also questioned on a continuous slider from 0 to 1000. The purpose of this question was to provide some insight into the applicability of our results to real-world applications.
- “*How successful do you think this robot would be in teaching cooking tasks to another robot?*” (rated on a continuous scale from 0 to 1000)
- “*If you had this robot in your home, how likely would you be to use it to assist you with chores?*” (rated on a continuous scale from 0 to 1000)

Two attention checks were included here: one asked if participants believed Tuesdays are weekends in the USA and Canada, and the other one asked about the different types of food the robot practised preparing. On the last page, participants were asked how much they like cooking and doing laundry, and what proportion of their “weekly meals”/“monthly laundry” they cook/do themselves, to further control for possible confounding factors.

4.3 Participants

Amazon Mechanical Turk was used to collect data from 221 participants. Participants were recruited from the USA and Canada and had completed more than 50 tasks already with at least a 95% approval rate. A 2 USD base payment, plus a 1 USD bonus prorated based on the percentage of the questions they answered was paid to the participants. The experiment took an average of 15–20 minutes to complete. Our recruitment started by paying half of those amounts, but we realized that the task was taking participants longer than we initially estimated, so we increased the payment. The study received full ethics clearance from the University of Waterloo Human Research Ethics Committee.

All the responses were reviewed to check participants’ attention. In addition to the attention checks in the questionnaires presented after each practising round, and those in the DT (i.e., Disposition of Trust) and post-experimental questionnaires, we considered the criteria used in [1] regarding the spotted errors in the practising rounds as extra attention checks. For better consistency within the experimental groups that involved ending errors, we filtered data to only include individuals who rated those errors at least 750/1000 in severity. Taking all these factors into account, the data from 176 participants were included in the analyses. Table 1 summarizes the characteristics of our sample. It is worthwhile noting that this sample size allows us to capture only intermediate to large effects ($f = 0.25$), at $\alpha = 0.05$ and $1 - \beta = 0.95$.

5 RESULTS

When asking participants, after training the robot, whether they would want the robot to cook their dinner and do their laundry in future, both ending error and verbal expressions seem to have affected the choices. Figure 4 plots the percentage of answers in each group. Responses are categorized based on either positive or negative intentions toward using the robot (i.e., having the robot involved in the task, or ignoring it). When there was no ending error, participants’ intention to use seemed to be lower in the modes

Table 1: Distribution and characteristics (age and gender) of participants in every group

Group	Robot behaviour	Robot verbal mode	Frq.	Age	Gender
1	With ending error	NS	30	[23,70], M=39.0	M:12 F:16
2		SI	28	[18,56], M=33.6	M:16 F:12
3		SM	30	[19,65], M=35.8	M:15 F:15
4	Without ending error	NS	28	[20,67], M=37.6	M:17 F:11
5		SI	29	[19,66], M=33.0	M:23 F:6
6		SM	31	[19,56], M=31.4	M:12 F:18
Total:			176	[18,70], M=35.1	M:101F:72

in which the robot speaks (SI and especially SM) compared to the mode without speech (NS). Contrary to that trend, when there was an ending error, participants appeared to have a higher intention to use the robot with verbal expressiveness modes that included robot speech rather than when it was silent. According to chi-squared tests, introducing an ending error significantly reduced the intention to use the robot for both cooking and laundry tasks when it was silent ($\chi^2(1) = 11.621, p < .001$ and $\chi^2(1) = 7.095, p < .001$, respectively). However, in the verbal expressiveness modes with the robot speaking, we could not detect any significant differences between conditions with and without the ending error.

We used Generalized Linear Models (GLMs) [32] with a binomial family to further explore these two measures while accounting for the confounding factors. The verbal expressiveness mode, ending error condition, participants’ demographics (i.e., age and gender), personality (based on TIPI) and how much they liked cooking/laundry, as well as their disposition of trust (based on the DT questionnaire) were initially included in every model. In the end, only a subset of those factors was kept in the models to minimize the Akaike’s Information Criterion (AIC) [8]. Interaction effects were studied subsequently. The final models for each task and for each ending error condition are shown in Tables 2, 3, 4, and 5.

