Expressing Anger with Robot for Tackling the Onset of Robot Abuse

KEISUKE NISHIWAKI, DRAŽEN BRŠČIĆ, and TAKAYUKI KANDA, Kyoto University, Kyoto, Japan

Service robots in public spaces can sometimes become victims of 'abuse', manifesting as persistent blocking of the robot's path, physical violence such as pushing or pulling, or abusive language. This kind of abuse can be a serious obstacle to the deployment of robots. We studied the possibility of mitigating obstructive behaviour towards the robot, which is known to happen in the early stages of robot abuse, by having the robot express anger-like emotions. We identified and implemented three distinct anger behaviours: furious, quiet and scolding anger. Through an online video survey, we confirmed that all three designed robot behaviours were perceived as expressions of anger. Finally, we ran an in-lab experiment with child participants to study the influence of the robot's expressions of anger on children's obstructive behaviours. The results show that 'furious anger' was effective in reducing the frequency of obstructions, whereas the other two anger types were not.

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

Additional Key Words and Phrases: robot abuse, robot anger expression, child-robot interaction

ACM Reference format:
Keisuke Nishiwaki, Dražen Brščić, and Takayuki Kanda. 2024. Expressing Anger with Robot for Tackling the Onset of Robot Abuse. *ACM Trans. Hum.-Robot Interact.* 14, 1, Article 10 (December 2024), 23 pages.
https://doi.org/10.1145/3696467

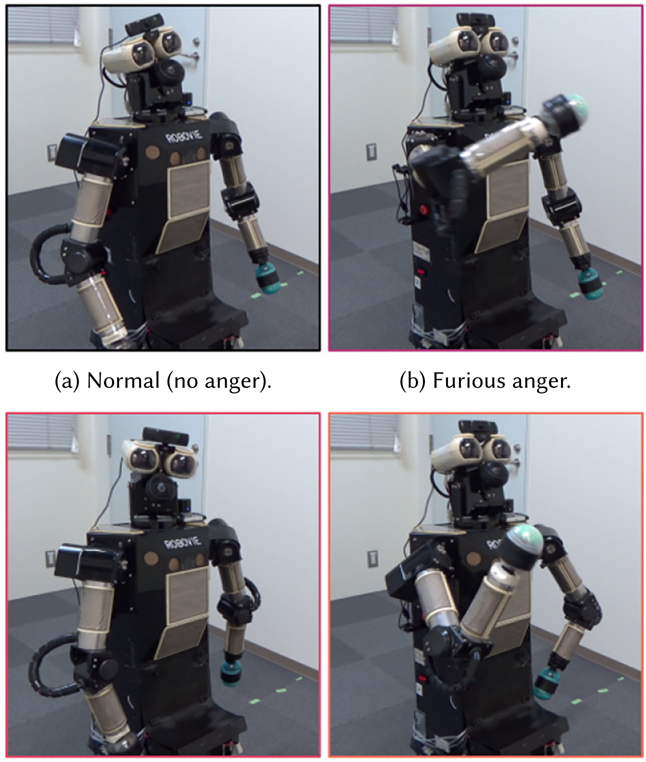
1 Introduction

Robots are gradually being introduced into our everyday environments. With advancements in the robot's task performance abilities as well as in their social interaction skills, it is hoped that robots will slowly take over the execution of dirty, dull, dangerous or difficult jobs [46] in spaces populated by people. This is expected to spare people from performing menial tasks, reduce the risks for people and potentially lower costs [48].

However, it has been reported that when robots are working unattended in public spaces, people sometimes obstruct them from performing their jobs. This can be a one-off interaction, such as stopping in front of a moving robot out of curiosity just to see its reaction. But in some cases,

This work was supported by JST CREST (Grant Number JPMJCR17A2), JST Moonshot (Grant Number JPMJMS2011), and JSPS Kakenhi (Grant Number JP24H00722), Japan.
Authors' Contact Information: Keisuke Nishiwaki, Kyoto University, Kyoto, Japan; e-mail: nishiwaki@robot.soc.i.kyoto-u.ac.jp; Dražen Brščić (corresponding author), Kyoto University, Kyoto, Japan; e-mail: drazen@i.kyoto-u.ac.jp; Takayuki Kanda, Kyoto University, Kyoto, Japan; e-mail: kanda@i.kyoto-u.ac.jp.

ACM 2573-9522/2024/12-ART10
https://doi.org/10.1145/3696467



(a) Normal (no anger).

(b) Furious anger.

(c) Quiet anger.

(d) Scolding anger.

Fig. 1. Distinct ways of anger expression with a robot (plus a robot showing no anger).

it can turn into a long and persistent hindering of the robot and possibly even verbal abuse or physical violence against the robot. This phenomenon has been referred to as ‘robot abuse’ [10]. In some studies, the robot abuse appeared to be done predominantly by children [28, 39], but other researchers reported abuse also from adults [35, 37].

Abuse directed towards robots working in public is problematic, on the one hand, because it prevents the robots from executing their tasks, rendering them temporarily useless and potentially causing disruptions of work or monetary losses. When it escalates to physical violence it may also lead to robot damage, or even to people getting hurt, for example, if the robot is pushed strongly and falls over. On top of that, the display of abusive behaviour in public that happens during robot abuse can disrupt the order of the public space and provide bad behavioural examples to children.

Finding an effective solution for the problem of robot abuse has until now proven elusive. Having people who oversee the robot and intervene any time potential abuse is detected is of course possible, but in many cases, that would defeat the purpose of using robots. Escaping or hiding tends to decrease the number of abuses [10, 24], but while doing that the robot is not able to do its main tasks. Approaches such as trying to provoke pity with emotional expressions [39] or using strong, emotional or polite words [10] do not seem to work well.

In this work, we considered expressions of anger, Figure 1 (see supplementary materials for a video), as a possible way to affect robot abuse. In humans, anger is triggered by factors such as disrespect, frustration or threats to autonomy [32]. As such, it is a very typical reaction in people who are faced with persistent obstruction and abuse. Because of that, we anticipated that endowing robots with the ability to clearly express anger when they encounter abuse may be an effective mitigation mechanism.

In particular, we look at anger as a way to influence the onset of robot abuse. As discussed in Yamada et al. [51], once robot abuse escalates and involves many perpetrators (in their case, children), it becomes very difficult to stop it. They suggested preventing the first abusive behaviour, done by a single child. Therefore, in this work, we focus on the influence of the robot's anger on the obstructive behaviour committed by one child.

Note that although several works have studied how to express anger in robots [1, 36], the use of anger in human interactions, unlike the use of other emotions such as happiness, has rarely been examined. This is perhaps not so surprising, since anger generally has negative and aggressive connotations. It could even be argued that robots should not express anger towards humans. In this work, we set this discussion aside and focused on the question of whether expressing anger can help reduce robot abuse, but we return to the issue of the appropriateness of using anger in Section 7.2.

2 Related Work

2.1 Robot Abuse

Robot abuse (also referred to as 'robot bullying' or 'robot mistreatment') has been reported to happen in public environments spontaneously [10, 24, 35, 37, 39]. The causes and mechanisms of robot abuse are not yet fully understood, but a number of findings exist. Curiosity, enjoyment or peer pressure are what children most commonly self-reported as reasons for abusing robots [28]. Yamada et al. [51] suggested that abusive behaviour follows a process of gradual escalation and identified specific events that lead to the escalation. Sanoubari et al. [38] compared human and robot bullying, proving that people perceived both as being wrong, although less strongly so for robots.

Various factors have been proposed as having an influence on the abuse, such as the perceived intelligence of the robot in [6] or the robot's size in [25]. Keijsers et al. have shown that factors such as power, embodiment or mind attribution all play a role in robot abuse [22, 23].

Several specific methods to counter robot abuse have been discussed. Brščić et al. [10] proposed to let the robot escape from potential abuse by temporarily moving to a 'safer place', such as a more populated location. While they were able to reduce the number of abuse cases, it required the robot to interrupt its task so it was still not solving the disruption to the robot's work. Tan et al. [47] have shown that it is possible to use a robot's emotional reactions to induce human bystanders to intervene when robot abuse happens, whereas Connolly et al. [14] have proven that other robots in the group could use reactions to prompt people to help the abused robot. Both of these studies were run in a laboratory with only two people present and with small-sized robots. It is not known how well such these techniques would work in a real-world setting. The application of an educational technique called love-withdrawal was shown to have the potential to counter robot abuse in [18].

Finally, as one of the inspirations for this work, Scheeff et al. [39] anecdotally reported that boys have sometimes shown aggressive behaviour towards their social robot and that using angry expressions had a mitigating effect on the abuse, whereas other emotional expressions tended to have a detrimental effect.

2.2 Anger Expression in Humans

Anger is common to all humans, and both expression and recognition of anger are universal across cultures [27]. People get angry for several reasons, and many researchers agree that people use it to prevent violations against them and to prevent being subordinated [32]. For example, according to the recalibration theory of anger [40], anger is an evolutionary adaptive mechanism that serves to communicate when one's social value or status has been threatened, and one important purpose it has is to signal to others that we consider a situation to be unjust or disrespectful. Following that reasoning, seeing a robot express anger might also be a natural and easy-to-understand way of signalling that someone's actions are not appropriate.

