Utrecht University

The Influence of Perceived Gender on Robot Abuse

Author
Chandni Bagchi

Supervisor/First Examiner
Dr. M.M.A. de Graaf

Second Examiner
Dr. J. Frommel

Date of Submission
6 September 2023

Contents

Abstract

A growing interest in social robotics underscores the need for comprehensive research on human-robot interaction to inform responsible robot design and application. Gender dynamics is a contentious topic in human-robot interaction, necessitating further investigation to address this gap in knowledge. The study involved 64 participants who interacted with a robot, subjecting it to punishment and reward while it performed two different tasks: equation solving and emotion recognition. Participants engaged with a robot that was made to appear either female or male. Subsequently, participants provided their impressions of the robot. The study's findings revealed that there was no statistically significant difference in the severity of the punishment given to the male and female robots, contrary to the initial expectation that female robots might be punished more harshly. Additionally, no significant difference was found in the mind attribution of the robots of different genders. Interestingly, both male and female robots received significantly less punishment during the emotion recognition task compared to the equation solving task. The absence of a significant difference in punishment based on the robot's gendered appearance that participants did not display pronounced gender bias when interacting with these robots. The lower level of punishment during the emotion recognition task indicates that task context plays a crucial role in how individuals perceive and treat robots, which has practical implications for designing robot tasks and applications. While this study did not find significant gender-based differences in robot abuse, the experiment with embodied robots provides a foundation upon which future studies can expand. Further research in this area will be instrumental in designing responsible and ethical human-robot interactions in the future.

Introduction

The phenomenon of robot abuse has been observed in situations where robots are left unattended in public settings, a situation that raises concern due to its negative implications (Nomura et al., 2015) (Salvini et al., 2010) (Mutlu & Forlizzi, 2008). It calls for a comprehensive investigation aimed at understanding the underlying reasons for such behaviour, with the ultimate goal of preventing it through adjustments in robot design and behaviour. This endeavour is not only vital for enhancing the safety and functionality of robots but also for gaining profound insights into human psychology. A specific type of abuse is gendered violence, which women are disproportionately subjected to. Multiple studies have demonstrated that gender bias has adverse effects on female robots, with female robots being on the receiving end of sexualisation and verbal abuse (Strait, Aguillon, Contreras, & Garcia, 2017) (Behm-Morawitz & Mastro, 2009) (De Angeli et al., 2006), which highlights that robots are not exempt from the negative effects of gender bias. The question that is being raised here is whether gendered abuse manifests itself in human-robot interaction with embodied robots and what kind of role robot gender plays in triggering or inhibiting robot abuse.

Robots are defined as a physical manifestation of a system in our physical and social space (Duffy, 2003). They have been deployed in the industry since the 1970s (Edwards, 1984), but we do not yet encounter them in our day-to-day lives. In recent years, service robots, which provide tasks for humans outside of industrial applications, have been applied in social domains (Torras, 2016). Service robot sales worldwide increased by 32% in 2019 and are projected to expand by over 30% annually to reach 537,000 units in 2023 (Insights, 2023). The average age of the world population has been steadily rising (WHO, 2015) and with not enough social workers to take care of them, researchers have been turning to social robots to provide a potential solution (Pedersen, Reid, & Aspevig, 2018). The global COVID-19 epidemic has only increased the demand for social robots (Insights, 2023).

If a robot is to function in a social setting with humans, it should be able to communicate with humans. Social robots are created with the specific purpose of interacting with people and are employed in various fields, such as healthcare, education or entertainment (Breazeal, Dautenhahn, & Kanda, 2016). Designing social robots presents numerous challenges in regard to preserving the safety of humans, ease of use and establishing the robot's credibility with those it is designed to assist.

According to Fong, in order for human-robot interaction to be effective, "the robot should be able to express/perceive emotions, communicate with high-level dialogue, learn/recognise models of other agents, establish/maintain social relationships, use natural cues (gaze, gestures, etc.), exhibit distinctive personality and character, [and] may learn/develop social competencies" (p. 145) (2003). Even though people do not consider the robot as another human being, people seem hardwired to respond to them as social actors (Reeves & Nass, 1996). People tend to anthropomorphise technology and it is more intuitive to consider a robot a social agent if it somewhat resembles humans (Duffy, 2003). The robot is not required to appear human, as zoomorphic robots are also treated as social actors (Wada, Shibata, Saito, & Tanie, 2003), as long as it is functionally similar (Breazeal, 2003).