Without any ending error, the only factors affecting intention to use the robot in the cooking task were participants’ gender and their agreeableness. In that case, male participants and those with more agreeable personalities were more likely to want to use the robot ($se = 0.57, z = 2.633, p < .01$ and $se = 0.20, z = 2.414, p < .05$, respectively). We could not detect any factor significantly influencing the intention to use the robot in the laundry task when no ending error was present.

With an ending error, however, the robot’s verbal expressiveness mode could impact participants’ intention to use it. Compared with when the robot was silent (NS), participants were significantly more intent to use the robot ($se = 0.62, z = 2.083, p < .05$) for a cooking task when it spoke and mentioned its error (SM). A trend, approaching significance, was also observed regarding the same task suggesting that when the robot was only speaking (SI) participants’ intention to use it improved ($se = 0.63, z = 1.622, p = .052$) compared with the silent robot. Concerning the laundry task, both these trends were significant: SM and SI expression modes could improve intention to use ($se = 0.71, z = 1.843, p < .05$ and $se = 0.73, z = 2.178, p < .05$, respectively) compared with the NS mode. A significant effect of participants’ gender was also observed

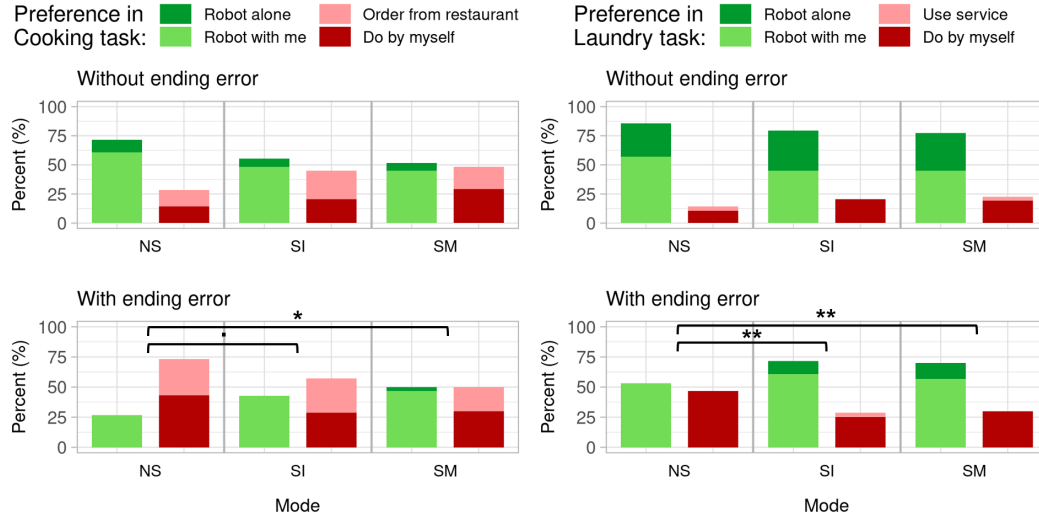


Figure 4: Preferences of participants in different groups regarding the robot doing their cooking and laundry tasks. · = $p < .1$; * = $p < .05$; ** = $p < .01$.

Table 2: Generalized Linear Model (GLM) predicting participants' preference in cooking task without ending error

Covariate	Estimate	SE	z	p
Age	-0.045	0.024	-1.820	.068 ·
GenderMale	1.507	0.572	2.633	.009 **
GenderNA	16.110	1455	0.011	.991
Competence	0.472	0.300	1.572	.116
Like cooking	-0.002	0.001	-1.930	.054 ·
Agreeableness	0.486	0.201	2.414	.016 *

GenderFemale is the baseline.

· = $p < .1$; * = $p < .05$; ** = $p < .01$.

Table 3: GLM predicting participants' preference in cooking task with ending error

Covariate	Estimate	SE	z	p
ModeSI	1.020	0.629	1.622	.052 ·
ModeSM	1.285	0.617	2.083	.019 *
Trusting Stance	0.424	0.236	1.801	.072 ·
Extraversion	0.238	0.155	1.537	.124
Agreeableness	0.162	0.207	0.783	.434

ModeNS and GenderFemale are the baselines.