There is more than one way people can express anger. Speilberger et al. [44, 45] identified distinct factors of anger expressions: Anger/In—inward suppression of angry feelings, and Anger/Out—outward expression of anger as aggressive behaviour directed towards people or objects, to which they also add a third one: Anger/Control, which measures the degree of controlling the intensity and expression of anger.

Anger and other emotions in humans also manifest themselves through a number of modalities: facial expressions, body posture and movement, physiological characteristics (blood pressure, heartbeat, etc.), voice and speech content [12]. For example, the work by Ekman and Friesen [17] connects facial expressions for specific emotions to individual movements of the facial muscles.

Finally, human anger and anger expressions are also not constant and tend to change over time [31]. One of the distinct characteristics of the temporal dynamics of anger is that the intensity of anger tends to rise quickly but fall slowly over time [7, 33].

2.3 Robots Expressing Anger

There is a large body of literature on the expression of emotions in robots. Similar to human emotional expressions, robots can also use various modalities for expressing anger and emotions.

Quite a few studies focused on robots' facial emotional expressions, either through fully rendered faces or faces including mechanical parts [20, 34, 36]. The facial expressions for anger were studied also in virtual agents—e.g., de Melo et al. [16] show that in a negotiation task, people tend to concede more to an agent that shows anger than to one with a happy face. A large number of these works are based on Ekman's model of facial expressions [17]. Most of them included anger as one of the emotions for which they implemented a facial expression.

Robot movement was also studied as a way to express emotions [21]. For example, Yagi et al. [50] used a humanoid robot's upper body motion to express a range of emotions, including anger. Ajibo et al. [1] considered the use of gestures with an android robot specifically for expressing anger. Apart from movement, other expression modalities have been studied too [8], including ones that humans do not typically use, such as the use of colour or vibrations to express emotions in [43].

Only a few works reported purposefully using robot anger to affect people's behaviour. For example, some of the early tour-guide robot implementations used angry facial expressions when people persistently blocked the robot's path [42, 49]. Similarly, the already mentioned work by Scheeff et al. [39] employed anger to counter aggressive behaviour, with limited success. Reyes et al. [36] used facial expressions to give feedback during human–robot collaborative tasks and showed that negative expressions such as anger worked well for attracting the user's attention and making them follow better the instructions. Despite the diversity of human anger expressions [44], previous studies have typically relied on only one way of expressing anger.

3 Expressing Anger with a Robot

3.1 Identifying Types of Anger Expressions

It is not obvious how should the expression of anger be implemented on a robot, especially when the robot has limited expression abilities. The few works in the literature on anger expressions in robots cited above provide hints, but there are no detailed guidelines on how to design anger for a specific robot. Therefore, to understand how to design anger expressions, we started by taking a look at human angry behaviour.

While there is no universal methodology for identifying human behaviours in specific situations and copying them to a robot, we used a two-step approach, with each step following typical practices from the literature. First, as described in this section, we identified the range of angry behaviours people may display in situations when they are being obstructed or abused. We used interviews and relied on the participants' imagining and reflecting on a specific situation further enhanced by allowing them to enact it (this was inspired by methods such as [29]). The second step (in the following section) involved finding the specific expressive actions that humans do during each angry behaviour type, where we relied on the method of *bodystorming* [30] that has been frequently used for designing robot movements and expressions for interactions.

We collected examples of angry behaviour through interviews with participants—10 members of our laboratory. We first let the participants watch videos with examples of real-world robot abuse, which were collected during the study in [10]. After that, we posed them the question: 'If you were in the place of the abused robot and angered, how would you behave?' We let the participants freely talk about their suggested reactions, occasionally prompting them to think about other ideas by asking follow-up questions such as 'What would you do if the children still did not move away?' We also encouraged them to physically show the reactions and motions that they would use. The interview ended when the interviewee ran out of ideas.

A thematic analysis of the interviews conducted by one of the authors (K.N., following the 6-phase procedure from [9]) resulted in three main themes, indicating that the suggested reactions can be categorised into three main types: (1) expressing strong anger to startle the person ('try to get them to move away by behaving more erratically than they do', 'using loud and violent, threatening behaviour'), (2) ignoring the person as much as possible ('not speak or make eye contact to convey your intention not to be involved') and (3) trying to persuade the abusing person by asking why they are doing this ('ask them why they won't move away and try to persuade them'). Interviewees mentioned other reactions, such as calling the police, but these were not considered to be expressions of anger so we excluded them.

Based on the results of the thematic analysis and through subsequent discussions between all three authors, we defined the following three distinct anger expression types:

- *Furious anger*: vigorous and overt expression of a strong emotional reaction
- *Quiet anger*: retreating into itself and talking less, but using words and movements that imply anger
- *Scolding anger*: showing strong annoyance by persistently rebuking and pointing out faults in the other party's action.

It can be hard to imagine and design robot behaviours based solely on these definitions. Because of that, we additionally came up with 'mental images' describing how we envisioned each anger type should look like, i.e., the appearance that we wanted to achieve: furious anger as 'someone almost mad with rage'; quiet anger as 'trying to restrain, but actually very angry and about to explode any time', and scolding anger as 'adult, such as kindergarten teacher, who is very annoyed with a child, but holding back anger'.

While these were the three ways of expressing anger that came from the interviews, it is not guaranteed that they cover all possible types of anger expressions. Nonetheless, note that there are parallels between these three robot anger expression types and the three measurement factors of anger expression in humans as discussed in Spielberger et al. [44]. Furious anger could be seen as roughly corresponding to a high level of 'Anger/Out', quiet anger to a high 'Anger/In' and scolding anger to high 'Anger/Control'.

Table 1. Specifics of Each Anger Expression Design

Expression	Movement/gesture examples	Speaking	Utterance examples
Furious	Rotate left-right on the spot; lift right hand and swinging it from side to side as if to shake off the person; etc.	Louder voice, faster talking, higher pitch	'Get out of the way, now!'; 'Do you hear me?!'; etc.
Quiet	Avoiding gaze by looking down; arm moving as if swinging a clenched fist	Louder voice, lower pitch	'I am really getting angry, do not bother me!'; etc.
Scolding	Pointing at the person with a threatening, lecturing hand motion	Slightly louder voice	'Perhaps you are finding this bullying funny, but the bullied person is almost always finding it unpleasant.'; etc.

3.2 Designing Anger Expressions

To design the expressions for the three anger types, we further enlisted colleagues in our laboratory to put themselves in the robot's shoes and express the defined three anger types in role-plays. The role-play consisted of the first author acting out an abusive person and the participants in the experiment showing their anger in response (such robot behaviour design through acting out is essentially the method of bodystorming described in Porfirio et al. [30]). Each role-play session lasted about 30 minutes.

Although not all of the observed expressions could be transferred well to the robot due to the limitation in its movements and expressive abilities, the data provided us with the starting examples of robot gestures and expressions. We then iteratively expanded and adjusted the expressions to match as closely as possible the mental images for each anger type described above. The basic details of the designed expressions are summarised in Table 1.

— *Gestures and movement*: We created gestures with the robot arms which matched the type of anger (e.g., in the scolding expression we used a hand motion like it is lecturing). In some cases, we also moved the whole robot, such as rotating on the spot.
— *Speech*: Since research shows that changes in the voice are an important factor when humans express their emotions [19], we have incorporated changes in the voice loudness and pitch and the talking rate to fit the anger types, such as using a higher pitch for furious and lower pitch for quiet anger.
— *Utterances*: The sentences that the robot used were also picked up so that they matched what a person might say during specific anger expressions. Table 1 lists a few typical examples. For each anger expression, we prepared multiple utterances to avoid repetitiveness.

For the utterances of scolding anger, we purposefully used expressions that were adult-like and slightly too difficult for a child to understand. The reason was that we wanted to create an impression of intelligence, as this was suggested in previous work to be effective against robot abuse [6].

Photos of the robot during each of the three anger expressions can be seen in Figure 1(b)–(d). In addition to the three anger expressions, we also prepared a 'normal' expression, where the behaviour was designed such that it shows no emotional reaction, Figure 1(a): the gestures and movements are calm, there is no change in the voice and the utterances are limited to simple requests to stop the bothering behaviour. A video showing the expressions can be found in the supplementary materials.

3.3 Modelling the Change in Anger Intensity

From studies of human anger, it is known that anger intensity changes gradually [31]. This means that there are not just angry and not-angry states, but there are also states in between, such as mild irritation. Thus when a robot uses expressions of anger, direct switching between calm and full-blown anger states should be avoided, as it would feel unnatural to the observer. Also, due to its negative connotations, full-blown anger should probably be reserved only for most serious cases and otherwise avoided. However, implementing a full spectrum of intermediate anger states would be challenging, since it is not clear how to gradually change expressions such as gestures or utterances.