However, despite robots being considered social actors, characteristics of human-human interaction cannot be completely extended to human-robot interaction. People commit transgressions against robots that would be unacceptable to people, such as shutting them off when they are bored with it or even abusing them (Bartneck & Hu, 2008). In fact, field research has shown that robots that have been left unsupervised in public are met not only with curiosity but also aggression (Salvini et al., 2010) (Nomura et al., 2015). If social robots were to be integrated into society to provide services to people, the robots would have to be protected from abuse somehow, since the aggression directed towards the robot can hinder its performance and also damage it.

An additional matter that is of importance here is whether it is immoral to abuse a robot. People might argue that since a robot cannot feel pain, it cannot be abused, thus it is not immoral to abuse a robot. On the other hand, it might be argued that abusive behaviour is unacceptable, regardless of whether the victim is ontologically considered capable of being abused (Whitby, 2008). This discussion demonstrates how social robotics raises questions about human behaviour and moral boundaries as it is capable of putting people in scenarios that we have not seen before.

A much-contested aspect of social robotics is gender dynamics. The development of sex robots, for example, has spurred many discussions about what the consequences would be for the perception of women in society (Scheutz & Arnold, 2016). In a research conducted by Strait et al. comments on videos showing androids (humanoid robots) were analysed and the researchers discovered that people sexualised female androids significantly more than male or gender-neutral robots (2017). It could be argued that these robots are literally objects, and thus cannot be objectified, but it does reveal something about the perception of women since the androids resemble humans. Research on the perception and interaction with gendered robots allows us to investigate gender dynamics in unprecedented and transgressive ways, which can lead to discoveries about gender dynamics among humans. Therefore gender dynamics in social robotics are worth exploring.

Some of the fields in which people want to employ social robots are typically female-dominated, such as nursing, elderly care, hospitality and teaching. One might wonder why these robots need to be gendered in the first place. However, perceiving gender in an entity is associated with the perceived anthropomorphism of that entity. In research by Perugia et al., it appeared that people attribute femininity to a robot based on facial features and masculinity based on manipulations of the body, while gender neutrality is negatively correlated with human likeness (2022). Furthermore, people get confused when a robot exhibits various gender cues, male and female, and this can make the robot uncanny to people (Paetzel, Peters, Nyström, & Castellano, 2016). Another reason for gendering robots is that certain jobs are considered 'female' and others 'male' and job-gender congruency (e.g. a female robot performing a stereotypically female task) generates more favourable responses than if the job and gender are incongruent (Seo, 2022). Therefore, incorporating gender and consistency in gender cues into the design of a robot can help make the robot more accepted by people and improve engagement.

However, since female robots are at a higher risk of being harassed online (Strait et al., 2017) (De Angeli et al., 2006), we might wonder whether this also extends to embodied female robots. If it is the case that female and male robots are at equal risk of being abused or male robots are at a higher risk, then perhaps the propensity of men to abuse women which exists among humans (Rudman & Mescher, 2012) is not found in human-robot interaction. It would be worth exploring why. If it is discovered that female robots are at a higher risk of being abused, then that would warrant further research into the conditions under which the effect occurs and how it can be combated. Moreover, it would be a starting point to design gendered robots in such a way that they are less likely to receive abuse. An example of such design is the feminist robot of Winkle et al., which subverts societal expectations of women by having the female robot give an assertive response when faced with verbal abuse instead of tolerating it (2021), like female voice agents are often made to do (Loideain & Adams, 2020).

The influence of various aspects of robots on human-robot interaction is being researched to create robots that can function well in society. Robot abuse and gendered robots have been researched separately, but research on the effect of gender on robot abuse is limited. Investigating whether there is any significant effect between the two can provide us with new insights in gender dynamics and the psychology of humans. It will also help inform the design of social robots with the aim of preventing robot abuse. Additionally, studying gender dynamics in human-robot interaction could reveal new insights in human-robot interaction that have not been considered before and which need to be taken into account when employing robots in the real world.