· = $p < .1$; * = $p < .05$.

for this task, male participants were less intent to use the robot for doing their laundry ($se = 0.61, z = -2.741, p < .01$). Those with a higher trusting stance and also greater openness to experience were found to have a higher intention to use the robot for the laundry task ($se = 0.34, z = 2.146, p < .05$ and $se = 0.23, z = 3.050, p < .01$, respectively). On the other hand, participants with higher emotional stability tended to want to use the robot less ($se = 0.34, z = -2.146, p < .05$).

There were four other measures, rated on continuous scales, evaluating aspects of the teaching scenario and the robot. A similar

Table 4: GLM predicting participants' preference in laundry task without ending error

Covariate	Estimate	SE	z	p
Often cooking	-0.002	0.002	-1.508	.132
Extraversion	0.349	0.198	1.761	.078 ·
Agreeableness	0.364	0.248	1.469	.142
Emotional Stability	0.473	0.247	1.912	.056 ·

· = $p < .1$.

Table 5: GLM predicting participants' preference in laundry task with ending error

Covariate	Estimate	SE	z	p
ModeSI	1.586	0.728	2.178	.015 *
ModeSM	1.302	0.707	1.843	.033 *
GenderMale	-1.664	0.607	-2.741	.006 **
GenderNA	0.212	1.640	0.129	.897
Benevolence	-0.493	0.315	-1.562	.118
Trusting Stance	0.730	0.340	2.146	.032 *
Emotional Stability	-0.431	0.210	-2.048	.041 *
Openness to Experiences	0.694	0.228	3.050	.002 **

ModeNS and GenderFemale are the baselines.

* = $p < .05$; ** = $p < .01$.

approach, this time with Linear Mixed-effects Models (LMMs) [6] was used to check for the significant effects while considering the other possible confounding factors that were included in the GLMs. We found that the perceived realism of the teaching scenario (rated 656/1000 on average) was only correlated with participants' agreeableness ($se = 13.21, t = 1.988, p < .05$) and their level of trust in other people's competencies ($se = 18.99, t = 2.524, p < .05$). The

Table 6: Linear Mixed-effects Model (LMM) predicting perceived success of the robot in teaching another robot

Covariate	Estimate	SE	t	p
ConditionWith ending error	-253.00	35.75	-7.078	.000 ***
ModeSI	-45.22	46.26	-0.978	.330
ModeSM	92.55	45.22	2.047	.042 *
Age	-2.94	1.66	-1.786	.076 ·
Extraversion	32.91	11.20	2.939	.004 **
Agreeableness	32.50	14.50	2.242	.026 *
Competence	42.84	20.04	2.138	.034 *

ModeNS and ConditionWithout ending error are the baselines.

· = $p < .1$; * = $p < .05$; ** = $p < .01$; *** = $p < .001$.

Table 7: LMM predicting participants' expected likelihood of using the robot to assist with chores

Covariate	Estimate	SE	t	p
ConditionWith ending error	-325.16	70.79	-4.594	.000 ***
ModeSI	-192.95	73.21	-2.635	.009 **
ModeSM	-168.68	70.69	-2.386	.018 *
Openness to Experiences	32.11	15.63	2.055	.042 *
Competence	39.99	22.91	1.746	.083 ·
ConditionWith ending error: ModeSI	229.06	101.32	2.261	.025 *
ConditionWith ending error: ModeSM	297.23	98.39	3.021	.003 **

ModeNS and ConditionWithout ending error are the baselines.

· = $p < .1$; * = $p < .05$; ** = $p < .01$; *** = $p < .001$.

perceived improvement of the robot (rated 622/1000 on average) was found to be affected only by its verbal expressiveness mode. When the robot was speaking but ignored its errors, participants perceived less improvement in its behaviour compared to when it was silent ($se = 43.973, t = -2.176, p < .05$).

Because the perceived success of the robot in teaching cooking tasks to another robot and the expected likelihood of using the robot for assistance in the future were affected by multiple factors, the LMMs predicting those two measures are presented in Tables 6 and 7. The former measure was significantly reduced by the ending error of the robot ($se = 35.746, t = -7.078, p < .001$). A robot that was speaking and mentioning its errors was perceived as being more successful in teaching another robot compared to both other modes of speaking, i.e., ignoring the errors ($se = 43.44, t = 3.172, p < .001$) or remaining silent ($se = 45.215, t = 2.047, p < .05$). Here, there was almost no difference between the silent robot and speaking robot when ignoring the errors ($se = 46.26, t = 0.978, p = .835$). More extroverted participants ($se = 11.20, t = 2.939, p < .01$), those with more agreeableness ($se = 14.50, t = 2.242, p < .05$), and participants who trusted more in other persons' competencies ($se = 20.04, t = 2.138, p < .05$) also perceived the robot as more successful in teaching.

