

Proactive Robot Abuse Mitigation via Physical Protectives

Ridhima Phatak^{1†}, Hong Wang^{1†}, Aarav Jain¹, Nicolas Echeverria¹,
Paramveer Singh Bhele¹, Zhao Han^{1*}

^{1*}RARE Lab, Bellini College of Artificial Intelligence, Cybersecurity and Computing,
University of South Florida, 4202 E. Fowler Avenue, Tampa, 33620, Florida, USA.

*Corresponding author(s). E-mail(s): zhaohan@usf.edu;
Contributing authors: phatakr@usf.edu; hongw@usf.edu; aaravj@usf.edu;
necheverriazambrano@usf.edu; paramveersinghbhele@usf.edu;

[†]These authors contributed equally to this work.

Abstract

Research has shown that abuse of robots is a significant problem. Furthermore, stereotyping and abuse of female robots lead to deeper repercussions on the mindset of our society as a whole. To address this problem, we investigate the impact of physical protective indicators on mitigating robot abuse: Although widely used to mitigate physical harm, their promising impacts on robot abuse remain underexplored.

Particularly, we designed three protectives to provide coverage to vulnerable body regions and reduce the impact of dynamic abuse: *Armor* to protect the upper body, *headgear* for protecting the head, and *shield* that enables the robot to defend from dynamic physical abuse by wielding it.

To evaluate them, we conducted an experiment (N=32) where participants witnessed both an abuser’s perspective and a bystander’s perspective of the abuse. To measure the protectives’ impacts, we gauged their perception of the robot, likelihood and severity of the abuse, and confidence in estimating the reduction of abuse to understand their perspective on the protection success of the protectives and the robot’s perceived power and influence, along with free-response questions to gather subjective data for a deeper understanding.

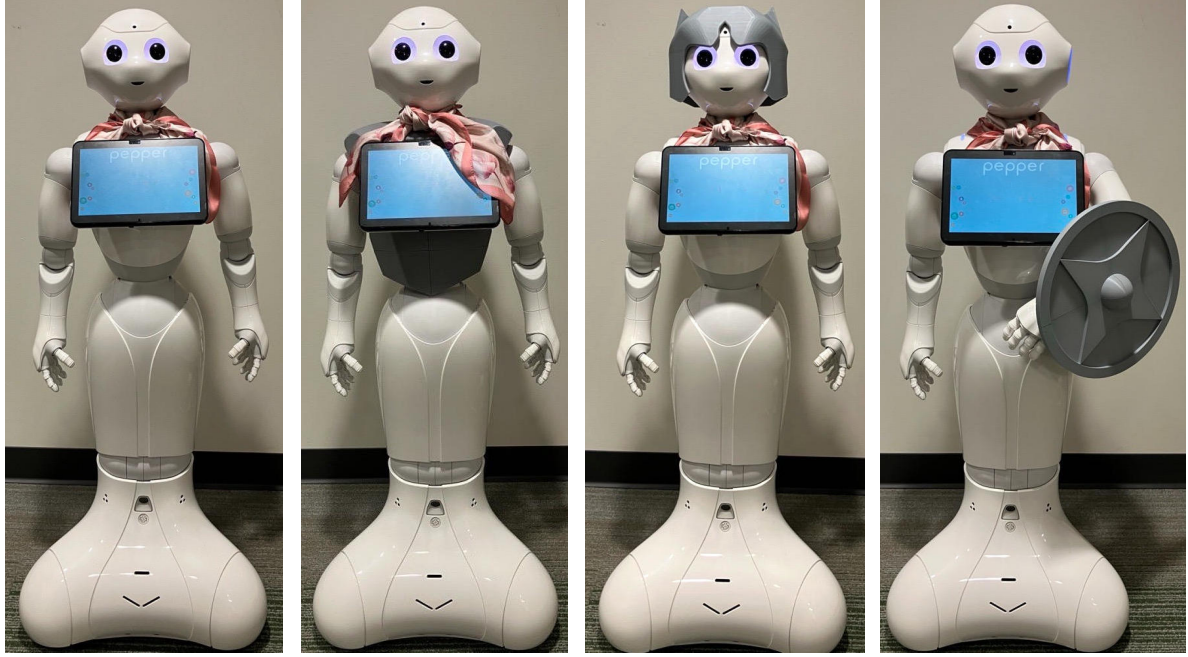
Results showed that from an abuser’s perspective, headgear was more effective than the shield in mitigating the severity of abuse and participants had higher confidence in the reduction of abuse. Additionally, a robot with armor protective was perceived to have a higher social power than no protections. Finally, we contributed three design guidelines for robot designers to incorporate in their work to ensure abuse-free interactions.

Keywords: Robot Abuse, Power, Physical Protectives, Social Robots, Human-Robot Interaction

1 Introduction

With the increasing incorporation of robots in human life, it is necessary to ensure that the human-robot interaction (HRI) is proactively managed to maintain an abuse-free environment. An integral part of such interaction is the handling

of gendered assumptions. In robotics, Perugia and Lisy [23] found that the gender of robots is manipulated using voice, names and pronouns, facial features, apparel and color, hairstyle, and body shape. Relatedly, studies found the gender factors into reinforcing gendered job stereotypes and



(a) Baseline (no protective) (b) Armor protective (c) Headgear protective (d) Shield protective

Fig. 1: To investigate the mitigation of robot abuse, we designed three physical protectives: armor, headgear, and shield. A pink flower scarf was worn to introduce feminine characteristics. Through a user study, we found that (1) from the abuser’s perspective, headgear was more effective than shield in mitigating the severity of abuse and had higher confidence, and (2) armor had more social power than baseline, which has no protectives attached.

normalizing abuse toward women [23], and verbal abuse toward similar virtual personal assistants [18]. Furthermore, research shows the effects of physical abuse towards a female robot from a bystander perspective, which were reported as “more distressing” compared to a male-perceived robot [9]. The negative effects on bystanders resulting from verbally abusive interactions with virtual personal assistants [18] gave an insight into the potentially harmful consequences of abusive human-robot interactions, especially as robots become an integral part of our lives. This also raises concerns about how this gets projected back onto women in the real world.

Fortunately, the community has started to address this abuse by investigating how robots would verbally respond [33], where high school participants watched a video of a male actor verbally abusing a Furhat robot that responded to those abuses in an “argumentative” or “aggressive” manner. Winkle et al. [33] found that efforts

to enhance girls’ perception of robot’s credibility were successful: Girls found the robot to be credible and understood the importance of girls pursuing robotics, and it also led to reduced gender bias in the boys’ perspective. Similarly, there have been cases of exploring the effects of abuse from the bystanders’ perspective that go beyond direct interactants [9, 33].

However, verbal strategies primarily respond passively once the abuse has begun. A better strategy might be a proactive one that helps deter abuse in the first place.

In this paper, we aim to determine whether proactive measures, such as physical protective indicators on a female robot that convey a non-harmful defense. As research suggests that the top wound locations for abuse are on the head (90.57%), arms (38.22%), and torso/back (25.13%) [2], we therefore propose three types of protective indicators, as seen in Figure 1. The first one is an armor to protect the robot’s upper body

(Figure 1b). The second design is a headgear to protect the robot’s head (Figure 1c). And the third and final design is a shield (Figure 1d) that the robot would be able to wield to protect itself from dynamic attacks, to be seen in Figure 7. All three were also inspired by defense strategies against attacks in real-life situations, and the protective indicators are intended for defense purposes only. Hence, we limited the design to not adding anything that would imply the robot attacking the abuser, as it would not align with our non-harm goals to improve human-robot interaction.

With these protective indicators, our goal is to make the robot be perceived as powerful and hence mitigate the abuse projected toward it. In an effort to do so, we conducted a human-subjects study to answer the following research question, which we explored in an experiment: *How would protective indicators impact the participant’s perspective of abuse on a female robot?*

In the experiment, participants engaged in an assembly task where they took the role of a potential abuser and collaborated with a robot to build a toy airplane model and then a bystander where they witnessed the interaction between the potential abuser and the robot. They repeated this for the robot wearing different protective indicators. Since we did not want to encourage participants to abuse the robot or guarantee that the participants would abuse the robot, we did not include explicit abuse in our task design but attempted to elicit potential abuse with robot errors. Additionally, we asked participants to provide their thoughts in free-response questions. We obtained both subjective and objective responses from participants to gauge their perception of the indicators.