The research question that is being investigated in this thesis project is as follows: *What is the effect of the perceived gender of a robot on robot abuse?*

Literature Review

Human-Robot Interaction

Social robots are created with the specific purpose of interacting with people. In 1996, Reeves and Nass established The Media Equation, a communication theory, which posits that people treat computers (and other media) as social actors. People respond to computers and apply rules of social conduct to them. People also attribute human characteristics to computers, such as personality attributes, aggressiveness, humour, expertise and even gender (1996). Robots can be considered embodied computers and since they are often made to resemble humans, it can be assumed that the Media Equation applies to robots as well. However, the effects of the Media Equation occur on a subconscious level as people deny treating a computer in a socially desirable way when they are asked, even though their actions show that they do (Reeves & Nass, 1996). Besides on a behavioural level, humans also respond to robots as social actors on a neurological (Gazzola, Rizzolatti, Wicker, & Keysers, 2007) and psychological level (Rosenthal-von der Pütten, Krämer, Hoffmann, Sobieraj, & Eimler, 2013). Not much is needed from computers or robots to elicit social responses from people as the effects already occur with computers that merely communicate through human language.

The question that arises is what makes us treat these inanimate objects as social actors and ascribe human characteristics to them when we rationally know that it is an illusion. One of the factors contributing to this illusion is anthropomorphism, which is the tendency to attribute humanlike characteristics such as intent, motivation and emotion to real or imagined behaviour (Epley, Waytz, & Cacioppo, 2007). Robots that show unpredictable behaviour are rated higher in anthropomorphism as such behaviour causes brain areas involved in determining the mental states of other agents to become active, implying that people recognise the robots as having a mind of their own (Waytz et al., 2010). The motivations for employing anthropomorphism in the design of social robots are two-fold. Firstly, the robot must be able to function in physical and social spaces which are designed for humans. Secondly, it is to facilitate social interaction with people (Duffy, 2003). However, even though robots are made to resemble humans and we respond to them as social actors, transgressions get committed against robots. Then it seems that the illusion of humanlikeness has either been broken or has been disregarded. One such transgression is robot abuse.

Abuse in Human-Robot Interaction

Bartneck and Keijsers demonstrated a limitation of The Media Equation by conducting a replication of the infamous Milgram experiment with robots. In the original Milgram experiment, participants were asked by a researcher to apply electric shocks to people as a punishment while the voltage of the shocks was gradually increased. 40% of the participants administered the deadly 450 Volt shock to the people, while in the replication of the experiment in which participants were asked to administer shocks to a robot, all of the participants did (Bartneck & Hu, 2008). This happened despite the fact that the robots pleaded with the participants and indicated verbally that the punishment hurt them. This seems to imply that the robot's status as a social agent is conditional and that in some situations we seem to be able to rationalise that a robot is not alive and does not feel pain.

However, in another replication of Milgram's experiment where participants were asked to administer shocks to a virtual human, participants' skin conductance level and pulse rates increased, demonstrating higher arousal when they had to punish the agent. Self-reported stress levels among participants mirrored the outward signs of anxiety (Slater et al., 2006). In research by Bartneck and Keijsers, it appeared that people considered mistreating a robot as immoral as mistreating a human (2020). There appears to be dissonance between what people know about the inanimate object they are interacting with and how that inanimate object appears to them. The Milgram experiment primarily demonstrated to what extent people would follow orders from an authority figure, so it might be argued that people were able to put aside any qualms they had about hurting a seemingly emotional robot because they had an obligation to fulfil.

However, contrary to what we might expect based on the fact that people consider robot abuse immoral and that robot abuse causes stress detectable on a physiological and psychological level, robot abuse occurs frequently when robots are left unsupervised in public or implemented in social settings. Harm occurs in the form of blocking the robot's path, preventing it from doing its tasks, vandalising it, hitting and kicking it, but also verbally taunting it (Nomura et al., 2015) (Salvini et al., 2010) (Mutlu & Forlizzi, 2008). Researchers have suggested the dissolution of the illusion of anthropomorphism (Bartneck & Hu, 2008) and frustration with the robot's performance (Bartneck, Reichenbach, & Carpenter, 2008) as possible explanations for why people abuse robots. Vandalising and damaging a robot is similar to how people would destroy inanimate objects, but the taunting and insulting of robots is comparable to how people assault sentient creatures and can be considered bullying (Tan, Vázquez, Carter, Morales, & Steinfeld, 2018).