Finally, in the direct measure of the likelihood of using the robot in the future, we found an effect suggesting that ending error could reduce this measure as well ($se = 70.79, t = -4.594, p < .001$). According to the LMM (see Table 7), the interactions between the robot verbal expressiveness mode and ending error condition were significant. Therefore, one-sided pair-wise tests with respect to

each level of ending error were conducted to study the differences. It was found that when there was an ending error, a robot utilizing speech and mentioning the errors scored a higher likelihood of use ($se = 69.27, t = 1.856, p < .05$) compared to the silent robot. The same trend, close to being statistically significant, was also observed for the robot that was speaking but ignoring its errors compared with a silent one ($se = 70.22, t = 1.316, p = .095$). In contrast, when there was no ending error, a robot utilizing speech and mentioning the error resulted in a lower likelihood of intention to use ($se = 70.70, t = -2.386, p < .01$) compared to the silent one. The same effect was observed for the robot that was speaking but ignoring its errors compared to a silent one ($se = 73.21, t = -2.635, p < .01$). Participants' openness to experiences was the only personality trait affecting the likelihood of using the robot in the future, with a positive correlation ($se = 15.63, t = 2.055, p < .05$).

6 DISCUSSION

The aim of this study was to examine whether and how different types of verbal expressiveness of a student robot would influence participants' (as human teachers) intention to use the robot after the occurrence of an error at the end of the teaching process. We compared three different verbal expressiveness modes: being silent (NS), speaking but ignoring the errors (SI), and finally speaking while mentioning the errors (SM).

In the absence of an ending error, i.e., when the teaching process proceeded smoothly and ended flawlessly, a silent robot was more likely to be used in the future and might be more preferred to cook dinner and do laundry according to participants' ratings. Considering two other modes that included speech, the robot in the silent mode was more efficient in terms of task completion time since there was no pause after completing the task for giving the explanations or discussing the next step. Note, the silent robot might have been less distracting because it was merely practising performing the task without any need for the teacher to pay attention to the verbal cues, as also confirmed in the study of Fraczak et al. [15]. In the case there was no ending error, and thus teaching appeared to be successful, those positive aspects associated with remaining silent could have been more important for participants to decide to use the robot. Besides, by justifying every error in the SM mode, even when they did not happen at a critical moment (e.g., some errors occurred during early stages of learning and may therefore have been forgotten later on), the robot revealed some information about the limited reliability of its internal mechanisms and alerted the participants that those can fail in some situations. Similarly in [14], Engelhardt et al. found that including apology and justification for failures leads to reduced scores in robots' perceived intelligence. This negative impression of the robot's reliability and intelligence could have possibly contributed to lowering intention to use in our discussed case [21, 23].

In the presence of an ending error, observing participants' preferences for using the robot in both cooking and laundry tasks, and also considering the results from the direct question about the likelihood of using the robot in the future suggest that robots with speech expressions (both SI and SM) were more likely to be used rather than a silent one. As we expected, expressing acknowledgement of the error, justification, and an apology, resulted in a

noticeable improvement in the intention to use the robot after an error (**H2** confirmed). However, there were no statistically significant differences between the SI and the SM mode. Therefore, we reject **H1**, which predicted robot speech without mentioning the error will not improve participants' intention to use it after an error. We found that speaking of any kind could possibly improve the intention to use.

Although lack of statistical power for detecting smaller effects might be a reason why this study could not find any differences between SI and SM modes, the HRI literature can suggest some plausible explanations for the positive impact of robot speech regardless of its content. It is vital in developing and maintaining human-human relationships to verbally communicate feelings, thoughts, opinions, and information to reduce uncertainty [11]. Lee et al. have demonstrated that such self-disclosures expressed also by conversational agents are positively correlated with interactional enjoyment [27]. The connection between robot speech and perceived enjoyment can also be explained by the fact that when an agent talks, it might be seen as more human-like, and that could positively affect the enjoyment of the interaction [17, 42]. Enjoyment is one of the key factors determining the intention to use technology [44, 45], especially conversational agents [22], and likewise robotic systems [27]. Improved enjoyment, even in the SI mode, might have led to the higher intention to use compared with the silent (NS) mode.