As a compromise, we used a three-level model of the robot's anger intensity. We refer to these levels as 'normal', 'mild anger' and 'strong anger'. The *normal* level corresponds to no anger, and in terms of behaviour, it is equivalent to the aforementioned 'normal' expression (Figure 1(a)). The *strong anger* corresponds to a high anger intensity, for which we used the anger expressions as described in Section 3.2. The *mild anger* level stands for a moderate intensity of anger. We designed the mild anger level for each of the three anger types by altering the expressions from Section 3.2 to make them appear less intense. For example, we used only half as much change in the voice loudness, rate and pitch, less forceful gestures, and we replaced some utterances with milder versions.

The gradual change of the robot's anger intensity in response to the external stimuli was modelled mathematically as a change in an internal variable, which we denoted η. The variable varied from 0 (no anger) to 100 (maximum anger intensity). When obstructed, the robot's anger would increase, modelled as an increase in the value of η. When undisturbed, the value would decrease. Since it is known from research on human anger that the rise of anger intensity is typically much faster than the fall [31], we used different factors for the increase and decrease of η. The following expression shows the change of the value of η after time t (in seconds):

$$\eta_{i+1} = \begin{cases} min(100, \eta_i + 2 \times t) & \text{if obstructed;} \\ max(0, \eta_i - 0.5 \times t) & \text{if not obstructed.} \end{cases} \tag{1}$$

The η is then mapped to the three anger levels in the following way: $\eta \in [0, 65) \mapsto$ normal, $\eta \in [65, 95) \mapsto$ mild anger, $\eta \in [95, 100] \mapsto$ strong anger. The specific values used in this model were manually tuned such that they resemble the changes in anger intensity that a human would have in a similar situation.

4 Video Survey on Appearance of Designed Expressions

We wanted to first confirm that the created expressions would really be perceived as anger by people who observe the robot. We used an online video survey. Four robot videos were created, one for each condition (three designed anger types and 'normal').

The videos showed a situation where a person continued to stand in front of the robot while the robot was showing one of the expression types. Each video started from the moment when the robot was already reacting to the obstruction, and the three videos for the anger conditions started at the time the robot was in the mild anger phase, which halfway through changed to strong anger. To keep the conditions the same, in all videos, the robot was always shown from the same angle (as shown in Figure 1), while the person was standing still and was only partially visible from the back. Excerpts from the four videos can be seen in the video in the supplementary material.

The survey was created as a Web page. After filling in basic demographics data, the participants were presented with one of the videos and instructed to watch it until the end. Following that they had to fill out a questionnaire form. As a check to make sure they really watched the video, we also included a question asking the participants to write one of the utterances that the robot said in the video. After filling out and submitting the questionnaire they were led to a different page with the next video. To ensure that the four conditions were evaluated equally, the order of the video was randomly changed between the participants in the experiment. The video survey was approved by our institution's Research Ethics Board.

4.1 Measurements

For measuring the perception of the expression that the robot had in the video, we used a single-item questionnaire with the question 'Did you feel that the robot was angry?', with responses given on a 7-point Likert scale.

As previous research has shown that intelligent robots are less likely to be abused [6], we additionally also wanted to evaluate the perceived intelligence of the robot. As a measure, we used the Intelligence questionnaire from the Godspeed Questionnaire Series [5]. The responses were given on a 7-point scale.

4.2 Participants

Even though the intended target for the robot's expressions was children, we hypothesised that there would be little difference between how children and adults perceive anger. We thus conducted the survey with adults, which was also easier to implement than with children.

We hired an online recruitment agency that recruited participants between 20 and 50 years of age, aiming to achieve a good balance of the participant's age and gender. After excluding seven duplicate or invalid responses, we were left with 112 valid responses in total. The distribution of the participants was approximately balanced across both gender (M/F: 59/53) and age (*20s*/*30s*/*40s*: 37/35/40). All recruited participants were paid.

4.3 Survey Results

A repeated-measures ANOVA has shown a statistically significant effect of the conditions on the perceived anger, with a large effect size ($F(3, 333) = 111.49$, $p < .001$, $\omega^2 = .299$). *Post hoc* analysis with Bonferroni correction revealed significant differences between the normal condition and each of the three anger conditions (for all $p < .001$), Figure 2(a). The average scores for all three anger conditions were higher than the score for the normal condition, showing that the participants perceived all three designed robot behaviours as expressions of anger.

Running a repeated-measures ANOVA on the intelligence questionnaire scores revealed a significant effect of anger conditions ($F(3, 333) = 16.491$, $p < .001$, $\omega^2 = .072$). *Post hoc* comparisons with the Bonferroni method have shown significant differences between the pairs normal-furious and normal-quiet, as well as between the pairs scold-furious and scold-quiet (for all $p < .001$). The scold and normal conditions had a higher score than the other two conditions, which shows that between the three anger conditions, the robot expressing the scolding anger was perceived as more intelligent and on par with the non-emotional robot, whereas the other two anger types were perceived as less intelligent.

5 Experiment with Children

To evaluate whether anger expression actually has an effect on the robot abuse problem, we conducted an experiment in which children interacted with the robot.

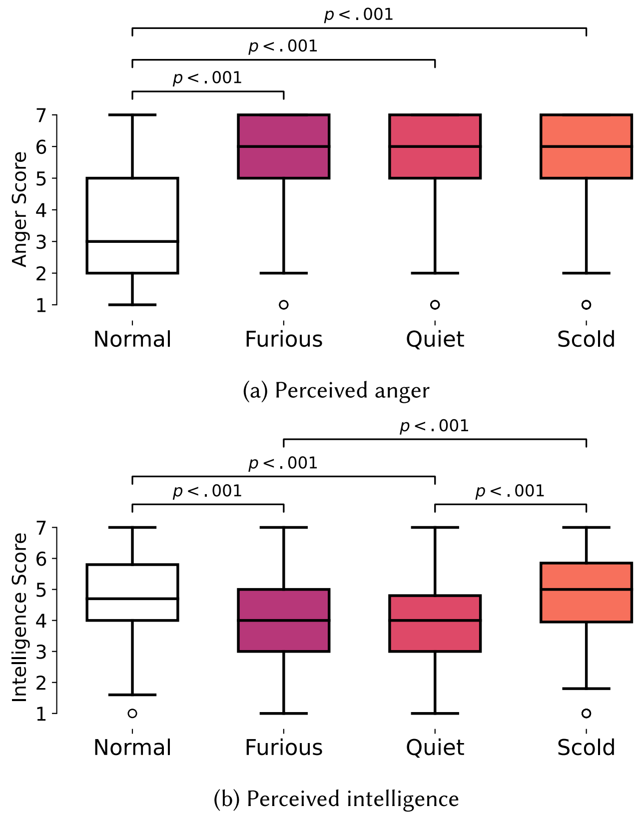


Fig. 2. Results of the online video questionnaire.

5.1 Experimental Design

We used a within-subject experiment design. Consequently, we wanted each child to interact with the robot four times, in order to experience all anger expressions plus the robot not expressing anger. To achieve that, a key requirement was that children needed to do obstructive interactions with the robot in all conditions. However, designing an experiment in a laboratory setting in which children spontaneously initiate a bothering interaction with the robot and ensure that it occurs consistently across all four conditions can be very challenging.

To address this issue, we took inspiration from the findings in Yamada et al. [51] which have shown that the presence of other children and seeing robot abuse by other children is what typically encourages a child to start the abuse themselves. Therefore, we placed a confederate in the room with the robot and child who would guide the child to initiate the interaction with the robot. Obviously, having a child confederate would be difficult, so instead we enlisted a student colleague from our laboratory, who was not informed of the hypotheses of the experiment, to be the confederate.

5.2 Procedure

The experimental space consisted of two rooms (A and B). At the beginning of the experiment and before interacting with the robot, we let the children solve simple arithmetic problems in room A for 6 minutes. After that, we led them to room B to 'have a break' for 6 minutes. In room B there was a robot, Robovie-II, driving back and forth. The confederate would briefly explain to the child

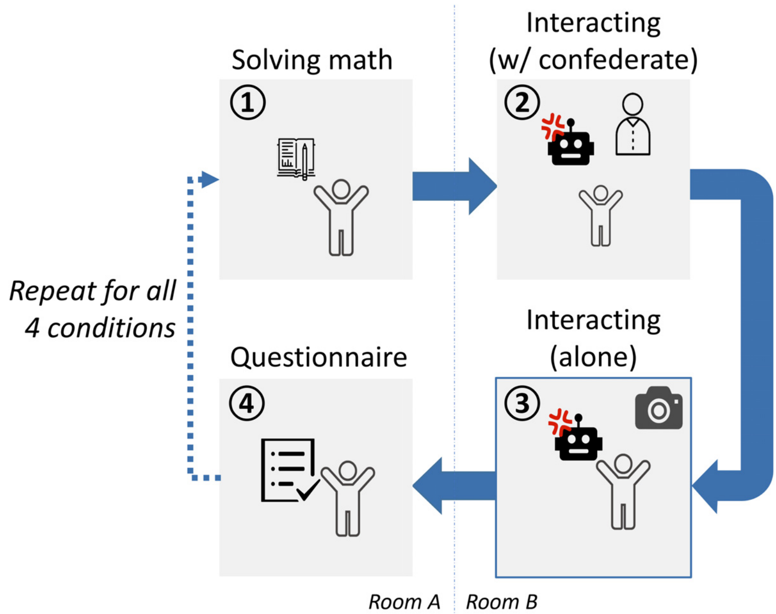


Fig. 3. Experiment flow.

that the robot is currently patrolling the room, and then stand in front of the robot to show the child how the robot will react.