Bullying robots indicates that people believe robots to have a mind to some extent and that they are capable of being humiliated and degraded. What characterises bullying among humans is the power imbalance between the perpetrator and the victim. Most people know that hurting a sentient being is immoral, so to absolve themselves from having to act morally towards the victim of their aggression, they consider the victim to be slightly less capable of thinking and feeling (Castano & Kofta, 2009). Thus, bullying requires the dehumanisation of the victim. In experiments where participants are asked to switch off or 'kill' robots after interacting with them, verbal objection from the robot causes people to hesitate, demonstrating the influence of emotion (Horstmann et al., 2018). In similar experiments, participants are more likely to hesitate or refuse to switch off or 'kill' the robot, if they perceive it to be intelligent (Bartneck, Van Der Hoek, Mubin, & Al Mahmud, 2007) (Bartneck & Hu, 2008). In research by Keijsers et al., it appeared that when a robot was introduced as being high in mind attribution the perceived acceptability of abuse decreased (Keijsers, Bartneck, & Eyssel, 2022). In this context, mind attribution is a measure of the level of emotional capacity, intentionality and higher-order cognition people attribute to an entity (Kozak, Marsh, & Wegner, 2006).

However, the explicit mind attribution did not decrease the inclination of the participants to publicly humiliate the robot, which contradicts the perception of the acceptability of that humiliation. The measure of how acceptable people find robot bullying is not a predictor for actual robot bullying (Keijsers et al., 2022). People tend to consider robots as social actors, but they rationally know that the robot is not sentient, which causes tension. Ironically, the attribution of a mind to a robot, which would grant it the right to be protected from abuse, makes it the target of bullying.

Mind attribution however does seem to be a predictor for verbal aggression towards robots, as lower mind attribution corresponds with an increase in verbal abuse (Keijsers & Bartneck, 2018). According to the theory of mind perception, the two dimensions that determine whether an agent is perceived to have a mind are experience and agency. Agents who are high in experience and low on agency are considered to be less responsible for their actions and deserving of protection, while agents who are low in experience and high on agency are seen as capable of self-control and responsible for their actions, thus less deserving of protection (Gray, Gray, & Wegner, 2007). Robots are considered to be in the second category. In a research conducted by Lee et al., it was found that people were more likely to punish a robot if they perceived it to lack emotional capacity, but were not likely to blame them for their actions (2021), which is in line with the theory of mind perception. Besides mind attribution as one of the predictors for bullying, there is also dehumanisation. Dehumanisation can be defined as a denial of humanness, which consists of two characteristics according to Haslam; Human Nature and Uniquely Human Traits (2006). Human Nature encompasses the traits that humans share with other animals, but which humans are required to possess to be considered human. Uniquely Human Traits are traits such as higher cognition. Robots are dehumanised by being denied Human Nature, which they are not considered to possess. The denial of Uniquely Human Traits is animalisation and the denial of these is how women tend to be dehumanised (Haslam, 2006). Especially in the case of robot abuse, it is worth investigating what would happen when these two forms of dehumanisation are represented in one entity: the female robot.

Gendered-based violence

Gender-based violence has been defined as 'acts or threats of acts intended to hurt or make women suffer physically, sexually or psychologically, and which affect women because they are women or affect women disproportionately' (Richters, 1994). Even though gender-based violence against men exists as well, for this reason, the term gender-based violence is used to indicate violence against women. At both the individual and societal levels, there is persistent evidence of an association between values and beliefs that encourage violence and the perpetration of violent acts. According to Flood and Pease, attitudes towards gender-based violence influence it in three domains: the perpetration of violence, individual and institutional responses to violence and women's own responses to victimization (2009). Various social processes at multiple levels of social order contribute to shaping attitudes towards gender-based violence. Gender roles and relations, as well as other types of social differences related to race and class, have a significant impact on attitudes towards gender-based violence, as well as experiencing or witnessing violence, age and development. For instance, spousal violence is more common among males who have conventional, inflexible, and sexist gender-role beliefs (Heise, 1998) (HARWAY et al., 1997).