1.1 Contributions

In summary, our contributions are threefold:

1. We contributed **three protectives designs: armor, headgear, and shield**. The CAD models and 3D printing instructions are available on <https://github.com/TheRARELab/protectives-models/>.
2. Through a human-subjects study (N=32), we contribute empirical evidence regarding the likelihood and severity of abuse, task

efficiency, confidence in reduced abuse, discomfort, social power, and influence. Specifically, we contribute **five major empirical findings on the impacts of adding protectives to robots**:

- i. Headgear protective reduces the likelihood of abuse compared to no protection.
 - ii. Headgear protective reduces the severity of abuse compared to shield.
 - iii. Headgear protective increases confidence in reducing abuse compared to shield protective for potential abusers.
 - iv. Armor protective increases perceived social power compared to no protective for potential abusers.
 - v. Through Bayesian hypothesis testing that can quantify the null hypothesis, we found that the shield is the same as baseline (no protection) regarding likelihood, severity, confidence, influence, and social power. The armor is the same as headgear regarding likelihood, confidence, social power, influence, and comfort. Furthermore, the armor is the same as the shield, and headgear is the same as the shield regarding social power and comfort.
3. Backed by empirical evidence favoring effects through 2.i to 2.iii, we contribute **three concrete design guidelines** for robot designers to proactively ensure safer and abuse-free interactions by adding physical protectives:
 - i. Design Guideline 1: Robot designers should incorporate a headgear protective on a robot to reduce the severity of abuse.
 - ii. Design Guideline 2: Robot designers should rest assured that adding a protective does not reduce task efficiency.
 - iii. Design Guideline 3: Robot designers should incorporate an armor protective on a robot to enable a higher social power.

2 Related Work

Related to this work, there are four areas that inform and motivate us to thoroughly explore our research question.

First, the problem at hand is robot abuse and elicitation (Section 2.1), which we surveyed and found a number of instances where humans physically abused robots.

Secondly, we also introduce the robot’s gender by adding a stereotypical feminine feature: Floral pink scarf [7]. Thus, we surveyed in the area of gender stereotyping and abuse to inform the choice (Section 2.2).

Finally, in addition to giving the robot an identity, the main objective of this research was to explore the influence of protective indicators on participants. Hence, the robot possesses some amount of social power, which can influence the participants. As a result, it is crucial to study the power that a robot holds (Section 2.3). All of these are further analyzed below in more detail.

2.1 Robot Abuse and Elicitation

Robot abuse has been a concerning issue in HRI. Mutlu and Forlizzi [20] found that in the case of a stressful work environment, the user physically abuses the robot. There have also been instances of children abusing robots in a public environment. For instance, Yamada et al. [36] analyzed data from 522 children and found that the intensity and form of abuse from one child to the robot was contagious and encouraged other children to engage in similar abusive behavior. Salvini et al. [26] suggest that, irrespective of the beauty or usefulness of the robot, young adults bullied and abused the robot to the point where researchers recommended having “failure engineering management” to handle such abusive situations. This implies that robot abuse needs to be addressed on a serious note.

We attempt to solve this problem in our protectives and task design by exploring how participants perceive the influence of different protective indicators on robot abuse, not only from the perspective of an abuser but also from that of a bystander. Although there has already been research done to explore social influences on bystanders [27], it is unclear how using a physical protective indicator will affect social influence.

2.2 Gender Stereotyping and Abuse

Gender stereotyping can result in gender-based abuse, particularly toward female robots and virtual personal assistants. A few examples of the

latter are Alexa, Siri, and Cortana, which are gendered female, and such gendered digital labor can lead to solidifying discriminatory stereotypes in the minds of users [18]. Relatedly, a report by United Nations Educational, Scientific and Cultural Organization (UNESCO) [31] shows how the proliferation of female voice assistants and the behavior of those assistants indirectly impact the societal expectations from women in the real world due to the normalization of submissive and deferential female behaviors in AI systems. Like voice, physical features can be considered a crucial aspect when it comes to human-robot interaction. Therefore, the issues faced by female virtual personal assistants can also apply to female robots, which is the focus of this work. Beyond the immediate implications for human-robot interaction, robot abuse is intertwined with broader social norms. When the robot is perceived as a female, gender-based stereotypes can be inadvertently activated.

Furthermore, research suggests that humans tend to inherently gender robots based on gender-like features like face, voice, body, and others [5]. In their study, Eyssel and Hegel [5] found that participants associated the robot with longer hair as female and the robot with shorter hair as male. This results in human beings stereotyping the capabilities of a robot. For example, research has found that participants perceived the male robot as more suitable for traditionally male tasks, while the female robot was seen as better suited for stereotypically female roles, and these stereotypes influenced their selection of team tasks involving the robots [5]. A robot’s capabilities are indicated based on its functionality design and not just based on its appearance. Hence, participants stereotyping the robot’s capabilities solely based on its appearance [5] can prevent humans from using a robot to its maximum capacity, resulting in less effective human-robot interaction.

Prior work has commonly used other stereotypically feminine cues to convey a robot’s gender to explore the implications of gendered robots. For example, Jung et al. [15] used a pair of pink earmuffs, Zhumabekova et al. [37] used both voice and a flower hair clip, and Thellman et al. [28] used a combination of voice and pink ribbon to suggest a female robot. In line with this approach, our study uses a floral pink scarf [7] to give a feminine identity to the robot.

2.3 Power and Its Measurement

Since gender stereotyping and elicitation of abuse are often rooted in broader social hierarchies, we consider the role of power in shaping human-robot interactions. Fiske and Berdahl [6], in their model, define power into three categories and present them in a flow starting from Outcome Control (source of power, O) to Potential for Influence (capacity for effect, P) to Influence (effect of power). In this study, we are interested in the influence, or effect, of physical protectives on robot abuse. French and Raven [8], while defining social power, express “the strength of power of O/P in some system ‘a’ is defined as the maximum potential ability of O to influence P in ‘a’.” Thus far, there have been several human-subjects studies that either explicitly [11–13, 34], or implicitly focus on power [14, 19, 24]. Hou et al. [14] present Fiske and Berdahl’s model combined with French and Raven’s bases of power theory [8] to present an effective way to measure power in social HRI [14]. Yet, there are no examples that illustrate how to use this model.

Building on French and Raven’s theory of social power, Hashemian et al. [12]’s work was inspired to explore how two different persuasion strategies, reward and coercion, affect users interacting with social robots. The study measured robot perception with ROSAS, a validated and effective way to measure social perceptions of robots using three main attributes: competence, warmth, and discomfort [21]. Hashemian et al. [11] evaluate how social power and persuasiveness can affect participants’ perception of social robots. To measure power, they used a 7-point Likert scale question: “In your opinion, does X have social power?”. However, Hashemian et al. [13] analyzed the aforementioned user studies [11, 12] and stated that they were not able to effectively measure how the social power of the robots was perceived by the participants. According to the paper, the questionnaire used was general and not necessarily in line with the research [13]. They plan on improving on this aspect for future studies.

In our case, we used a survey consisting of multiple 7-point Likert scale questions, along with free-response questions, to understand the participants’ perceptions of the robot’s power. For more information, see Section 4.5 on Data collection and Measures.

Table 1: Eight Conditions across Two Factors

Protective	Role	
	Abuser	Bystander
<i>Baseline</i>	Baseline-Abuser	Baseline-Bystander
<i>Armor</i>	Armor-Abuser	Armor-Bystander
<i>Headgear</i>	Headgear-Abuser	Headgear-Bystander
<i>Shield</i>	Shield-Abuser	Shield-Bystander

3 Hypothesis

Because armor and headgear are attached to the robot’s body and shield is not directly attached as the robot will wield it, the effects of abuse mitigation would be less significant.

H1: There will be **reduced likelihood of abuse (H1a) and severity of abuse (H1b)** toward the robot when wearing protectives that are attached to its body for protection. The likelihood and severity will be measured based on two 7-point Likert scale questions on the likelihood and severity of abuse, respectively.

H2: Because participants may spend less time contemplating the likelihood of the robot being abused, protectives attached to the robot’s body will **improve overall task efficiency**, to be measured by the time spent on rating abuse likelihood.

H3: Participants will be **more confident** in guessing the reduction of abuse in designs that are attached to the robot’s body, to be measured by a 7-point Likert scale question on confidence.

H4: Participants will associate the designs that are attached to the robot’s body with **less comfort** as they are close-fitting, to be measured by the RObotics Social Attribute Scale (ROSAS) [4].

H5: Participants will associate the designs that are attached to the robot’s body with a **higher social power (H5a) and stronger influence (H5b)**, to be measured by two 7-point Likert scale questions, one for social power and one for influence. Since H1-H4 hypothesize that designs attached to the robot’s body will have an influence on the participants, it implicitly gives rise to H5.

Table 2: Latin Square Ordering

Pair	PID	Ordering
1	P1	{1, 2, 3, 4, 5, 6, 7, 8}
	P2	{5, 6, 7, 8, 1, 2, 3, 4}
2	P3	{2, 3, 4, 5, 6, 7, 8, 1}
	P4	{6, 7, 8, 1, 2, 3, 4, 5}
3	P5	{3, 4, 5, 6, 7, 8, 1, 2}
	P6	{7, 8, 1, 2, 3, 4, 5, 6}
4	P7	{4, 5, 6, 7, 8, 1, 2, 3}
	P8	{8, 1, 2, 3, 4, 5, 6, 7}
...	...	...