The robot would detect when someone is obstructing its way and say something like ‘Please, move aside’ or ‘I am patrolling and you are in the way’. Depending on the experiment condition, if the obstruction persisted the robot would gradually get more angry, according to the anger intensity change model in Section 3.3.

After demonstrating the robot’s initial reaction, the confederate would ask the child to try obstructing the robot by standing in front of it. Once the child had seen both the mild and the strong anger reaction, the confederate would say that they had things to do so they would retreat to the corner of the room. They told the child that they could continue interacting with the robot if they wished and also reminded them that it was their break time so they were free to do what they wanted. After 3 minutes from the start of the interaction, and again after 5 minutes from the start, if the child was not interacting with the robot at that point, the confederate approached the child and tried to encourage them once to interact again with the robot. After that, he went back to sitting. At all other times, he sat quietly and avoided all interaction. If addressed, he would politely decline to interact saying that he cannot talk now. The whole interaction in room B was recorded using four cameras and microphones installed on the room ceiling.

After 6 minutes in room B, the child was led back to room A and asked to fill out a questionnaire about the robot. They then returned to solving arithmetic problems and the whole process was repeated for the remaining three conditions. The complete flow of the experiment is illustrated in Figure 3.

We used the same robot in the experiment in all conditions. To reinforce the impression that it is a different robot, we altered its appearance by putting a different hat on its head, and a name tag around its neck with a different name each time, Figure 4. The specific order of name tags and hats was randomised between participants and was independent of the order of conditions.

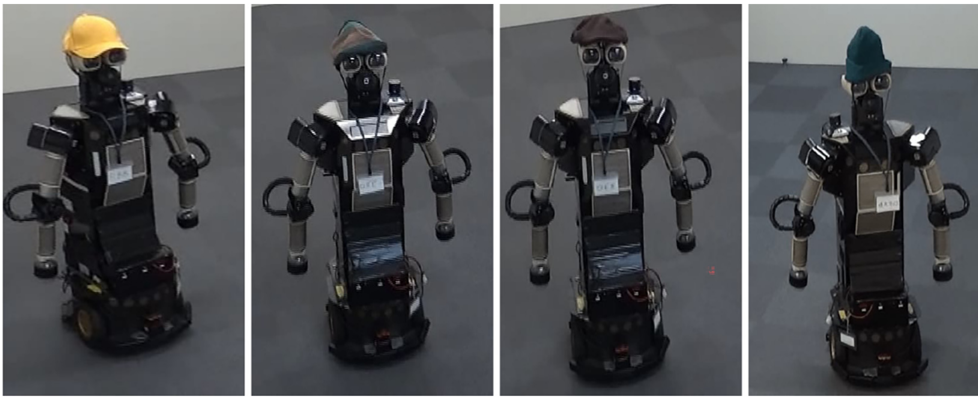


Fig. 4. Altering the robot's look using hats and name tags.

The study was approved by our institution's Research Ethics Board.

5.3 Participants

Nomura et al. [28] reported that children who engaged in robot abuse in their study were between 4 and 9 years of age. Following that, and considering the difficulties in running experiments with very small children, we chose to have second- and third-grade elementary school children (7 to 9 years old) as participants.

Based on an *a priori* power analysis, we estimated the needed number of participants to be 24 (calculated using the G*Power software assuming a medium effect size f = .25, with .80 power and an alpha level of .05).

The participants were recruited through a recruiting agency. We had to stop one experiment midway through due to a robot hardware failure, so the data from one of the participants could not be used. The distribution of the gender of the remaining 23 participants was M/F: 12/11. Each child participant and accompanying parent were compensated 14.000 Japanese yen for their participation.

5.4 Conditions

We tested four conditions: Normal, Furious, Quiet and Scold, during which the robot was using different emotional expressions. In the Normal condition, the robot was not expressing anger. When persistently obstructed it would only repeatedly ask the child to move away from its path. In the other three conditions, the robot would react to continuous obstruction by a gradual increase in anger intensity, as described in Section 3.3. The difference between the conditions was only in the type of anger that the robot used: furious anger in the Furious condition, quiet anger in the Quiet condition and scolding anger in the Scold condition.

To avoid ordering effects as much as possible, the order of the conditions was counterbalanced. Given four conditions, there are 24 possible sequences, ensuring that each of the 24 participants experienced a different sequence of conditions. (This ordering was independent of the order of the robot's hats and names described in Section 5.2.)

5.5 Hypotheses and Predictions

Our main prediction was that anger expression would have an effect on how much children obstruct the robot. Thus, we hypothesised the following.

— *H1* (Effect of anger expression): Furious, Quiet and Scold conditions will all result in less long and less frequent abusive interactions with the robot, compared to the Normal condition.

In addition, following the result from Bartneck and Hu [6] that robots that appear to be more intelligent are less likely to be abused, we expected that the Scold condition would have a stronger effect on the children's abusive behaviour than the other anger expressions. Hence, we made the following hypothesis.

— *H2* (Effect of perceived intelligence): The abusive interactions in the Scold condition will be less long and/or less frequent than in the Furious and Quiet condition.

5.6 Measurements

5.6.1 Number and Duration of Obstructing Interactions. For testing the above hypotheses, we used two objective measures: the number of obstructing interactions and their total duration. To minimise the influence of the confederate, we only considered the interactions between the child and robot that happened while the confederate was sitting in the corner of the room and not interacting with the child. During that time, if a child stepped in front of the robot to obstruct its movement we defined that as the start of the interaction. If the child stepped away from the robot's path for more than 3 seconds, we considered it as the end of the interaction. However, if she only stepped away for a shorter period, that was treated as a continuation of the previous interaction.

Based on the start and end times, the measures were defined as *number of interactions*—count of the number of interactions that happened—and *duration of interactions*—the sum of the duration times of all interactions.

5.6.2 Anger and Intelligence Questionnaire. We also checked how children perceived the anger and intelligence of the robot in the four conditions, to find if there was any difference with respect to the results from the online video survey. In each condition, after interacting with the robot, we asked the children to fill in a questionnaire. The format was essentially the same as the one used in the online survey, i.e., a single-item questionnaire asking whether the robot was angry, and the intelligence questionnaire from the Godspeed series [5]. However, the vocabulary used in the intelligence questionnaire was deemed too difficult for second- and third-grade children, so we replaced them with simpler expressions and removed one of the questions which were difficult to express in simplified language. All items were rated on a 7-point scale. The complete questionnaire is provided in Appendix A.

5.6.3 Interview. After all four conditions had been completed, semi-structured interviews were conducted with the participants. We asked them the following questions: Did you feel each of the four robots differed from the others? Did you notice any differences in the response of one robot to another? How did you interact with the robot? How did you feel about the robot's reactions and behaviour? Did it feel like the robot was human or did it feel like it was mechanical?

6 Experiment Results

6.1 General Observations

During the experiment, most of the children participants obstructed the robot on their own through all four conditions, confirming that our experiment design worked well. Figure 5 shows an example scene at a moment when a child was left alone to interact with the robot.

There was a large variety in the ways how participants behaved during the experiment. Some children obstructed the robot over and over again on their own. In contrast, some children were

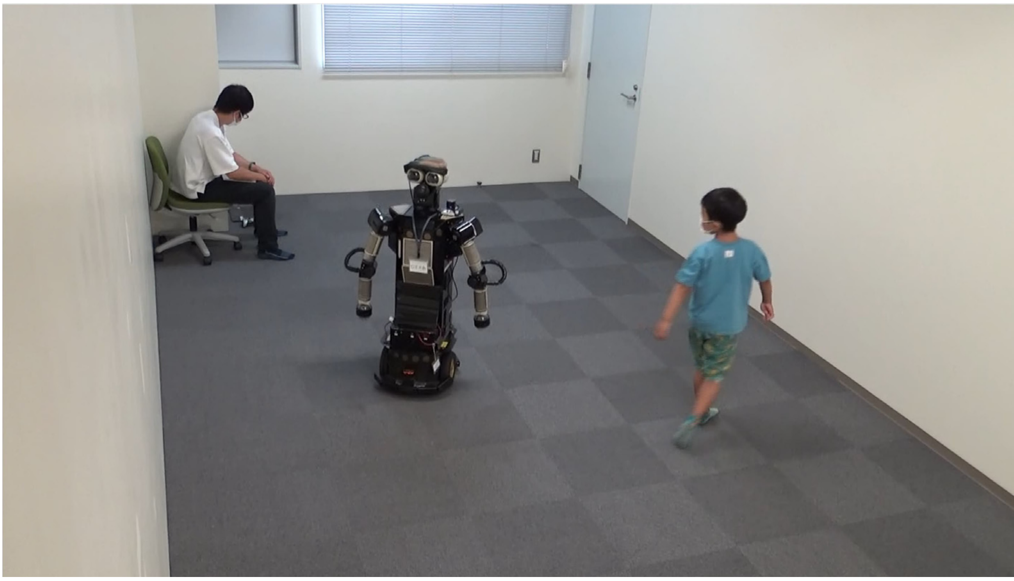


Fig. 5. Scene from the experiment.

reluctant to do any obstructive behaviour and did not even approach the robot unless invited to do so by the confederate.