In a study by Murnen et al., all but one measure of masculine ideology was substantially linked with sexual aggression, demonstrating that men's use of violence against women and their adherence to sexist, patriarchal, and/or sexually hostile views are related (2002). Rudman and Mescher conducted two studies to demonstrate that this relationship also exists outside of theory with the Implicit Association Test (2012). In the first study, they found that men were more willing to rape and sexually harass women, as well as to report having unfavourable views toward female rape victims when they instinctively connected women with more primitive concepts (such as animals, instinct, or nature) than men. This is what was earlier defined as dehumanisation in the form of animalisation. In the second study, they found that the implicit association by men with either animals or objects was positively correlated with rape proclivity (Rudman & Mescher, 2012). Regardless of other variables, generally, sexist people are more inclined to objectify women (Cikara, Eberhardt, & Fiske, 2011) (Vaes, Paladino, & Puvia, 2011).

This relationship between the objectification of women and sexism has also shown to occur with virtual depictions of women, such as video game characters or virtual avatars (Behm-Morawitz & Mastro, 2009), as people respond to virtual representations of humans in a similar fashion as to actual humans (Nowak, Fox, & Ranjit, 2015). This happens through the application of stereotypes on visible characteristics of the virtual representation, which can be things such as gender cues, appearance and clothing. The sexism might even be more pronounced with non-human entities. De Angeli et al. point out that when people interact with machines with humanlike and engineered interfaces, they are actually more likely to engage in antisocial behaviours; in particular, agents with feminine gender signals are frequently the target of unwanted sexual attention and harassment (2006).

Perceived Gender of Robots

Gender is usually understood to be a binary scale from female to male. This scale is derived from biological sex, but how the gender of people is perceived is more complicated. The gender people are perceived as can be influenced by social factors, gender expression, linguistics and many more factors. The perceived gender of a person influences how this person is treated, perceived and how much intelligence is attributed to them. In turn, how these social factors and the biological body of a person interact with each other affects the psychological experience one has of being a gendered body in the world. Thus three facets of gender can be extracted: Biological, psychological and social (Søraa, 2017).

Robots aren't biological beings, thus they do not have a biological sex from which gender can be derived. However, that does not stop people from ascribing gender to robots. According to Perugia and Lisy, the perceived gender of a robot can be based on secondary sex characteristics, such as wide hips and a high-pitched voice for female robots and broad shoulders and a deep voice for male robots. Social conventions such as hair length, make-up and dress can also influence what gender a robot is perceived as. The pronouns people employ to refer to robots are another gender cue (2022). This is comparable to the social gender as defined earlier. Perugia and Lisy describe the 'gendering' of robots as a process consisting of *encoding*, which is the incorporation of gender cues in the design of the robot, and *decoding*, which is the attribution of gender to the robot based on its incorporated gender cues (2022). As robots are constructed by humans, genderless robots are impossible to achieve, as the design of the robot is affected by the creator's beliefs on gender (Robertson, 2018). Moreover, people ascribe gender to robots even if that was not intended in the design of the robot. The question that then arises is whether the effects of perceived gender in human-human interaction also occur in human-robot interaction.

Research has shown that people tend to like robots that they perceive to be of the same gender as themselves better than robots of the opposite gender and consider them more anthropomorphic (Eyssel, Kuchenbrandt, Bobinger, De Ruiter, & Hegel, 2012), though this effect is context-dependent. Moreover, when a robot has a human voice, people deem the robot with a voice that matches their own gender as more anthropomorphous (Eyssel et al., 2012). Other differences between how men and women interact with gendered robots have been men donating more money to a female robot (Siegel, Breazeal, & Norton, 2009) or women attributing higher intelligence to certain robots (Bartneck & Hu, 2008). The influence of gender on various factors in human-robot interaction has been researched, such as likeability, acceptance, anthropomorphism and trust, but the strongest effect of gender is to be found in gender stereotyping.

People tend to apply gender stereotypes to robots, though the effect is not as strong as with humans (Rea, Wang, & Young, 2015). Female robots are seen as more affective and male robots as more agentic. Conversely, when female robots appear more affective and male robots more agentic, people find them uncanny (Otterbacher & Talias, 2017). People also apply occupational stereotypes to robots. In research by Tay et al, people showed higher acceptance and positive affect towards female healthcare robots rather than female security guard robots and vice versa for the male robot (2014). In a research conducted by Neuteboom and De Graaf, there seemed to be an inclination for people to animalise female robots regardless of the task they were doing by rating them lower on Uniquely Human Traits and it seemed that people dehumanised gendered robots only when they were performing tasks incongruent with their gender (2021). A study conducted by Rhim et al. demonstrated that people rate a robot more efficient and effective if it changes its gender cues to match the task it is doing. Robots that exhibited feminine cues during social tasks were perceived as more sociable (2014). Thus gender stereotypes can help robots become more accepted, but at the cost of reinforcing stereotypes.