Nevertheless, mentioning its errors was the only robot verbal expressiveness mode that could significantly improve the perceived success of the robot in teaching the cooking tasks to another robot as a proxy of the perceived competence of the robot. This implies that participants could detect some differences in terms of robot attributes between the SM mode and the two other ones, especially SI. Further investigation is indeed required to explore the effects of the content of the speech on the intention to use.

Note, for future applications, a robot being seen as more human-like could in fact be detrimental since no robots at present can, autonomously, show truly human-like intelligence, social interaction, and communication capabilities, which could elicit the uncanny valley effect [10, 28, 31]. There is a fine line in trying to improve human-robot interaction, and not to deceive users by pretending that the robot possesses abilities that it actually lacks.

7 LIMITATIONS AND FUTURE WORK

Due to COVID-19, we faced the impossibility of carrying out in-person studies as originally planned. However, the crowd-sourcing framework had benefits too, allowing recruitment of and data collection from a large number of participants safely during the lockdown situation. In addition, no direct face-to-face engagement of the participants with the experimenters may reduce the risk of experimenter bias [35]. This approach also allowed us to control for potential differences that could add noise in a real scenario (e.g., the angle of the camera was fixed, so we were sure that the participants can always see the labels). To evaluate the experience of participants as teachers, they were asked about the realism of the learning scenario, and the rating for that measure was not found to be affected by any factor other than individual personality differences. Comments referring to a tangible aspect of the scenario submitted by a number of participants, e.g., "*It tried to poison me*

twice", may also illustrate their overall believable feeling about the task. It is important to note that the learning interaction we studied has two aspects: the way people experience the robot, and the way the robot learns. The first aspect can remain unchanged in a real-world scenario with robot motor actions pre-programmed for object manipulation, leaving the participants with only the responsibility of showing the robot which items to add (similar to our study). For the second aspect, however, we used pre-recorded video instead of a real learning platform to allow the remote study with a large number of participants. Indeed, in-person studies are needed to validate the findings of our study.

Also, the virtual format of the experiment conducted here produces further limitations. Most importantly, as mentioned before, we had to replace the real voice of the robot with speech bubbles. After hearing the robot's verbal expressions generated by its built-in text-to-speech module, participants may view the robot differently, compared with reading the contents from a speech bubble. In the latter case, the form of expressions used may remind participants of what they see in comic books or cartoons. The presence of real characters in those media can possibly influence the expectations of people about the robot. Making use of the robot's own capabilities, although they might be more limited, can help resolve this problem. The choice of robot speech is also a sensitive issue [3, 5, 41]. We tried to present the robot as gender- and culturally-neutral as possible, but a real or artificial robot voice is likely to be interpreted according to gender, age, culture, accent, etc.

8 CONCLUSION

Adding to previous research showing that even errors made by a trainee robot during the learning process can negatively impact intention to use the robot, this study tested whether those negative impacts can be mitigated by the use of appropriate verbal expressions. We observed that including robot speech in the interaction might not always be a good strategy, especially in the case of having a learner robot that may show some faulty actions in the earlier stages of learning, but does not make any errors beyond that point. Without any errors after completely learning a task, apology, acknowledgement of past errors, and justification for those errors led to a trend suggesting that a silent robot was more likely to be used later. Still, when errors occurred after the teaching interaction, when the robot might have been expected to perform correctly, verbal expressions were significant in increasing intention to use the robot for the future tasks.

The challenge is that when a robot is learning a task, it cannot predict whether errors will happen later or not. So, depending on parameters such as the nature of the learning task and the method used, if the robot anticipates that errors are not likely to happen, it might be better for the robot to remain silent after its early performances that may contain errors. Otherwise, a verbal expression mode that acknowledges errors could enhance participants' intention to use if anything unexpected happens at a later time.

ACKNOWLEDGMENTS

This research was undertaken, in part, thanks to funding from the Canada 150 Research Chairs Program.