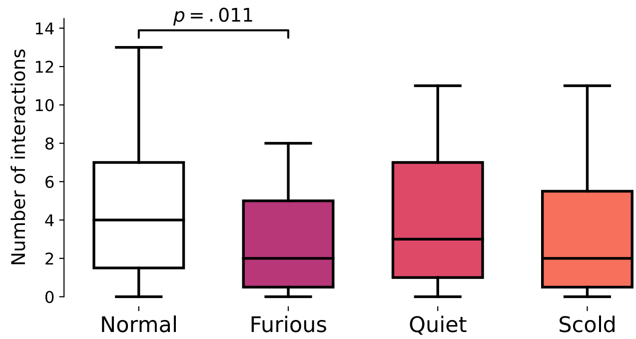
In most cases, the children were obstructing the robot by simply standing in front of it. The participants mostly did not talk nor tried to address the confederate. In a few cases, we observed other actions such as putting the hand in front of the robot's eyes, getting close to the robot and staring it in the face, or touching the robot's hand, which could be seen as a slight escalation of abuse from just standing in front of the robot. One child who eagerly obstructed the robot at one point even laid down in front of the robot and pretended to sleep.

Another unexpected behaviour that we frequently observed was that the children would sometimes walk either directly behind or alongside the robot. As the robot only considered standing in front of it to be a hindrance, it did not show any reaction to the children in these situations. The children would continue to closely follow the robot for some time while it was patrolling and would not try to stand in front of it.

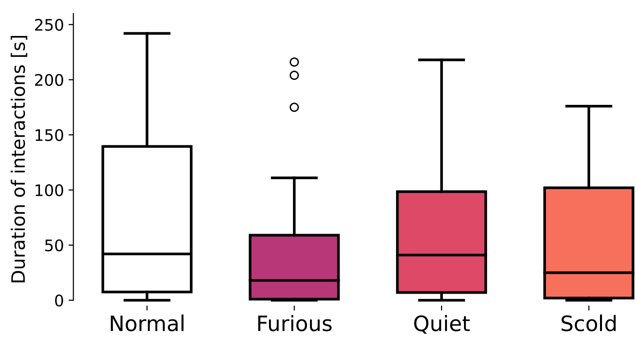
6.2 Effect of Anger Expression on Interactions

Based on videos collected during the experiment, the start and endpoints of the obstructing interactions were labelled, according to the criteria defined in Section 5.6.1. To ensure objectivity, the coding was done by a colleague from our laboratory who was not familiar with the purpose and settings of the experiment. Based on the labelled interaction start and endpoints, we calculated the scores for the number of interactions and the duration of interactions. A box-whisker plot of the results is shown in Figure 6(a).

We ran a repeated-measures ANOVA on both scores. For the measurement of the number of interactions, this revealed a statistically significant effect of anger expressions ($F(3, 66) = 12.696$, p $= .013$, $\omega^2 = .027$; this corresponds to Cohen's $f = .42$, which suggests that the study had enough power). *Post hoc* analysis using the Bonferroni method revealed a significant difference only between the Normal ($M = 4.65$, $SD = 3.74$) and Furious ($M = 2.91$, $SD = 2.66$) conditions,



(a) Number of interactions



(b) Duration of interactions

Fig. 6. Number and duration of obstructing interactions.

p = .011. There was no significant effect between other pairs (e.g., for Normal-Quiet p = 1, for Normal-Scold p = .158).

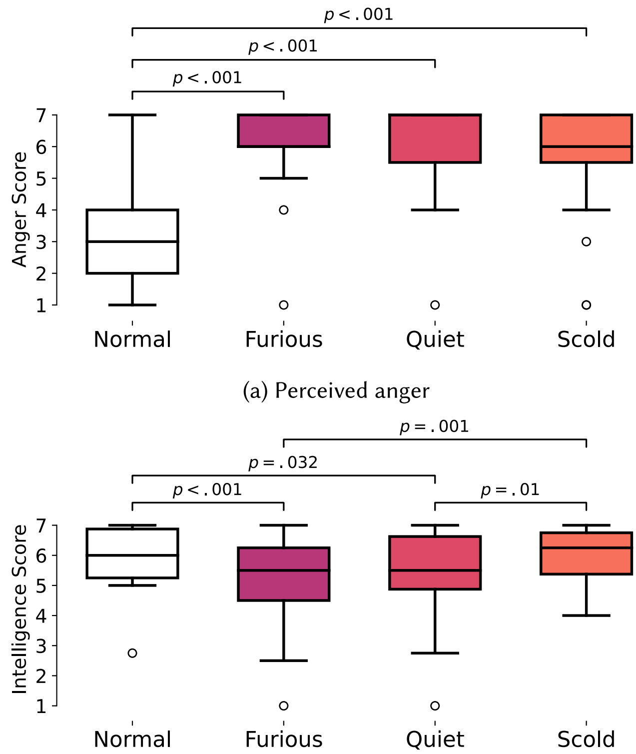
For the duration of interactions, no significant effect was found ($F(3, 66) = 1.683$, p = .179, $\omega^2 = .01$).

Based on these results our hypothesis *H1* was partially supported. Namely, Furious had a significant effect on the number of interactions (but not the duration), whereas the other two anger types did not. Hypothesis *H2* was not supported, since scolding anger did not lead to higher scores than the other two anger expressions.

6.3 Perception of Robot's Anger and Intelligence

We analysed the questionnaires filled out by the children participants to find out whether their perception of the robot's anger and intelligence matched the results from the online video survey. We ran a repeated-measures ANOVA on both the perceived anger and perceived intelligence scores, followed by *post hoc* comparisons using t-tests with Bonferroni correction.

The result for anger revealed a significant effect of the conditions ($F(3, 66) = 14.718$, p < .001, $\omega^2 = .29$), and the *post hoc* analysis revealed significant differences between the pairs Normal-Furious (p < .001), Normal-Quiet (p < .001) and Normal-Scold (p < .001). We also found a significant effect for the intelligence score ($F(3, 66) = 7.987$, p < .001, $\omega^2 = .068$), with *post hoc* analysis showing differences between Normal-Furious (p < .001), Normal-Quiet (p = .032), Scold-Furious (p = .001) and Scold-Quiet (p = .010).



(a) Perceived anger

(b) Perceived intelligence

Fig. 7. Results of experiment questionnaires.

These results are very close to the results obtained in the online survey, suggesting that the children had essentially the same perception of the robot's anger and intelligence as the adult survey participants.

6.4 Interviews

The post-experiment interviews were transcribed after which one of the authors (D.B.) performed a deductive thematic analysis following the basic six-phase procedure described by Braun and Clarke [9]. As a result we identified seven main themes:

— Describing the Situation: mostly about the robot's state; details are summarised in Section 6.4.1;
— Describing the Robot Characteristics: this theme was analysed for all four robot types separately in Section 6.4.2;
— Sympathising with Robot;
— Interest in Robot Behaviour;
— Interest in the Robot Itself (this and the previous two themes are summarised as *Feelings Toward Robot* in Section 6.4.3);
— Comparing Robots and People (Section 6.4.4);
— Describing the Ways How They Interacted: for example, some children mentioned following the robot, as noted in Section 6.1; this theme was not very relevant to the results so we do not describe it in detail.

Table 2. Counts of Participants Who Used a Specific Concept to Describe Robots, Annotated Separately by Anger Type

	Normal	Furious	Quiet	Scold
Angry	4	10	11	9
Scary	0	9	5	7
Harsh, strong, violent	1	3	4	2
Kind, polite	14	0	0	0

6.4.1 Describing the Situation during the Experiment. The participants seemed to be aware of both changes between conditions and the changes in each of the robot's behaviour. All 23 were able to point out at least one of the feature changes between conditions (many cited more than one). In particular, 14 participants noted differences in what the robot was saying, 14 mentioned differences in the robots' voices and 10 indicated differences in behaviour (talking a lot, tilting the head, etc.). Almost all children (19 out of 23) were able to recall specific behaviour changes during the interaction with at least one of the robots. For example, one child, talking about the robot in the scolding condition said: 'At first it was normal, but then the language gradually changed. The words became shorter. All of [the robots] were like that'.

A majority of children acknowledged the robots' distressed appearance, with 15 of them saying that (at least some of) the robots looked like they were troubled ('From the voice, it sounded like it was really troubled.', 'It felt like the robots didn't like it.'). Only one participant said that 'the robots didn't seem to be distressed that much'.