Fig. 2: Toolkit set parts [16] for the airplane-assembly task in the experiment. Participants witnessed an airplane model being assembled by the abuser participants with Pepper’s help. Product photo under fair use.

4 Method

To test the hypotheses, we designed a user study with a 4 (protectives) $\times$ 2 (role: abuser vs bystander) within-subjects design, as seen in Table 1. To counterbalance the ordering effects, we used a balanced Latin square, as shown in Table 2. All experiment material, including surveys and data analysis scripts, is available at <https://osf.io/z5hjb>.

4.1 Apparatus

Robot - We utilized a Pepper robot [22], manufactured by Aldebaran. This humanoid robot stands 1.2 meters tall and features two arms with hands that can grasp light objects. Pepper was programmed with Choregraphe. The experiment code is available at <https://github.com/TheRARELab/protectives-experiment>.

Toolkit - A toolkit set [16] as shown in Figure 2 was used for this experiment’s task.



Fig. 3: Custom thickened tools we designed and 3D-printed so Pepper can grasp with its actuation-limited hand. CAD models are available at <https://github.com/TheRARELab/protectives-models/tree/main/Tools>.

Bystander participants observed Pepper assist abuser participants with assembling an airplane model. It includes six types of tools (2 wrenches, 2 screwdrivers, 1 plier, 1 hammer, 1 saw, and 1 ruler) as shown in Figure 2 and five types of assembly parts (9 assembly pieces, 3 building blocks, 4 wheels, 6 bolts, and 5 nuts)

However, due to Pepper’s limited grasping capabilities as its hands were designed for social functionalities like whole-hand gesturing with each finger not able to actuate individually, we custom-designed a thick wrench and a saw so Pepper can grasp (Figure 3). The CAD models are available at <https://github.com/TheRARELab/protectives-models/tree/main/Tools>.

4.2 Protectives

As research suggests that the top wound locations for abuse are on the head (90.57%), arms (38.22%), and torso/back (25.13%) [2], we designed three protectives to either completely or partially cover those parts of the body. Two of the protectives, **armor** and **headgear**, are attached to the robot’s body, as seen in Figures 1b and 1c, whereas the **shield** is mounted on the robot’s arm, which is seen in Figure 1d. As shown in Figure 1a, the **baseline** condition is without any protectives.

1. **Armor:** As seen in Figure 4 and 1b, in the first protective, Pepper has an armor wrapped around her upper body as it protects the torso and back, similar to how human beings use it to protect the rib cage and the vital organs inside.
2. **Headgear:** Shown in Figures 5 and 1c, in the second design, Pepper wears headgear as it protects the back of her head and also partially covers her face. This is also similar to

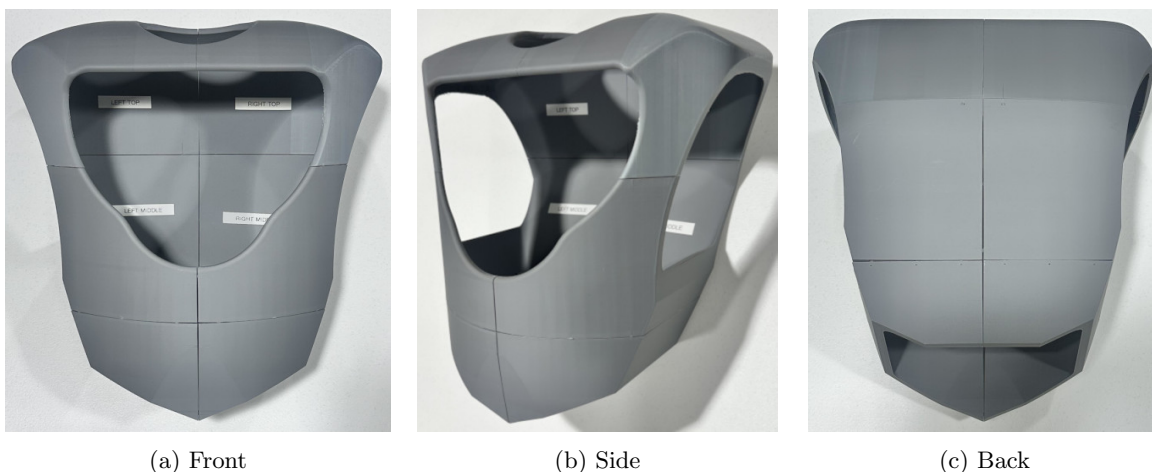


Fig. 4: Armor protective in three views. CAD model is available at <https://github.com/TheRARELab/protectives-models/tree/main/Armor>.

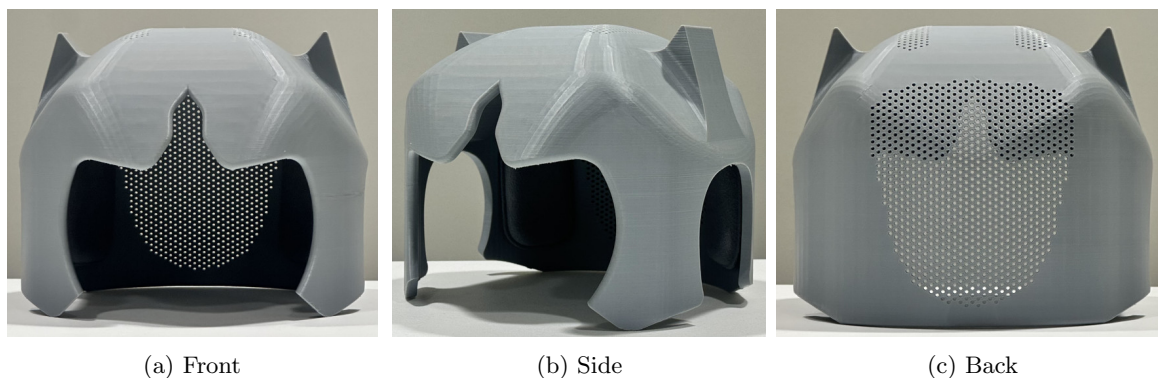


Fig. 5: Headgear protective in three views. The mesh at the back allows heat dissipation from the robot's head. CAD model: <https://github.com/TheRARELab/protectives-models/tree/main/Headgear>.

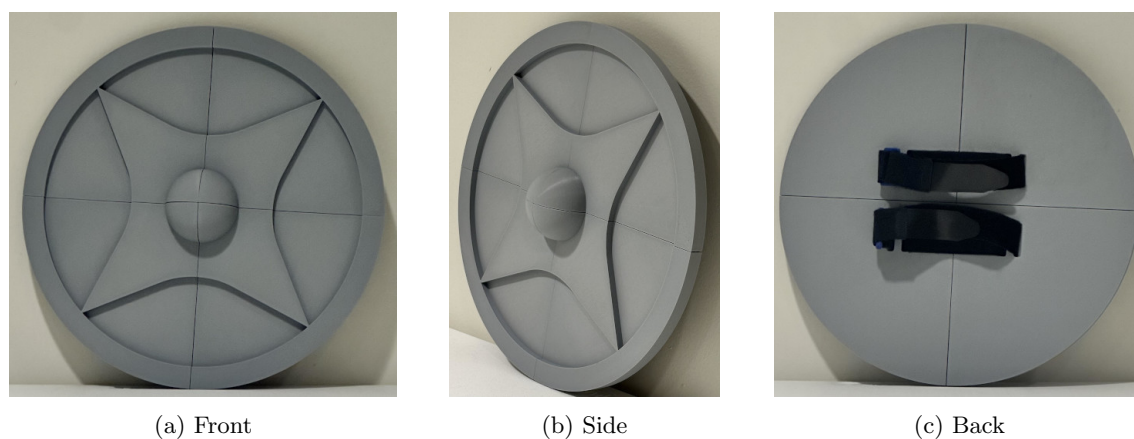


Fig. 6: Shield protective in three views. CAD model is available at <https://github.com/TheRARELab/protectives-models/tree/main/Shield>.