REFERENCES

- [1] Pourya Aliasghari, Moojan Ghafurian, Chrystopher L. Nehaniv, and Kerstin Dautenhahn. 2021. Effect of Domestic Trainee Robots' Errors on Human Teachers' Trust. In *2021 30th IEEE International Conference on Robot Human Interactive Communication (RO-MAN)*. 81–88. <https://doi.org/10.1109/RO-MAN50785.2021.9515510>
- [2] Pourya Aliasghari, Moojan Ghafurian, Chrystopher L. Nehaniv, and Kerstin Dautenhahn. (in press). Impact of Nonverbal Robot Behaviour on Human Teachers' Perceptions of a Learner Robot. Accepted for publication in *Interaction Studies*.
- [3] Sean Andrist, Micheline Ziadde, Halim Boukaram, Bilge Mutlu, and Majd Sakr. 2015. Effects of Culture on the Credibility of Robot Speech: A Comparison between English and Arabic. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)*. ACM, 157–164. <https://doi.org/10.1145/2696454.2696464>
- [4] Sule Anjomshoae, Amro Najjar, Davide Calvaresi, and Kary Främling. 2019. Explainable Agents and Robots: Results from a Systematic Literature Review. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '19)*. International Foundation for Autonomous Agents and Multiagent Systems, 1078–1088. <https://doi.org/10.5555/3306127.3331806>
- [5] Franziska Babel, Johannes Kraus, Linda Miller, Matthias Kraus, Nicolas Wagner, Wolfgang Minker, and Martin Baumann. 2021. Small Talk with a Robot? The Impact of Dialog Content, Talk Initiative, and Gaze Behavior of a Social Robot on Trust, Acceptance, and Proximity. *International Journal of Social Robotics* (01 2021). <https://doi.org/10.1007/s12369-020-00730-0>
- [6] Douglas Bates, Martin Mächler, Benjamin M. Bolker, and Steven C. Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1 (Oct. 2015). <https://doi.org/10.18637/jss.v067.i01>
- [7] Aude Billard, Yann Epars, Sylvain Calinon, Stefan Schaal, and Gordon Cheng. 2004. Discovering optimal imitation strategies. *Robotics and Autonomous Systems* 47, 2–3 (2004), 69–77. <https://doi.org/10.1016/j.robot.2004.03.002>
- [8] Hamparsum Bozdogan. 1987. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika* 52, 3 (1987), 345–370. <https://doi.org/10.1007/BF02294361>
- [9] Sylvain Calinon and Aude Billard. 2007. Learning of gestures by imitation in a humanoid robot. In *Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social and Communicative Dimensions*, Chrystopher L. Nehaniv and Kerstin Dautenhahn (Eds.). Cambridge University Press, 153–178. <https://doi.org/10.1017/CBO9780511489808.012>
- [10] Elizabeth Cha, Anca D. Dragan, and Siddhartha S. Srinivasa. 2015. Perceived robot capability. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 541–548. <https://doi.org/10.1109/ROMAN.2015.7333656>
- [11] Nancy L. Collins and Lynn Carol Miller. 1994. Self-Disclosure and Liking: A Meta-Analytic Review. *Psychological Bulletin* 116, 3 (1994), 457–475. <https://doi.org/10.1037/0033-2909.116.3.457>
- [12] Filipa Correia, Carla Guerra, Samuel Mascarenhas, Francisco S. Melo, and Ana Paiva. 2018. Exploring the Impact of Fault Justification in Human-Robot Trust. In *Proceedings of the 17th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS '18)*. International Foundation for Autonomous Agents and Multiagent Systems, 507–513. <https://doi.org/10.5555/3237383.32374598>
- [13] Munjal Desai, Kristen Stubbs, Aaron Steinfeld, and Holly Yanco. 2009. Creating trustworthy robots: Lessons and inspirations from automated systems. In *Proceedings of the AISB Convention: New Frontiers in Human-Robot Interaction*.
- [14] Sara Engelhardt, Emmeli Hansson, and Iolanda Leite. 2017. Better Faulty than Sorry: Investigating Social Recovery Strategies to Minimize the Impact of Failure in Human-Robot Interaction. *CEUR Workshop Proceedings* 1943 (2017), 19–27.
- [15] Piotr Fratzczak, Yee Mey Goh, Peter Kinnell, Laura Justham, and Andrea Soltoggio. 2021. Robot apology as a post-accident trust-recovery control strategy in industrial human-robot interaction. *International Journal of Industrial Ergonomics* 82 (2021), 103078. <https://doi.org/10.1016/j.ergon.2020.103078>
- [16] Metta G, Natale L, Nori F, Sandini G, Vernon D, Fadiga L, von Hofsten C, Rosander K, Lopes M, Santos-Victor J, Bernardino A, and Montesano L. 2010. The iCub humanoid robot: An open-systems platform for research in cognitive development. *Neural networks : the official journal of the International Neural Network Society* 23, 8–9 (10 2010), 1125–1134. <https://doi.org/10.1016/J.NEUNET.2010.08.010>
- [17] Moojan Ghafurian, Neil Budnarain, and Jesse Hoey. 2019. Improving Human-ness of Virtual Agents and Users' Cooperation through Emotions. (2019). arXiv:1903.03980
- [18] Samuel D. Gosling, Peter J. Rentfrow, and William B. Swann. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality* 37, 6 (2003), 504–528. [https://doi.org/10.1016/S0092-6566\(03\)00046-1](https://doi.org/10.1016/S0092-6566(03)00046-1)
- [19] Adriana Hamacher, Nadia Bianchi-Berthouze, Anthony G. Pipe, and Kerstin Eder. 2016. Believing in BERT: Using expressive communication to enhance trust and counteract operational error in physical human-robot interaction. *25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2016* (2016), 493–500. <https://doi.org/10.1109/ROMAN.2016.7745163>