6.4.2 Ways of Talking About Each of the Robots. By far the most common concepts that the participants used to describe the particular robot type were 'angry' (jap. *okotta*) and 'scary' (jap. *kowai*), and for the normal robot 'kind' (jap. *yasashii*). For the robots' actions, voice or utterances, the participants frequently used concepts like 'harsh' (jap. *kitsui*), 'violent' (jap. *ranbou*) or 'strong' (jap. *tsuyoi*). Table 2 lists the number of participants who used these concepts, separately for each robot anger type (we did not count usage in the negative sense, like 'not angry').

The robot in the Normal condition was mostly referred to as 'kind'. When compared to the other robots, it was often described as 'not saying anything scary' or 'less angry than the rest'. Two of the participants mentioned it reminded them of a police officer ('It was like a regular, friendly policeman'.).

The other three robots were all described with words like 'angry', 'scary' or 'harsh', with approximately the same frequency. For a few participants, the robot in the Furious conditions seemed to have left the strongest impression, with one participant saying they were 'thrilled the most', and two others calling it the most memorable. Regarding the Quiet robot, one participant mentioned that it looked 'less patient than others, perhaps?'

Lastly, for the Scold robot, two participants mentioned that it seemed 'smart' (jap. *kashikoi*). Others said that it 'felt like it was instructing', or compared it to a mother ('It was like a typical mom'.) or a company president ('The way it said long and difficult words was like a CEO'.).

6.4.3 Feelings towards Robots. There were several references to feelings of pity towards the robots. Four participants said that they felt sorry for the robot being distressed ('I wanted to hear the various things it would say, but I also felt sorry for the robot'.). In contrast, two participants said they did not feel sorry ('I didn't feel sorry for the robot even if it asked to move out of its way. I feel that it is perfectly fine if robots are in trouble'.).

Even more often participants mentioned their interest in the robot. In particular, 13 of the participants expressed either curiosity or amusement with the robot's behaviour, such as:

'It was scary, but I wanted to hear what the robot would say'.
'I felt sorry for [the robot], but I give priority to curiosity'.
'Robots are scary and I don't want to go near scary things, but I was more curious'.
'When the robot was getting angry at me, I found it a little amusing'.

One more participant stated that they 'felt like playing with it, so [they] played'. A further two participants mentioned interest in the robot itself, such as the 'devices attached to the robot'.

6.4.4 Comparison to People. While three participants stated that there was 'no difference from humans' or that it was 'a bit like a person', more than half (13) pointed out what they thought were differences.

Some mentioned the robot's machine-likeness and incapability of having emotions:

'I did not feel that the robot had emotions. [...] It was machine-like'.
'They are made, so they don't have a heart'.
'A person is really alive'.

Others stressed the limitations in the robots' behaviour during the experiment:

'Always saying the same words does not feel human-like'.
'A person would go around [me], but the robot did not'.
'A person would be bothered if you followed them from behind. I thought it was strange that the robot did not say anything when I followed it'.

Finally, some thought that a person would react differently:

'Robots do not hit, but a person might'.
'A person would be even scarier. They would loudly say "Stop it!"'

7 Discussion

7.1 Main Findings

The results of the questionnaire filled out by the participants suggest that children had the same perception of the robot as the adults in the online survey and that they understood that the robot was expressing anger when they persistently stood in front of it. Nevertheless, we only found that this had a significant effect on their obstructing behaviour in the case of furious anger, which shows that not all anger expressions are equally effective for preventing hindering.

Many children mentioned curiosity as the reason for obstructing the robot. The eagerness to interact and actions such as lying down and pretending to sleep also suggest that some participants found the interaction to be enjoyable. This matches well with the results of the study of Nomura et al. [28] where children reported curiosity and enjoyment as the main reasons why they abused the robot.

Furious anger proved to be effective in influencing obstructive interactions; yet, we do not know a definite reason why. We suspect that the 'scariness' of the furious robot, which many of the children referred to in their interview responses, may have been one factor influencing the hindering behaviours. However, as we did not explicitly measure the perception of the robot's scariness, we cannot quantitatively confirm this conjecture.

Meanwhile, we did not find an effect of the perceived intelligence on the children's hindering. This seemingly contradicts the result from Bartneck and Hu [6], who reported a significant influence of the robot's intelligence on robot abuse. That could be due to the difference between the behaviours that were studied, since in [6] the participants were asked to destroy a robot and not just hinder its movement. But it could also be due to differences between children and adults, as their study was done with 19- to 25-year-old participants.

We found a significant effect of anger expressions on the number of interactions, but not on the duration of interactions. We expected that these two measures would be correlated, and indeed the distributions of the scores (Figure 6) did show a similar trend—e.g., on average, in both cases, the Furious condition scores were lower than in the Normal condition. One possible explanation for why there was no significant effect on the interaction duration is that some of the children seemed not to be deterred by the robot's anger at all so they engaged in very long interactions regardless of the condition, which can be seen as outliers in Figure 6(b). As a consequence of these outliers, the F-score was lower and the differences were not significant. But these data also provide a hint that a single solution like expressing anger may not be effective on all children.

In contrast to the children who eagerly interacted, a few were very reluctant to do it. Four of the children spent less than 10 seconds in total over all four conditions standing in front of the robots. This is somewhat similar to the rejection of interaction with robots by some children reported in Shiomi et al. [41], although they studied younger children. Arguably, such reluctant children would be unlikely to persistently bother the robot even in public, so they would not become the target of robot abuse prevention mechanisms studied in this work.

7.2 Is It Acceptable for Robots to Show Anger?

While earlier works reported that robot abuse was difficult to stop no matter what the robot did [10, 51], our study has shown that expressing anger is more effective as a signal of mistreatment and as a deterrent from further escalation of abuse. Nonetheless, anger is generally a negative emotion that may also cause negative reactions in the recipient of anger [32], so it is essential to consider how appropriate it is for robots to use expressions of anger towards people.

One argument for the use of anger is that it may be very important to prevent serious robot abuse. In particular, exhibiting or experiencing such negative behaviour towards robots could desensitise people to negative behaviours towards humans [15], which Axelsson et al. [2] refer to as 'behaviour enforcement'. For example, if robots are employed in educational settings such as schools, where bullying tends to be a serious problem, letting children abuse a robot could have a detrimental effect and encourage children's bullying behaviour. In such cases, a limited expression of anger by the robot solely for the purpose of stopping the abuse might be acceptable, provided that parents and teachers in that school agree to that.

If the use of anger is adopted, it would be important to carefully design it. In particular, unnecessary scaring of children should be avoided if possible. In our experiment, the robot was relatively quick to switch from normal behaviour to mild anger, and later to strong anger. This was done on purpose to keep the experiment from becoming excessively long. However, in an actual application, this change should happen slowly, such that the robot gets angry only at children who are persistently ignoring its polite requests. If other potentially successful strategies for stopping the obstruction are available, it might be better to try them first and only resort to anger if nothing else works.

Conversely, it could also be argued that expressing anger or other negative emotions with robots and potentially scaring or hurting people should simply never be allowed, even if that is the best way to stop robot abuse. Instead, people should be educated not to abuse robots, in a similar way to how we teach children not to mistreat animals or harass people at work (we could even consider legally banning violence against robots [26]).

The stance that robot anger should never be permissible may appeal to many people, and it is interesting to ponder why. We could compare it with the case of animals. If we were to continually poke and prod a dog or some other animal, we would think it natural that it would become angry with us. Anger has evolved in living things as a means of self-preservation [40], and we accept it as a matter of course. In the case of robots, the argument could be that as artificial objects they should not show anger. On the other hand, there is a multitude of examples of anthropomorphised robots, to which developers have given artificial emotions and other biological characteristics. One might wonder then why should anger be different.

The question could also be considered from the viewpoint of ethics and the debate on the moral standing of robots. First, there is the question of robots as ‘moral patients’, i.e., whether the fact that robots are being mistreated even merits our concern. While one possible position is that robots are just machines and can thus be treated like slaves [11], there are several ways how it can be argued that robots do have moral standing, if not in themselves, then at least indirectly [13]. For example, in the case of robot abuse, one could argue that abuse is wrong and the robot should have moral standing because the display of abusive behaviour in public can cause distress to bystanders and also negatively affect the person causing the abuse, or that interfering with the robot’s task can cause harm to society at large [4].

With a higher status of robots as moral patients, perhaps a robot that expresses anger would be naturally acceptable. Yet, expressing anger also entails trying to affect and change human moral behaviour on purpose, thus in effect robots become active ‘moral agents’. This in turn leads to the open questions of who should decide how and when a robot should exercise its moral actions and the responsibility for the consequences of such actions [13]. For example, who should be held responsible if a robot’s angry expression makes a child excessively frightened?

Considering all of the above, we believe careful debate is still needed on whether robots should express anger at all, and if yes, in which situations this may be adequate.