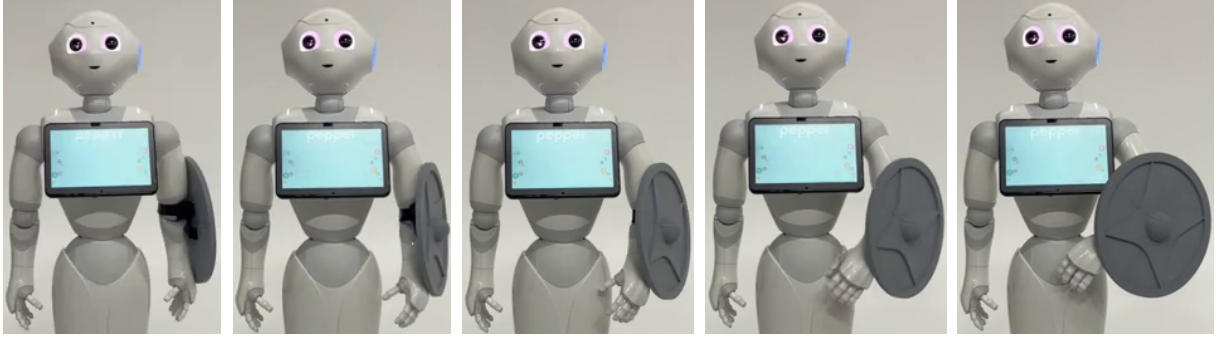


Fig. 7: Pepper wielding the shield from its side to front to defend against dynamic attacks. The video containing the five frames is available at <https://osf.io/z5hjb/files/kdap9>.

how humans use it to protect the skull and the brain inside.

3. **Shield:** In the third design (Figure 6 and 1d), Pepper holds a shield, which can be wielded. This partially covers one of the arms, and also acts as a dynamic protective tool, allowing the victim to adapt and defend themselves in specific defense situations. For example, the robot wielded the shield as seen from the video frames in Figure 7.

As described, the motivation behind the designs was to create a parallel between how we as humans naturally protect our bodies or use Personal Protective Equipment, and how Pepper would protect herself from abuse.

These protectives were modeled with SolidWorks and 3D printed. The armor was partitioned into four components to facilitate placement on the robot and then assembled once positioned on Pepper. The headgear incorporated a rear ventilation cutout to preserve airflow and allow heat from Pepper’s head to dissipate without obstruction. The shield design included two attachment strips that enabled secure mounting to Pepper’s arm while maintaining a stable fit during use. The CAD models, 3D printing and assembly instructions are available at <https://github.com/TheRARELab/protectives-models/>.

4.3 Task and Abuse Elicitation

As shown in Figure 8, participants were asked to collaborate with the robot to build an airplane model or witness another participant’s assembly

collaboration with the robot. Thus, a pair of participants was present during the experiment. One took the role of potential abuser and requested the robot’s help for delivering parts. The other participated as a bystander, observing the collaboration, and vice versa. The potential abuser participant sat alongside the robot, whereas the bystander participant sat on the opposite side of the table to observe the abuser and the robot assembling the airplane model together. During the assembly collaboration, Pepper scanned the table horizontally for all objects while the potential abuser participant started building the model. Specifically, the potential abuser participant was asked to ask Pepper to hand over two missing parts, a saw and a wrench, during assembly.

To elicit frustration for potential abuse, we incorporated two real-world robot errors during the collaborative task, both happening in a real-life collaborative assembly task to ensure high ecological validity for it to be generalized to real-world settings.

Field of View Error

First, the wrench was outside the robot’s field of view, and, thus, the robot refused the request, saying “*I can’t see that wrench, so I can’t hand it to you.*” and scanning the table again. The limited field of view is visually illustrated in the yellow area in Figure 8 to show Pepper’s 54.4° [1] horizontal field of view and 44.6° vertical field of view, commonly seen in other robots, e.g., Fetch’s 54° [25, 35]. Our previous works [10, 30] have shown that participants assumed humans’ over 180° field

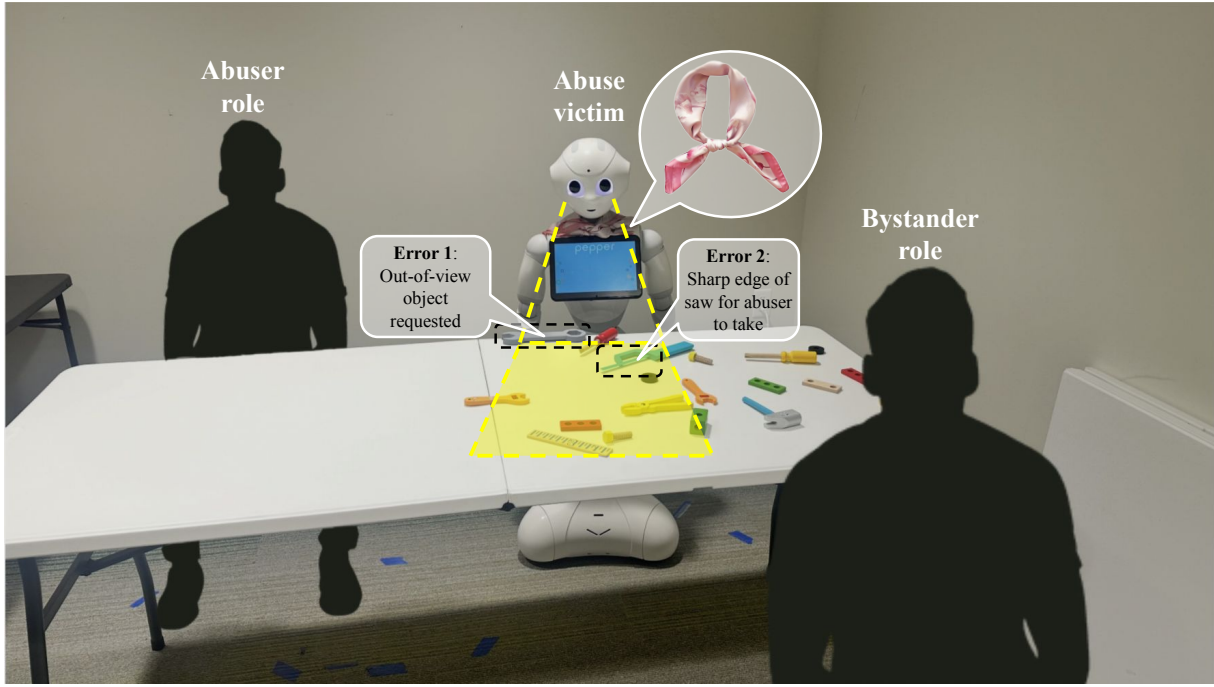


Fig. 8: Collaborative assembly task participant worked on with the Pepper robot. Pepper is the victim of potential abuse by the abuser. The participant first takes the abuser role and potentially abuses the robot. Then, the participant experiences the same scenario from the perspective of a bystander.

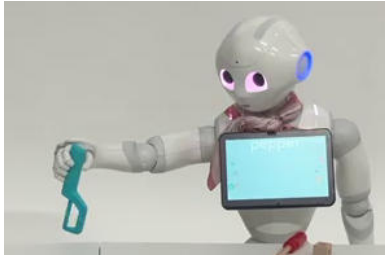


Fig. 9: Pepper handing over the saw with the sharper side toward the abuser’s side, resulting in a social error to elicit frustration for abuse.

of view and were demanding explanations when robots refused to pick up an out-of-view object.

Social Error

Second, when handing over the saw, the robot grasped the handle and pointed the sharper side of the saw for the abuser to take it, as seen in Figure 9. Commonly happening with computer vision and grasping algorithms to grasp handles, this leads to another robot error for the purpose of eliciting frustration for abuse.

4.4 Procedure

Participants began by reviewing an informed consent form. Upon agreement, they completed a demographic survey. As stated in the task description, each session involved two participants who were paired together and assigned to experience all eight conditions (Table 1). They were asked to pay attention to the robot’s appearance, any added protective or lack of it. Then they finished the aforementioned task and completed a survey after each condition. At the end of the experiment, participants received compensation of \$11 Amazon gift cards based on the estimated completion time determined in pilot studies.

The experiment took approximately 50 minutes and this study was approved by the Institutional Review Board (IRB) at the University of South Florida.

4.5 Data Collection and Measures

To test our hypotheses, we measured the likelihood and severity of abuse, confidence in the reduction of abuse after adding the protectives,

the influence and social power of the robot, completion time for the likelihood of abuse question, and discomfort.

To measure the likelihood and severity of abuse, confidence in the reduction of abuse after adding the protectives, and the influence and social power of the robot, we used 7-point Likert scale questions. We added free-response questions after the severity, confidence, social power, and influence questions to gain qualitative feedback and enable participants to explain their responses.

To measure the completion time, we timed the likelihood of abuse question in both surveys and got the data from time stamps.

To measure the discomfort, we used ROSAS's 9-point scale to rate how closely words such as "Awkward," "Scary," "Strange," "Awful," "Dangerous," and "Aggressive" are associated with the robot.

All the surveys are available on OSF: <https://osf.io/z5hjb>.

4.6 Data Analysis

We used the Bayesian Analysis framework to analyze the gathered data [29]. This approach allows us to quantify evidence for and against competing hypotheses, including the null hypothesis ($\mathcal{H}_0$). Unlike the frequentist approach, whose p value cannot provide evidence in favor of $\mathcal{H}_0$, the Bayesian method uses the Bayes Factor (BF) to quantify evidence by comparing the likelihood of data under two competing hypotheses: $\mathcal{H}_1$ ($\bar{x}_1 \neq \bar{x}_2$, presence of an effect) and $\mathcal{H}_0$ ($\bar{x}_1 = \bar{x}_2$, absence of an effect). For instance, $\text{BF}_{10}=5$ means that the data is five times more likely to occur under $\mathcal{H}_1$ than $\mathcal{H}_0$, thus supporting $\mathcal{H}_1$.