- [20] Cory J. Hayes, Maryam Moosaei, and Laurel D. Riek. 2016. Exploring implicit human responses to robot mistakes in a learning from demonstration task. *25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2016* (2016), 246–252. <https://doi.org/10.1109/ROMAN.2016.7745138>
- [21] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. 2010. Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model. *International Journal of Social Robotics* 2, 4 (2010), 361–375. <https://doi.org/10.1007/s12369-010-0068-5>
- [22] Marcel Heerink, Ben Kröse, Bob Wielinga, and Vanessa Evers. 2008. Enjoyment Intention to Use and Actual Use of a Conversational Robot by Elderly People. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction (HRI '08)*. ACM, 113–120. <https://doi.org/10.1145/1349822.1349838>
- [23] Aya Hussein, Sondoss Elsawah, and Hussein A. Abbass. 2020. Trust Mediating Reliability–Reliance Relationship in Supervisory Control of Human–Swarm Interactions. *Human Factors* 62, 8 (2020), 1237–1248. <https://doi.org/10.1177/0018720819879273>
- [24] Dimosthenis Kontogiorgos, Andre Pereira, Boran Sahindal, Sanne van Waveren, and Joakim Gustafson. 2020. Behavioural Responses to Robot Conversational Failures. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20)*. Association for Computing Machinery, 53–62. <https://doi.org/10.1145/3319502.3374782>
- [25] John D. Lee and Katrina A. See. 2004. Trust in automation: Designing for appropriate reliance. *Human Factors* 46, 1 (2004), 50–80. <https://doi.org/10.1518/hfes.46.1.50.30392>
- [26] Min Kyung Lee, Sara Kiessler, Jodi Forlizzi, Siddhartha Srinivasa, and Paul Rybski. 2010. Gracefully Mitigating Breakdowns in Robotic Services. In *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI '10)*. IEEE Press, 203–210. <https://doi.org/10.5555/1734454.1734544>
- [27] SeoYoung Lee and Junho Choi. 2017. Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies* 103 (2017), 95–105. <https://doi.org/10.1016/j.ijhcs.2017.02.005>
- [28] Kf MacDorman. 2006. Subjective Ratings of Robot Video Clips for Human Likeness, Familiarity, and Eeriness: An Exploration of the Uncanny Valley. *ICCS/CogSci-2006 long symposium*, 26–29. <http://www.coli.uni-saarland.de/courses/agentinteraction/contents/papers/MacDorman06short.pdf>
- [29] Conor McGinn and Ilaria Torre. 2019. Can you Tell the Robot by the Voice? An Exploratory Study on the Role of Voice in the Perception of Robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 211–221. <https://doi.org/10.1109/HRI.2019.8673305>
- [30] D. Harrison McKnight, Vivek Choudhury, and Charles Kacmar. 2002. Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. *Information Systems Research* 13, 3 (2002), 334–359. <https://doi.org/10.1287/isre.13.3.334.81>
- [31] Roger K. Moore. 2017. Appropriate Voices for Artefacts: Some Key Insights. *1st International Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots*. http://vihar-2017.vihar.org/assets/papers/VIHAR-2017_paper_8.pdf
- [32] J. A. Nelder and R. W. M. Wedderburn. 1972. Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)* 135, 3 (1972), 370–384.
- [33] Birthe Nessel, David A. Robb, José Lopes, and Helen Hastie. 2021. Transparency in HRI: Trust and Decision Making in the Face of Robot Errors. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21 Companion)*. ACM, 313–317. <https://doi.org/10.1145/3434074.3447183>
- [34] Andreea Niculescu, Betsy Dijk, Anton Nijholt, Haizhou Li, and Sweet See. 2013. Making Social Robots More Attractive: The Effects of Voice Pitch, Humor and Empathy. *International Journal of Social Robotics* 5 (04 2013), 171–191. <https://doi.org/10.1007/s12369-012-0171-x>
- [35] Gabriele Paolacci, Jesse Chandler, and Panagiotis G. Ipeirotis. 2010. Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5, 5 (2010), 411–419.
- [36] Paul Robinette, Ayanna M. Howard, and Alan R. Wagner. 2015. Timing Is Key for Robot Trust Repair. In *Social Robotics*. Springer International Publishing, Cham, 574–583. https://doi.org/10.1007/978-3-319-25554-5_57
- [37] Alessandra Rossi, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L. Walters. 2017. How the Timing and Magnitude of Robot Errors Influence Peoples' Trust of Robots in an Emergency Scenario. In *9th International Conference on Social Robotics*. Springer International Publishing, 42–52. <https://doi.org/10.1007/978-3-319-70022-9>
- [38] Alessandra Rossi, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L. Walters. 2018. The impact of peoples' personal dispositions and personalities on their trust of robots in an emergency scenario. *Paladyn, Journal of Behavioral Robotics* 9, 1 (2018), 137–154. <https://doi.org/10.1515/pjbr-2018-0010>
- [39] Anara Sandygulova and Gregory M. P. O'Hare. 2015. Children's Perception of Synthesized Voice: Robot's Gender, Age and Accent. In *7th International Conference on Social Robotics*. Springer International Publishing, 594–602. https://doi.org/10.1007/978-3-319-25554-5_59
- [40] Maia Stiber and Chien-Ming Huang. 2020. Not All Errors Are Created Equal: Exploring Human Responses to Robot Errors with Varying Severity. In *Companion*