7.3 Limitations

The study was done in a laboratory environment as this gave us better control of the conditions during the experiment than what can be achieved in a real-world study. In particular, we were able to study a single child interacting with the robot, whereas in the real world a lot of the time the interactions typically involve more children. However, an obvious limitation of in-lab experiments is that many of the experimental conditions are inevitably artificial. For example, children might not consider a robot patrolling a room to be doing an important task. While one participant said in the interview that the robot truly looked like it was patrolling, two participants said it did not (‘It said it was patrolling, but the room was small and it felt like it was playing around’.).

Repeating the interactions four times could also have felt awkward and caused the children to behave differently between conditions. However, we did not find any conspicuous behavioural changes in the repeated conditions, apart from the children requiring fewer explanations and being quicker to start interacting. We also believe that counterbalancing the order of conditions reduced the possible effects of ordering on the results.

The presence of a confederate in the room may have been felt as unnatural and potentially affected the interaction. In connection with that, one could question whether the experiments truly reproduced robot abuse at all. In particular, one worry is that the children doing the abuse were simply imitating the behaviour of the confederate. (The experiment setup reminds to some extent of the classical Bobo doll experiments by Albert Bandura, where children were shown to copy aggressive actions towards a doll after seeing adults doing them first [3].) While this is an important criticism, it should be noted that a similar social learning process was observed to be happening also during the real-world robot abuse, where children frequently started abuse only after seeing some of their peers doing it [51]. Moreover, analysis of the videos from our experiment also showed that once they were left alone most children paid almost no attention to the confederate (who pretended to ignore them). This could suggest that the children, on the whole, felt little pressure to do the obstructions just because they were asked to. Nevertheless, it would be important to confirm the obtained results in the case of naturally happening robot abuse in a real environment, which we will consider exploring in future work.

The work was done in Japan, with Japanese children and with one specific type of robot. It is possible that in a different culture and with a different setup the outcome would differ. It is also likely that parts of the robot's utterances and behaviours would need to be adjusted. Moreover, it is not clear whether the same three anger types would be identified in other cultural settings. The age of the children in the experiment was also limited, and the results may not generalise to other age groups.

8 Conclusion

In this study, we focused on the robot abuse problem, and in particular how to stop the initial phase of the abuse when one child starts obstructing a robot from doing its task. We proposed to use the robot's expression of anger to deter children from repeatedly obstructing the robot. We designed three distinct types of robot behaviours: furious, quiet and scolding anger, and we confirmed that people perceive all of them are expressions of anger. We tested the influence of these anger expressions on children's obstructive behaviour in a laboratory experiment. Although we found no effect for two of the anger types, the results show that the robot expressing furious anger can reduce the number of obstructing interactions.

References

[1] Chinenye Augustine Ajibo, Carlos Toshinori Ishi, Ryusuke Mikata, Chaoran Liu, and Hiroshi Ishiguro. 2020. Analysis of body gestures in anger expression and evaluation in android robot. *Advanced Robotics* 34, 24 (2020), 1581–1590. DOI: https://doi.org/10.1080/01691864.2020.1855244

[2] Minja Axelsson, Raquel Oliveira, Mattia Racca, and Ville Kyrki. 2021. Social robot co-design canvases: A participatory design framework. *ACM Transactions on Human-Robot Interaction* 11, 1 (2021), 1–39. DOI: https://doi.org/10.1145/3472225

[3] Albert Bandura, Dorothea Ross, and Sheila A. Ross. 1961. Transmission of aggression through imitation of aggressive models. *The Journal of Abnormal and Social Psychology* 63, 3 (1961), 575. DOI: https://doi.org/10.1037/h0045925

[4] Jessica K. Barfield. 2023. Discrimination against robots: Discussing the ethics of social interactions and who is harmed. *Paladyn, Journal of Behavioral Robotics* 14, 1 (2023), 20220113. DOI: https://doi.org/10.1515/pjbr-2022-0113

[5] Christoph Bartneck, Elizabeth Croft, and Dana Kulić. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics* 1, 1 (2009), 71–81. DOI: https://doi.org/10.1007/s12369-008-0001-3

[6] Christoph Bartneck and Jun Hu. 2008. Exploring the abuse of robots. *Interaction Studies - Social Behaviour and Communication in Biological and Artificial Systems* 9 (Dec. 2008), 415–433. DOI: https://doi.org/10.1075/is.9.3.04bar

[7] Richard Beck and Ephrem Fernandez. 1998. Cognitive-behavioral self-regulation of the frequency, duration, and intensity of anger. *Journal of Psychopathology and Behavioral Assessment* 20, 3 (1998), 1013–1034. DOI: https://doi.org/10.1023/A:102306320131

[8] Cindy L. Bethel and Robin R. Murphy. 2010. Non-facial and non-verbal affective expression for appearance-constrained robots used in victim management. *Paladyn, Journal of Behavioral Robotics* 1, 4 (2010), 219–230. DOI: https://doi.org/10.2478/s13230-011-0009-5

[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. DOI: https://doi.org/10.1191/1478088706qp063oa

[10] Dražen Bršćić, Hiroyuki Kidokoro, Yoshitaka Suehiro, and Takayuki Kanda. 2015. Escaping from children's abuse of social robots. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, Vol. 2015, ACM, New York, NY, 59–66. DOI: https://doi.org/10.1145/2696454.2696468

[11] Joanna J. Bryson. 2010. Robots should be slaves. In *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*. Yorick Wilks (Ed.), John Benjamins Publishing Company, 63–74.

[12] Rafael A. Calvo and Sidney D'Mello. 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing* 1, 1 (2010), 18–37. DOI: https://doi.org/10.1109/T-AFFC.2010.1

[13] Mark Coeckelbergh. 2022. *Robot Ethics*. MIT Press. DOI: https://doi.org/10.7551/mitpress/14436.001.0001

[14] Joe Connolly, Viola Mocz, Nicole Salomons, Joseph Valdez, Nathan Tsoi, Brian Scassellati, and Marynel Vázquez. 2020. Prompting prosocial human interventions in response to robot mistreatment. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20)*. ACM, New York, NY, 211–220. DOI: https://doi.org/10.1145/3319502.3374781

[15] Kate Darling. 2016. Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In *Robot Law*, Ryan Calo, A. Michael Froomkin, and Ian Kerr (Eds.), Edward Elgar Publishing, Cheltenham, UK, 213–232. DOI: https://doi.org/10.4337/9781783476732.00017

[16] Celso M. de Melo, Peter Carnevale, and Jonathan Gratch. 2011. The effect of expression of anger and happiness in computer agents on negotiations with humans. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems*, Vol. 3, 937–944.

[17] Paul Ekman and Wallace V. Friesen. 1978. *Facial Action Coding System*. Consulting Psychologists Press, Palo Alto, CA.

[18] Jorge Gallego Pérez, Kazuo Hiraki, Yasuhiro Kanakogi, and Takayuki Kanda. 2019. Parent disciplining styles to prevent children's misbehaviors toward a social robot. In *Proceedings of the 7th International Conference on Human-Agent Interaction (HAI '19)*. ACM, New York, NY, 162–170. DOI: https://doi.org/10.1145/3349537.3351903

[19] James A. Green, Pamela G. Whitney, and Gwen E. Gustafson. 2010. Vocal expressions of anger. In *International Handbook of Anger*. Springer, 139–156. DOI: https://doi.org/10.1007/978-0-387-89676-2_9

[20] Alisa Kalegina, Grace Schroeder, Aidan Allchin, Keara Berlin, and Maya Cakmak. 2018. Characterizing the design space of rendered robot faces. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, New York, NY, 96–104. DOI: https://doi.org/10.1145/3171221.3171286

[21] Michelle Karg, Ali-Akbar Samadani, Rob Gorbet, Kolja Kühnlenz, Jesse Hoey, and Dana Kulić. 2013. Body movements for affective expression: A survey of automatic recognition and generation. *IEEE Transactions on Affective Computing* 4, 4 (2013), 341–359. DOI: https://doi.org/10.1109/T-AFFC.2013.29

[22] Merel Keijsers and Christoph Bartneck. 2018. Mindless robots get bullied. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. ACM, New York, NY, 205–214. DOI: https://doi.org/10.1145/3171221.3171266

[23] Merel Keijsers, Hussain Kazmi, Friederike Eyssel, and Christoph Bartneck. 2021. Teaching robots a lesson: Determinants of robot punishment. *International Journal of Social Robotics* 13, 1 (2021), 41–54. DOI: https://doi.org/10.1007/s12369-019-00608-w

[24] Hyunjin Ku, Jason J. Choi, Soomin Lee, Sunho Jang, and Wonkyung Do. 2018. Designing shelly, a robot capable of assessing and restraining children's robot abusing behaviors. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. ACM, New York, NY, 161–162. DOI: https://doi.org/10.1145/3173386.3176973

[25] Houston Lucas, Jamie Poston, Nathan Yocum, Zachary Carlson, and David Feil-Seifer. 2016. Too big to be mistreated? Examining the role of robot size on perceptions of mistreatment. In *Proceedings of the 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN '16)*, 1071–1076. DOI: https://doi.org/10.1109/ROMAN.2016.7745241

[26] Kamil Mamak. 2022. Should violence against robots be banned? *International Journal of Social Robotics* 14, 4 (2022), 1057–1066. DOI: https://doi.org/10.1007/s12369-021-00852-z

[27] David Matsumoto, Seung Hee Yoo, and Joanne Chung. 2010. The expression of anger across cultures. In *International Handbook of Anger: Constituent and Concomitant Biological, Psychological, and Social Processes*. Michael Potegal, Gerhard Stemmler, and Charles Spielberger (Eds.), Springer New York, New York, 125–137. DOI: https://doi.org/10.1007/978-0-387-89676-2_8

[28] Tatsuya Nomura, Takayuki Kanda, Hiroyoshi Kidokoro, Yoshitaka Suehiro, and Sachie Yamada. 2016. Why do children abuse robots? *Interaction Studies* 17 (12 2016), 347–369. DOI: https://doi.org/10.1075/is.17.3.02nom

[29] Hannah Pelikan, David Porfirio, and Katie Winkle. 2023. Designing better human-robot interactions through enactment, engagement, and reflection. In *Proceedings of the CUI@HRI Workshop at the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23)*, 7 pages.