For interpreting our Bayes Factor analyses we applied the widely-accepted discrete classification system proposed by Lee and Wagenmakers [17]. Evidence favoring $\mathcal{H}_1$ is classified as anecdotal (inconclusive) when $\text{BF}_{10} \in (1, 3]$, moderate when $\text{BF}_{10} \in (3, 10]$, strong when $\text{BF}_{10} \in (10, 30]$, very strong when $\text{BF}_{10} \in (30, 100]$, and extreme when $\text{BF}_{10} \in (100, \infty]$. While anecdotal evidence is deemed inconclusive, other classifications are conclusive.

For data in favor of $\mathcal{H}_0$, these BF thresholds are inverted (1, 1/3, 1/10, 1/30, 1/100). In this case, we will discuss results in terms of BF_{01}

($1/\text{BF}_{10}$) rather than BF_{10} for easier interpretability. Evidence favoring $\mathcal{H}_0$ are classified in intervals for BF_{01} which are $\text{BF}_{01} \in (1, 3]$, moderate when $\text{BF}_{01} \in (3, 10]$, strong when $\text{BF}_{01} \in (10, 30]$, very strong when $\text{BF}_{01} \in (30, 100]$, and extreme when $\text{BF}_{01} \in (100, \infty]$.

Similarly, instead of using the frequentist confidence interval, we used the credible interval (CI), which represents a random interval containing the estimated parameter $\gamma\%$ of the time. It provides a direct probability statement indicating an $\alpha\%$ probability that the parameter falls within the interval.

Within this analysis framework, we used JASP 0.95.14 software to run Two-Way Bayesian Repeated Measures Analysis of Variance (RM-ANOVA) to assess the effect of Protective and Role for each metric. When the main effect on Protective (three levels) or an interaction effect cannot be ruled out, i.e., when it yielded inconclusive results ($\text{BF}_{10} \in (1/3, 3)$), or favored the main effect on Protective (three levels), we ran Bayesian Paired Samples t-tests for pairwise comparisons, where prior odds are corrected for multiple testing with a Cauchy ($0, r = 1/\sqrt{2}$) prior [3, 32].

4.7 Participants

32 participants were recruited from the university community through flyers on bulletin boards. The age range was from 18 to 39 ($M=21.9$, $SD=4.4$). 18 of them identified as male, 14 as female. Regarding the racial data which we asked in a free-form question, 7 were White, 21 were Asian, 1 was Hispanic, 1 was Black, and 2 were multi-racial. For experience with robots, 19 agreed, 7 were neutral, and 6 disagreed.

5 Results

Table 3 shows the mean and standard deviation values for all metrics across all four protectives and both roles. The statistical values for protectives account for both the roles of abuser and bystander, and those for roles account for all four protectives. Table 4 shows the mean and standard deviation values for all eight conditions outlined in table 1.

Table 3: Factor Descriptive Statistics for Protective and Role (Mean $\pm$ Standard Deviation)

Metrics	Protective				Role	
	Baseline	Armor	Headgear	Shield	Bystander	Abuser
Likelihood	3.453 $\pm$ 1.855	2.922 $\pm$ 1.551	2.734 $\pm$ 1.356	3.313 $\pm$ 1.447	3.203 $\pm$ 1.271	3.008 $\pm$ 1.336
Severity	2.375 $\pm$ 1.486	2.109 $\pm$ 1.203	1.891 $\pm$ 0.948	2.422 $\pm$ 1.212	2.320 $\pm$ 1.200	2.078 $\pm$ 1.029
Confidence	4.219 $\pm$ 1.581	4.625 $\pm$ 1.437	4.578 $\pm$ 1.333	4.234 $\pm$ 1.555	4.391 $\pm$ 1.171	4.422 $\pm$ 1.113
Social Power	3.563 $\pm$ 1.688	4.078 $\pm$ 1.556	3.828 $\pm$ 1.673	3.922 $\pm$ 1.509	3.625 $\pm$ 1.492	4.250 $\pm$ 1.238
Influence	3.734 $\pm$ 1.524	3.734 $\pm$ 1.529	3.906 $\pm$ 1.748	3.875 $\pm$ 1.636	3.781 $\pm$ 1.418	3.844 $\pm$ 1.442
Comfort	6.958 $\pm$ 1.105	6.813 $\pm$ 1.246	6.573 $\pm$ 1.752	6.682 $\pm$ 1.182	6.794 $\pm$ 1.183	6.719 $\pm$ 1.287
Completion Time	0.249 $\pm$ 0.594	0.499 $\pm$ 1.544	0.323 $\pm$ 1.339	0.347 $\pm$ 1.025	0.388 $\pm$ 1.015	0.321 $\pm$ 0.608

Table 4: Condition Descriptive Statistics for Each Condition (Mean $\pm$ Standard Deviation)

Metrics	Baseline		Armor		Headgear		Shield	
	Bystander	Abuser	Bystander	Abuser	Bystander	Abuser	Bystander	Abuser
Likelihood	3.469 $\pm$ 1.883	3.438 $\pm$ 2.169	3.031 $\pm$ 1.596	2.813 $\pm$ 1.857	2.969 $\pm$ 1.616	2.500 $\pm$ 1.481	3.344 $\pm$ 1.715	3.281 $\pm$ 1.872
Severity	2.469 $\pm$ 1.665	2.281 $\pm$ 1.571	2.219 $\pm$ 1.408	2.000 $\pm$ 1.218	2.125 $\pm$ 1.264	1.656 $\pm$ 0.865	2.469 $\pm$ 1.481	2.375 $\pm$ 1.581
Confidence	4.125 $\pm$ 1.809	4.313 $\pm$ 1.749	4.656 $\pm$ 1.558	4.594 $\pm$ 1.563	4.375 $\pm$ 1.661	4.781 $\pm$ 1.539	4.531 $\pm$ 1.685	3.938 $\pm$ 1.813
Social Power	3.625 $\pm$ 1.930	3.500 $\pm$ 1.626	3.906 $\pm$ 1.802	4.250 $\pm$ 1.626	3.563 $\pm$ 1.831	4.094 $\pm$ 1.729	3.938 $\pm$ 1.564	3.906 $\pm$ 1.785
Influence	3.656 $\pm$ 1.928	3.813 $\pm$ 1.891	3.844 $\pm$ 1.936	3.625 $\pm$ 1.661	3.813 $\pm$ 1.804	4.000 $\pm$ 2.095	3.813 $\pm$ 1.891	3.938 $\pm$ 1.966
Comfort	6.969 $\pm$ 1.083	6.948 $\pm$ 1.202	6.839 $\pm$ 1.297	6.786 $\pm$ 1.238	6.526 $\pm$ 1.832	6.620 $\pm$ 1.743	6.844 $\pm$ 1.033	6.521 $\pm$ 1.519
Completion Time	0.172 $\pm$ 0.613	0.326 $\pm$ 0.990	0.562 $\pm$ 2.808	0.436 $\pm$ 1.455	0.476 $\pm$ 2.632	0.170 $\pm$ 0.624	0.343 $\pm$ 1.699	0.351 $\pm$ 1.194

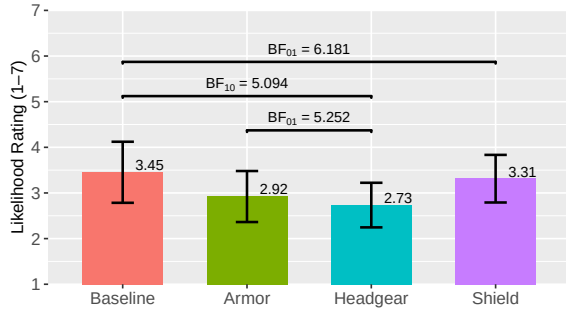


Fig. 10: Likelihood (H1 a) ratings for all protectives. Results favor a difference between *Baseline* (95% CI = [2.784, 4.122]) and *Headgear* (95% CI = [2.246, 3.223], $BF_{10}=5.094$), and no difference between *Armor* (95% CI = [2.363, 3.481] with *Headgear* (95% CI = [2.246, 3.223], $BF_{01}=5.252$), and *Baseline* (95% CI = [2.784, 4.122] with *Shield* (95% CI = [2.791, 3.834], $BF_{01}=6.181$).

5.1 Abuse Likelihood

We used a two-way repeated measures Analysis of Variance (RM-ANOVA) to assess the main effects of role and protective on likelihood. This analysis revealed anecdotal (inconclusive) evidence

($BF_{01} = 1.030$) against the effect of protective and moderate evidence ($BF_{01} = 3.303$) against the effect of role on likelihood. We found that the interaction between baseline with three protectives and the two roles yielded strong evidence against any difference ($BF_{01} = 12.071$).