- Publication of the 2020 International Conference on Multimodal Interaction (ICMI '20 Companion)*. ACM, 97–101. <https://doi.org/10.1145/3395035.3425245>
- [41] Cristen Torrey, Susan R. Fussell, and Sara Kiesler. 2013. How a robot should give advice. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 275–282. <https://doi.org/10.1109/HRI.2013.6483599>
 - [42] Michelle M. E. Van Pinxteren, Ruud W. H. Wetzels, Jessica Rüger, Mark Pluy-maekers, and Martin Wetzels. 2019. Trust in humanoid robots: implications for services marketing. *Journal of Services Marketing* 33, 4 (2019), 507–518. <https://doi.org/10.1108/JSM-01-2018-0045>
 - [43] Sanne van Waveren, Elizabeth J. Carter, and Iolanda Leite. 2019. Take One For the Team: The Effects of Error Severity in Collaborative Tasks with Social Robots. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. ACM, 151–158. <https://doi.org/10.1145/3308532.3329475>
 - [44] Viswanath Venkatesh. 2000. Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. *Information Systems Research* 11, 4 (2000), 342–365. <https://doi.org/10.1287/isre.11.4.342.11872>
 - [45] Viswanath Venkatesh, Michael G. Morris, Gordon B. Davis, and Fred D. Davis. 2003. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly* 27, 3 (2003), 425–478. <https://doi.org/10.2307/30036540>