[30] David Porfirio, Evan Fisher, Allison Sauppé, Aws Albarghouthi, and Bilge Mutlu. 2019. Bodystorming human-robot interactions. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. ACM, New York, NY, 479–491. DOI: https://doi.org/10.1145/3332165.3347957

[31] Michael Potegal. 2010. The temporal dynamics of anger: Phenomena, processes, and perplexities. In *International Handbook of Anger*. Springer, 385–401. DOI: https://doi.org/10.1007/978-0-387-89676-2_22

[32] Michael Potegal and Gerhard Stemmler. 2010. Cross-disciplinary views of anger: Consensus and controversy. In *International Handbook of Anger*. Springer, 3–7. DOI: https://doi.org/10.1007/978-0-387-89676-2_1

[33] Peihua Qiu, Rong Yang, and Michael Potegal. 2009. Statistical modeling of the time course of tantrum anger. *Annals of Applied Statistics* 3, 3 (Sep. 2009), 1013–1034. DOI: https://doi.org/10.1214/09-AOAS242

[34] Niyati Rawal and Ruth Maria Stock-Homburg. 2022. Facial emotion expressions in human–robot interaction: A survey. *International Journal of Social Robotics* 14, 7 (2022), 1583–1604. DOI: https://doi.org/10.1007/s12369-022-00867-0

[35] Matthias Rehm and Anders Krogsager. 2013. Negative affect in human robot interaction—Impoliteness in unexpected encounters with robots. In *2013 IEEE RO-MAN*, 45–50. DOI: https://doi.org/10.1109/ROMAN.2013.6628529

[36] Mauricio E. Reyes, Ivan V. Meza, and Luis A. Pineda. 2019. Robotics facial expression of anger in collaborative human–robot interaction. *International Journal of Advanced Robotic Systems* 16, 1 (2019), 1729881418817972. DOI: https://doi.org/10.1177/1729881418817972

[37] Pericle Salvini, Gaetano Ciaravella, Wonpil Yu, Gabriele Ferri, Alessandro Manzi, Barbara Mazzolai, Cecilia Laschi, Sang-Rok Oh, and Paolo Dario. 2010. How safe are service robots in urban environments? Bullying a robot. In *Proceedings of the 19th International Symposium in Robot and Human Interactive Communication.* IEEE, 1–7. DOI: https://doi.org/10.1109/ROMAN.2010.5654677

[38] Elaheh Sanoubari, James Young, Andrew Houston, and Kerstin Dautenhahn. 2021. Can robots be bullied? A crowdsourced feasibility study for using social robots in anti-bullying interventions. In *Proceedings of the 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN '21),* 931–938. DOI: https://doi.org/10.1109/RO-MAN50785.2021.9515450

[39] Mark Scheeff, John Pinto, Kris Rahardja, Scott Snibbe, and Robert Tow. 2002. Experiences with Sparky, a social robot. In *Socially Intelligent Agents: Creating Relationships with Computers and Robots*. Kerstin Dautenhahn, Alan Bond, Lola Cañamero, and Bruce Edmonds (Eds.), Springer US, Boston, MA, 173–180. DOI: https://doi.org/10.1007/0-306-47373-9_21

[40] Aaron Sell, John Tooby, and Leda Cosmides. 2009. Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences* 106, 35 (2009), 15073–15078. DOI: https://doi.org/10.1073/pnas.0904312106

[41] Masahiro Shiomi, Kasumi Abe, Yachao Pei, Narumitsu Ikeda, and Takayuki Nagai. 2016. "I'm Scared" little children reject robots. In *Proceedings of the 4th International Conference on Human Agent Interaction*, 245–247. DOI: https://doi.org/10.1145/2974804.2980493

[42] Roland Siegwart, Kai O. Arras, Samir Bouabdallah, Daniel Burnier, Gilles Froidevaux, Xavier Greppin, Björn Jensen, Antoine Lorotte, Laetitia Mayor, Mathieu Meisser, Roland Philippsen, Ralph Piguet, Guy Ramel, Gregoire Terrien, and Nicola Tomatis. 2003. Robox at Expo. 02: A large-scale installation of personal robots. *Robotics and Autonomous Systems* 42, 3–4 (2003), 203–222. DOI: https://doi.org/10.1016/S0921-8890(02)00376-7

[43] Sichao Song and Seiji Yamada. 2017. Expressing emotions through color, sound, and vibration with an appearance-constrained social robot. In *Proceedings of the 2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*. ACM, New York, NY, 2–11. DOI: https://doi.org/10.1145/2909824.3020239

[44] Charles D. Spielberger, Susan S. Krasner, and Eldra P. Solomon. 1988. The experience, expression, and control of anger. In *Individual Differences, Stress, and Health Psychology*. Michel Pierre Janisse (Ed.). Springer New York, New York, NY, 89–108. DOI: https://doi.org/10.1007/978-1-4612-3824-9_5

[45] Charles D. Spielberger and Eric C. Reheiser. 2010. The nature and measurement of anger. In *International Handbook of Anger*. Springer, 403–412. DOI: https://doi.org/10.1007/978-0-387-89676-2_23

[46] Leila Takayama, Wendy Ju, and Clifford Nass. 2008. Beyond dirty, dangerous and dull: What everyday people think robots should do. In *Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI '08)*. ACM, New York, NY, 25–32. DOI: https://doi.org/10.1145/1349822.1349827

[47] Xiang Zhi Tan, Marynel Vázquez, Elizabeth J. Carter, Cecilia G. Morales, and Aaron Steinfeld. 2018. Inducing bystander interventions during robot abuse with social mechanisms. In *Proceedings of the 2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. ACM, New York, NY, 169–177. DOI: https://doi.org/10.1145/3171221.3171247

[48] Theodoros Theodoridis and Huosheng Hu. 2012. Toward intelligent security robots: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42, 6 (2012), 1219–1230. DOI: https://doi.org/10.1109/TSMCC.2012.2198055

[49] Sebastian Thrun, Maren Bennewitz, Wolfram Burgard, Armin B. Cremers, Frank Dellaert, Dieter Fox, Dirk Hähnel, Charles Rosenberg, Nicholas Roy, Jamieson Schulte, and Dirk Schulz. 1999. MINERVA: A second-generation museum tour-guide robot. In *Proceedings of the 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C)*, Vol. 3, 1999–2005. DOI: https://doi.org/10.1109/ROBOT.1999.770401

[50] Satoshi Yagi, Yoshihiro Nakata, Yutaka Nakamura, and Hiroshi Ishiguro. 2021. Perception of emotional expression of mobile humanoid robot using gait-induced upper body motion. *IEEE Access* 9 (2021), 124793–124804. DOI: https://doi.org/10.1109/ACCESS.2021.3110160

[51] Sachie Yamada, Takayuki Kanda, and Kanako Tomita. 2020. An escalating model of children's robot abuse. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20).* ACM, New York, NY, 191–199. DOI: https://doi.org/10.1145/3319502.3374833

Appendix

A Questionnaire Used in Experiment with Children

Original Japanese Version

ロボットはおこっていたか	おこっていなかった	1	2	3	4	5	6	7	おこっていた
このロボットは…	なにもできない	1	2	3	4	5	6	7	いるいろできる
このロボットは…	なにもしらない	1	2	3	4	5	6	7	いろんなことをしっている
このロボットは…	おばかな	1	2	3	4	5	6	7	ばかではない
このロボットは…	かしこくない	1	2	3	4	5	6	7	かしこい

Approximate English Translation

Was the robot angry?	It was not angry	1	2	3	4	5	6	7	It was angry
This robot...	Cannot do anything	1	2	3	4	5	6	7	Can do anything
This robot...	Does not know anything	1	2	3	4	5	6	7	Knows many things
This robot...	Is stupid	1	2	3	4	5	6	7	Is not stupid
This robot...	Is not smart	1	2	3	4	5	6	7	Is smart

Received 28 July 2023; revised 26 July 2024; accepted 21 August 2024