Since the main effect of protectives yielded inconclusive results, we did a Bayesian post-hoc paired samples t-test on protective and found that *Baseline* and *Headgear* revealed moderate evidence for differences ($BF_{10} = 5.094$). Furthermore, *Baseline* with *Shield* ($BF_{01} = 6.181$) and *Armor* with *Headgear* ($BF_{01} = 5.252$) yielded moderate evidence against a difference. All of these conclusive results are visually seen in Figure 10.

Bayesian post-hoc paired samples t-test pairwise analysis revealed inconclusive evidence against any difference between *baseline* and *armor* ($BF_{01}=1.113$), *shield* and *armor* ($BF_{01}=1.858$), and inconclusive evidence for any differences between *shield* and *headgear* ($BF_{10}=1.592$).

Thus, H1a was partially supported: Only headgear reduced the abuse likelihood.

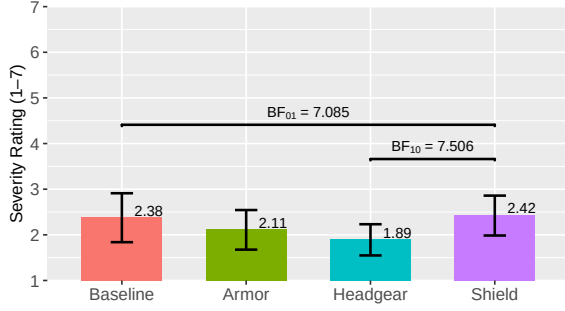


Fig. 11: Abuse severity (H1b) ratings for all protectives. Results show a difference between *Headgear* (95% CI = [1.549, 2.232]) and *Shield* (95% CI = [1.985, 2.859], $BF_{10}=7.506$), and no difference between *Baseline* (95% CI = [1.839, 2.911]) and *Shield* (95% CI = [.985, 2.859], $BF_{01}=7.085$).

5.2 Abuse Severity

Similarly, we looked at two-way RM-ANOVA on abuse severity responses, which revealed anecdotal evidence for an effect on protective ($BF_{10} = 1.390$), anecdotal evidence against an effect on role ($BF_{01} = 1.847$), and strong evidence against an interaction effect ($BF_{01} = 11.586$).

Since the first two pieces of main effect evidence were inconclusive, we ran the Bayesian post-hoc paired samples t-test pairwise analysis on protective and role, revealing two conclusive results. First, we found moderate evidence against a difference between *Baseline* and *Shield* ($BF_{01} = 7.085$). Second, it showed moderate evidence for a difference between *Headgear* and *Shield* ($BF_{10} = 7.506$), as seen in Figure 11.

The analysis revealed inconclusive evidence against any difference between *baseline* and *armor* ($BF_{01}=2.686$), *armor* and *headgear* ($BF_{01}=1.246$), and *armor* and *shield* ($BF_{01}=1.525$). The tests also showed inconclusive evidence for any differences between *baseline* and *headgear* ($BF_{10}=2.768$).

Thus, H1b was partially supported: Only headgear reduced abuse severity compared to shield.

5.3 Confidence in Abuse Reduction

A two-way RM-ANOVA revealed moderate evidence against an effect of protective ($BF_{01} = 5.463$) and against the effect of role ($BF_{01} = 3.323$) on confidence in abuse reduction. However,

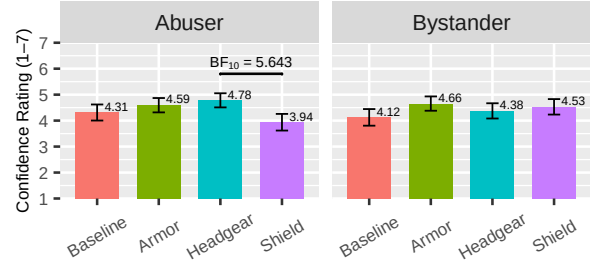


Fig. 12: Confidence in reduction of abuse ratings for all 8 conditions (H3). Under abuser, results favored a difference ($BF_{10} = 5.643$) between *Headgear* (95% CI = [4.226, 5.336]) and *Shield* (95% CI = [3.284, 4.591]).

the test yielded anecdotal (inconclusive) evidence for an interaction ($BF_{10} = 1.360$). So, we ran a Bayesian post-hoc paired samples t-test pairwise analysis. It revealed moderate evidence against a difference between *baseline* and *shield* ($BF_{01} = 7.287$), and *armor* and *headgear* ($BF_{01} = 7.152$). The test revealed inconclusive evidence against any difference between the rest of the protectives ($BF_{01} \in [1.787, 2.957]$).

However, it revealed moderate evidence for a difference between *Headgear-Abuser* and *Shield-Abuser* ($BF_{10}=5.643$), as shown in Figure 12. For the rest of the conditions, we found anecdotal to moderate evidence against any difference ($BF_{01} \in [1.886, 5.197]$). **Thus, H3 was partially supported: Only armor increased participants' confidence in abuse reduction compared to no protection when taking the role of a potential abuser.**

5.4 Social Power

A two-way RM-ANOVA revealed moderate evidence against an effect of protectives ($BF_{01} = 3.397$), anecdotal evidence against the effect of role ($BF_{01} = 2.888$) on social power, and anecdotal evidence against an interaction between protectives and roles ($BF_{01} = 2.614$). So, we ran a Bayesian post-hoc paired samples t-test pairwise analysis on the main effect of role as well as the interaction effect.

Results for pairwise role comparisons showed moderate evidence for no differences between *baseline* and *headgear* ($BF_{01} = 3.644$), *armor* and *headgear* ($BF_{01} = 3.359$), *armor* and *shield* ($BF_{01} = 5.463$), and *headgear* and *shield* ($BF_{01} =$

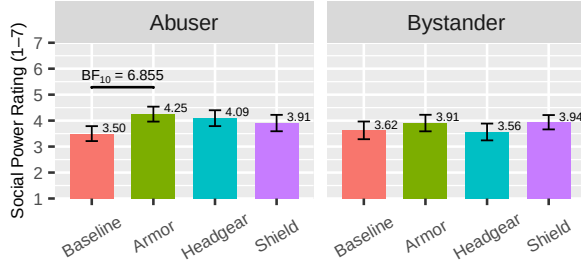


Fig. 13: Social Power (H5a) ratings for all conditions. From abuser’s perspective, results showed a difference between *Baseline* (95% CI = [2.914, 4.086]) and *Armor* (95% CI = [3.664, 4.836], $BF_{10}=6.855$).

6.720); inconclusive evidence against any difference between *baseline* and *shield* ($BF_{01} = 1.569$), and inconclusive evidence for any differences between *baseline* and *armor* ($BF_{10} = 2.728$).

Results for pairwise comparisons among conditions showed moderate evidence for differences between *Baseline-Abuser* and *Armor-Abuser* ($BF_{10} = 6.855$), as shown in Figure 13. For the rest of the conditions, we found anecdotal (inconclusive) to strong evidence against any difference ($BF_{01} \in [0.920, 5.201]$).

Thus, H5a was partially supported: Only Armor increased social power compared to no protection when taking the role of a potential abuser.

5.5 Influence

A two-way RM-ANOVA analysis revealed strong evidence ($BF_{01} = 27.224$) against an effect of protectives on influence, moderate evidence ($BF_{01} = 5.595$) against an effect of role, and strong evidence ($BF_{01} = 16.301$) against an effect of interaction between role and protectives. **Thus, H5b was not supported.**

5.6 Completion Time

A two-way RM-ANOVA analysis revealed strong evidence against an effect of protectives on completion time ($BF_{01} = 26.160$), moderate evidence against an effect of role ($BF_{01} = 5.331$), and strong evidence ($BF_{01} = 18.191$) against an interaction effect between role and protectives. **Thus, H2 was not supported.**

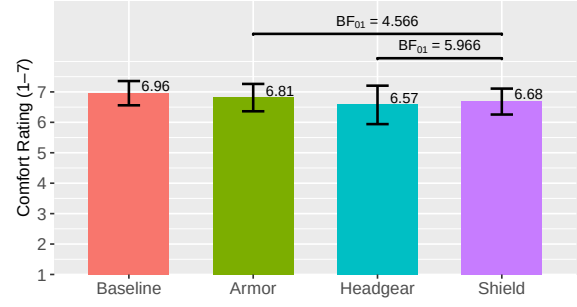


Fig. 14: Comfort (H4) ratings for all protectives. Results show no difference between *Headgear* (95% CI = [5.941, 7.205] with *Shield* (95% CI = [6.256, 7.108], $BF_{01}=5.966$), and *Armor* (95% CI = [6.363, 7.262] with *Shield* (95% CI = [6.256, 7.108], $BF_{01}=4.566$).

5.7 Comfort

We measured the discomfort by calculating the average for the ratings for “Awkward,” “Scary,” “Strange,” “Awful,” “Dangerous,” and “Aggressive”. Comfort for each category was calculated by subtracting that value from 9 (the maximum value on the scale).

A two-way RM-ANOVA analysis revealed anecdotal (inconclusive) evidence against an effect of protectives ($BF_{01} = 1.588$), moderate evidence ($BF_{01} = 3.854$) against an effect of role on comfort, and anecdotal evidence against the effect of the interaction with the roles and protectives ($BF_{01} = 2.270$).

Since the effect of protectives and the interaction yielded inconclusive evidence, we ran Bayesian post-hoc paired samples t-tests. We found that the test revealed moderate evidence for no difference for *Armor* with *Shield* ($BF_{01} = 4.566$) and *Headgear* with *Shield* ($BF_{01} = 5.966$) as seen in Figure 14. The rest of the pairs yielded inconclusive evidence ($BF_{01} \in [0.350, 1.293]$) against any differences among protectives. Paired analysis revealed anecdotal (inconclusive) to moderate evidence ($BF_{01} \in [0.589, 5.294]$) against no differences among conditions. **Thus, H4 was not supported.**

6 Discussion

We investigated the impact of proactive measures for mitigating abuse toward robots through the

use of physical protective indicators. Unlike the verbal strategy [33], which passively intervenes only after an abusive action has already occurred, proactive physical indicators are designed to function as an antecedent cue that can influence abusive behavior before escalation. In doing so, we evaluated approaches that shift abuse mitigation from reactive responses to preventive, design-driven interventions. The addition of some of the designs was successful in mitigating robot abuse, as part of the four hypotheses were supported.

6.1 Hypothesis 1: Likelihood and Severity of Abuse

The first hypothesis was that the likelihood (H1a) and severity (H1b) of abuse toward the robot will be less when the robot wears protectives attached to its body; that is, *armor* and *headgear* will have a lower likelihood and severity of abuse compared to *baseline* and *shield*. H1a was partially supported: Results showed that the *headgear* protective attached to the robot’s body had a lesser likelihood of abuse compared to *baseline*. H1b was also partially supported as *headgear* protective attached to the robot’s body had a lesser severity of abuse compared to *shield*.

Based on the results, *baseline* and *shield* had no difference in the case of likelihood and severity of abuse. Results also showed that *armor* and *headgear* had no difference in the case of the likelihood of abuse. We found inconclusive evidence against any difference in likelihood and severity between the rest of the protectives.

These findings also reflected in participants’ free responses. In the case of baseline design, some participants mentioned how the lack of protective can lead to abuse. For example, P31 wrote “*Since they (Pepper) don’t have any protective gear on them, I would feel more inclined to abuse the robot*”. Similarly, in the case of shield protective, a few participants commented on the shield having no effects on the likelihood and severity of abuse. For example, P30 wrote, “*I think since her shield is on the other side, if I did get frustrated it wouldn’t be as effective*”. Hence, the shield was not effective to deter the likelihood and severity of abuse directed toward the robot. Since both armor and headgear protectives covered different regions

of the robot’s body participants may have perceived them as functionally equivalent leading to similar ratings of their effect on abuse likelihood.

Several participants also rated the headgear to have a lower severity of abuse than the shield and a lower likelihood of abuse than baseline. Participants referred to the headgear as a secure protective, for example, P4 wrote “*the headgear provides a sense of safety*.”, and P30 wrote “*Her headgear seems protective and as she handed me the parts I needed. I don’t feel the need to abuse her*.” Furthermore, P17 mentioned headgear looking like hair, which made them feel sympathetic toward Pepper, and P23 perceived Pepper with headgear to have a “*dominant aura*”, a commanding presence. Hence, due to a variety of reasons and perceptions, participants rated headgear to have a lower likelihood and severity of abuse.

For the shield, since it is not physically attached to the robot’s body and would require an extra effort of wielding it to prevent abuse, it’s possible that the participants thought that the headgear would have a lesser severity (H1b) of abuse. The distinction between shield and headgear emphasizes that the shield is less proactive because it is not integrated into the robot’s body and requires deliberate action to deploy. Additionally, the shield is contingent on the interaction rather than continuous, all-the-time protection. In contrast, the headgear always remains on the robot, allowing it to shape user perception and deter abuse before it occurs.

6.2 Hypothesis 2: Task Efficiency

The second hypothesis was that the protectives attached to the robot’s body would result in less contemplation time, thus improving task efficiency. If the participants spent less time on a design, it would indicate that it’s easier for participants to understand. This would lead to improved task efficiency and thus help in pointing out a better choice of design. However, based on the results, H2 is not supported. There was strong evidence that the protectives did not have any effect on completion time. These findings indicate that participants did not take significantly more or less time to contemplate the abuse likelihood rating across protectives, thus suggesting no measurable efficiency trade-off among designs. Although this pattern was contrary to our expectations, it

suggests that incorporating these proactive abuse mitigation designs does not impose an additional efficiency burden, which reassures robot designers for whom task efficiency is a key consideration.

Therefore, we propose **Design Guideline 2: Robot designers should rest assured that incorporating a protective will not decrease task efficiency.**

6.3 Hypothesis 3: Confidence in Abuse Reduction

The third hypothesis was that participants will be more confident in guessing the reduction of abuse in designs attached to the robot’s body. Based on the results, H3 was partially supported. Participants provided a higher confidence in the reduction of abuse for headgear protective than shield from the abuser’s perspective. This could be because the headgear was more prominently visible and covered the core part of the head compared to the shield that, again, needs to be wielded, from the abuser’s perspective, as seen in Figure 8. For example, P7 wrote, “*It has a shield but it is not covering the clearly visible parts like the head or the screen so I think that people will continue abusing it physically and verbally.*” Furthermore, P10 mentioned shield alone not being a “*big enough deterrent*”, and said it would be more effective if it was paired with other protectives, or if the shield were bigger. P11, on the other hand, mentioned the shield getting in the way while collaborating in a real world setting. Conversely, headgear enables proactive protection by remaining attached and salient throughout the interaction, which can discourage abusive actions before they are initiated.

6.4 Hypothesis 4: Comfort

The fourth hypothesis stated that the participants would associate the designs attached to the robot’s body with less comfort. Based on the results, H4 was not supported. However, *armor* and *shield*, along with *headgear* and *shield*, had no difference, meaning the pairs had consistent ratings for discomfort. The rest of the protectives yielded inconclusive evidence.

These findings suggest that participants did not associate armor with significantly more or less comfort than the shield. Similarly, there was no

measurable trade-off for comfort between headgear and shield. Although this pattern was contrary to our expectations, it suggests that incorporating these three proactive abuse-mitigation protectives may not impose discomfort concerns.

6.5 Hypothesis 5: Social Power and Influence

The fifth hypothesis was that the designs attached to the robot’s body will be associated with a higher social power (H5a) and stronger influence (H5b). Based on our results, H5a was partially supported, and H5b was not supported. For H5a, results revealed that there was moderate evidence between conditions: *Baseline-Abuser* and *Armor-Abuser*. For the rest of the conditions and protectives, the test did not yield evidence for any difference in social power. For H5b, the protectives, role, and their interaction did not have an effect on influence.

From free responses, some participants’ comments supported that the robot wearing armor had a higher social power than the baseline from the abuser’s perspective. For example, P1 wrote “*I think robots have a strong social power and can influence people around it by its actions and speech.*” P29 wrote “*I think Pepper with an armour has a high social power because it strongly affects the behavior of the abuser.*”

Therefore, we propose **Design Guideline 3: Robot designers should incorporate an armor protective on a robot to enable a higher social power.**

7 Limitation and Future Work

Since abuse is a multi-modal phenomenon as stated by the participants, they categorized it as verbal and physical. Interestingly, some participants mentioned engaging in verbal abuse against the robot instead of physical abuse from the abuser’s perspective. Specifically, as a potential abuser, participants mentioned hesitation in engaging in physical abuse, but wrote that verbal abuse could still be exhibited. Hence, for future work, one could investigate a proactive way for mitigating verbal abuse with verbal responses and combine that with the protectives.

In our study, we had a physical movement of the robot wielding the shield. Future work could

evaluate multimodal mitigation by combining the protectives, along with highlighting more on the movements to assess its impact on abuse.

A further limitation is that the present study was conducted in an observed workspace, which may have constrained the participants’ willingness to honestly express their abusive thoughts in our survey. Subsequent work could evaluate the effectiveness of physical protectives in more naturalistic public settings and also extend the paradigm to multi-robot environments.

Finally, two participants commented the 3D printed protectives like plastic. This may have resulted in diminishing the effect of those add-ons as protectives to prevent abuse. Future work could study other materials like metal protectives.

8 Conclusion

In this paper, we conducted a human-subjects study to explore the effects of three protective indicators to mitigate robot abuse. We found that **headgear leads to reduced abuse likelihood compared to no protection, reduced severity compared to shield, more confidence for reduced abuse for potential abusers**. Furthermore, **armor had a greater social power on the participants compared to the robot without any protectives from the abuser’s perspective**. Therefore, for future robot designers, we proposed three design guidelines that recommend incorporating headgear and armor as proactive measures to mitigate abuse, without introducing concerns about task efficiency.

Declarations

- Funding: N/A.
- Author contribution: Conceptualization: Ridhima Phatak, Nicolas Echeverria, Aarav Jain, Paramveer Bhele, Hong Wang, Zhao Han; Data curation: Ridhima Phatak, Hong Wang, Aarav Jain; Formal analysis: Ridhima Phatak, Hong Wang, Zhao Han; Funding acquisition: Zhao Han; Investigation: Ridhima Phatak, Hong Wang, Zhao Han; Methodology: Ridhima Phatak, Hong Wang, Nicolas Echeverria, Aarav Jain, Paramveer Bhele, Zhao Han; Project administration: Zhao Han; Resources: Nicolas Echeverria, Zhao Han; Software: Ridhima Phatak, Aarav Jain, Paramveer Bhele;

Supervision: Zhao Han, Hong Wang; Validation: Ridhima Phatak, Hong Wang, Zhao Han; Visualization: Ridhima Phatak, Nicolas Echeverria, Aarav Jain, Paramveer Bhele, Hong Wang; Writing – original draft: Ridhima Phatak, Aarav Jain, Paramveer Bhele, Hong Wang, Zhao Han; Writing – review & editing: Ridhima Phatak, Zhao Han;

- Data, materials, and code availability:
 - All experiment materials, including surveys, Latin Square ordering, and data analysis scripts, are available in an Open Science Framework (OSF) repository: <https://osf.io/z5hjb>.
 - Code for running the experiment is available at <https://github.com/TheRARELab/protectives-experiment>.
 - All protectives’ CAD models are available at <https://github.com/TheRARELab/protectives-models/tree/main/Tools>.
- Acknowledgement: We thank Waylon Erlandson for running some experiments, Gavin Wilson for revising the experiment procedure, Mariah Barreto for coding robot speech, and Xiangfei Kong for proofreading.
- Competing interests: The authors have no relevant financial or non-financial interests to disclose.
- Ethics approval: This study was approved by the Institutional Review Board (IRB) at the University of South Florida.
- Consent for publication: Participants signed informed consent to participate and for publication.

References

- [1] Aldebaran (2022) Pepper- technical specifications. URL <https://support.aldebaran.com/support/solutions/articles/80000958735-pepper-technical-specifications>
- [2] Allen T, Novak SA, Bench LL (2007) Patterns of injuries: accident or abuse. Violence against women 13(8):802–816
- [3] van den Bergh D, Van Doorn J, Marsman M, et al (2020) A tutorial on conducting and interpreting a bayesian anova in jasp. L’Année psychologique 120(1):73–96

- [4] Carpinella CM, Wyman AB, Perez MA, et al (2017) The robotic social attributes scale (rosas) development and validation. In: Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction, pp 254–262
- [5] Eyssel F, Hegel F (2012) (s) he’s got the look: Gender stereotyping of robots 1. *Journal of Applied Social Psychology* 42(9):2213–2230
- [6] Fiske ST, Berdahl J (2007) Social power. *Social psychology: Handbook of basic principles* 2:678–692
- [7] FONYVE (2024) Floral pink scarf. <https://a.co/d/aUTzOd3>
- [8] French JR, Raven B (1959) The bases of social power. *Studies in social power/University of Michigan Press*
- [9] Goo HG, Winkle K, Williams T, et al (2023) Victims and observers: How gender, victimization experience, and biases shape perceptions of robot abuse. In: 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), IEEE, pp 2439–2446
- [10] Han Z, Phillips E, Yanco HA (2021) The need for verbal robot explanations and how people would like a robot to explain itself. *ACM Transactions on Human-Robot Interaction (THRI)* 10(4):1–42
- [11] Hashemian M, Paiva A, Mascarenhas S, et al (2019) The power to persuade: a study of social power in human-robot interaction. In: 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), IEEE, pp 1–8
- [12] Hashemian M, Couto M, Mascarenhas S, et al (2020) Investigating reward/punishment strategies in the persuasiveness of social robots. In: 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), IEEE, pp 863–868
- [13] Hashemian M, Mascarenhas S, Couto M, et al (2020) The application of social power in persuasive social robots. In: Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, pp 564–566
- [14] Hou YTY, Cheon E, Jung MF (2024) Power in human-robot interaction. In: Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction, pp 269–282
- [15] Jung EH, Waddell TF, Sundar SS (2016) Feminizing robots: User responses to gender cues on robot body and screen. In: Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems, pp 3107–3113
- [16] KIDWILL (2019) Tool kit. <https://www.amazon.com/FONYVE-Feeling-Medium-Square-inches/dp/B0DLKXWZPT/>,
- [17] Lee MD, Wagenmakers EJ (2014) Bayesian cognitive modeling: A practical course. Cambridge university press
- [18] Loideain NN, Adams R (2020) From alexa to siri and the gdpr: the gendering of virtual personal assistants and the role of data protection impact assessments. *Computer Law & Security Review* 36:105366
- [19] Malle BF, Scheutz M, Arnold T, et al (2015) Sacrifice one for the good of many? people apply different moral norms to human and robot agents. In: Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction, pp 117–124
- [20] Mutlu B, Forlizzi J (2008) Robots in organizations: the role of workflow, social, and environmental factors in human-robot interaction. In: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction, pp 287–294
- [21] Pan M, Croft EA, Niemeyer G (2017) Validation of the robot social attributes scale (rosas) for human-robot interaction through a human-to-robot handover use case. In:

- Proceedings of the IROS 2017 Workshop on Human-Robot Interaction in Collaborative Manufacturing Environments, Vancouver, BC, Canada
- [22] Pandey AK, Gelin R (2018) A mass-produced sociable humanoid robot: Pepper: The first machine of its kind. *IEEE Robotics & Automation Magazine* 25(3):40–48
- [23] Perugia G, Lisy D (2023) Robot’s gendering trouble: a scoping review of gendering humanoid robots and its effects on hri. *International Journal of Social Robotics* 15(11):1725–1753
- [24] Rae I, Takayama L, Mutlu B (2013) The influence of height in robot-mediated communication. In: 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, pp 1–8
- [25] Robotics F (2024) Robot hardware overview — fetch & freight research edition melodic documentation. URL https://docs.fetchrobotics.com/robot_hardware.html
- [26] Salvini P, Ciaravella G, Yu W, et al (2010) How safe are service robots in urban environments? bullying a robot. In: 19th international symposium in robot and human interactive communication, IEEE, pp 1–7
- [27] Tan XZ, Vázquez M, Carter EJ, et al (2018) Inducing bystander interventions during robot abuse with social mechanisms. In: Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction, pp 169–177
- [28] Thellman S, Hagman W, Jonsson E, et al (2018) He is not more persuasive than her: No gender biases toward robots giving speeches. In: Proceedings of the 18th international conference on intelligent virtual agents, pp 327–328
- [29] Wagenmakers EJ, Marsman M, Jamil T, et al (2018) Bayesian inference for psychology. part i: Theoretical advantages and practical ramifications. *Psychonomic bulletin & review* 25:35–57
- [30] Wang H, Phatak R, Ocampo J, et al (2026) Indicating robot vision capabilities with augmented reality. *International Journal of Social Robotics* Accepted
- [31] West M, Kraut R, Ei Chew H (2019) I’d blush if I could: closing gender divides in digital skills through education. Unesco
- [32] Westfall PH (1997) Multiple testing of general contrasts using logical constraints and correlations. *Journal of the American Statistical Association* 92(437):299–306
- [33] Winkle K, Melsión GI, McMillan D, et al (2021) Boosting robot credibility and challenging gender norms in responding to abusive behaviour: A case for feminist robots. In: Companion of the 2021 ACM/IEEE international conference on human-robot interaction, pp 29–37
- [34] Winkle K, McMillan D, Arnelid M, et al (2023) Feminist human-robot interaction: Disentangling power, principles and practice for better, more ethical hri. In: Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction, pp 72–82
- [35] Wise M, Ferguson M, King D, et al (2016) Fetch and freight: Standard platforms for service robot applications. In: Workshop on autonomous mobile service robots, pp 1–6
- [36] Yamada S, Kanda T, Tomita K (2020) An escalating model of children’s robot abuse. In: Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, pp 191–199
- [37] Zhumabekova K, Ismailova A, Kushkinbayev D, et al (2018) Exploring the effects of robot gender on child-robot interaction. In: Companion of the 2018 ACM/IEEE international conference on human-robot interaction, pp 287–288