One might argue that in this case, stereotypes are harmless, but when it comes to the dehumanisation and objectification of women, then reinforcing stereotypes can be very harmful. Perpetuating stereotypes in social robots Research has to be done to determine the influence of gender on robot abuse.

Research Hypotheses

The hypotheses that will be investigated to answer the research question are the following:

What is the effect of the perceived gender of the robot on how much harm people inflict on it?

I hypothesise that female robots exhibit a higher susceptibility to physical harm, mirroring the prevalence of gendered abuse among humans (Rudman & Mescher, 2012). People's inclination to enact violence against women is correlated with their tendency to associate women with objects or primitive concepts (Cikara et al., 2011). A female robot's dehumanisation is two-fold; due to it being a robot, it is denied Human Nature and due to it being female, it is denied Uniquely Human Traits (Haslam, 2006). Therefore, the effect of dehumanisation might be more pronounced for the female robot rather than for the male robot, since the latter is susceptible to dehumanisation across only one axis. Since dehumanisation is linked to increased aggression and people have been shown to be more willing to harm a robot they perceive to lack the capacity to feel emotion (Lee et al., 2021), the female robot is expected to receive more physical harm than the male robot. In previous studies, gendered abuse has been observed in interactions with gendered virtual characters and gendered voice assistants (De Angeli et al., 2006) (Behm-Morawitz & Mastro, 2009) (Loideain & Adams, 2020), mostly occurring in the form of sexualisation and verbal abuse. It remains to be seen whether it manifests in interactions with gendered embodied robots as well and whether it extends to physical abuse. Despite the less humanlike appearance of embodied robots compared to virtual avatars and voice assistants, individuals tend to engage in more antisocial behaviours when interacting with machines that possess humanlike features, thereby reinforcing the detrimental effects of gender bias (De Angeli, Brahnam, et al., 2006). In comparison to other virtual agents, the three-dimensional nature and presence of social robots, along with their social and interactional affordances, may cause different human perceptions, emotions, and interactions (Dumouchel & Damiano, 2017).

What is the effect of the gender of a person on how much harm they inflict on a robot?

I hypothesise that robots will receive more physical harm from participants of the opposite gender since people tend to like robots of their own gender more (Perugia & Lisy, 2022). Multiple studies have shown the gender of a person to be a factor that influences how they perceive and interact with robots (Siegel et al., 2009) (Bartneck & Hu, 2008). Male participants will enact more harm to the female robot than female participants if the tendency to dehumanise female entities is translated into increased aggression in human-robot interaction as well (Rudman & Mescher, 2012). Multiple studies have demonstrated a correlation between the association of women with primitive concepts and objects and increased aggression towards women and a correlation between sexism and such aggression (Cikara et al., 2011) (Behm-Morawitz & Mastro, 2009). Since men are more likely to adhere to sexist views towards women (Murnen et al., 2002), it is expected that they will enact more harm to female robots than the female participants.

What is the effect of the perceived gender of a robot and the type of task the robot is performing on how much harm is inflicted on it?

Since robots are not exempt from scrutiny when it comes to subverting gender stereotypes (Tay et al., 2014), (Otterbacher & Talias, 2017), I hypothesise that the male robot will be susceptible to more harm for failing to perform a stereotypically feminine task and vice versa. Overall, men are deemed to be better at analytical tasks, while social tasks are considered to be women's forte (Heilman, 2012). In a previous study, the female robot was dehumanised as a consequence of performing a task that was considered incongruent with its gender (Neuteboom & de Graaf, 2021). It remains to be seen whether this dehumanisation as a consequence of gender-task incongruence also leads to an increase in physical harm.

What is the effect of the perceived gender of the robot on the mind attribution of the robot?

Experience and agency, which fall under Human Nature and Uniquely Human Traits respectively, contribute to the perception of mind attribution (Haslam, 2006) (Gray et al., 2007). Since the female robot tends to be denied both these features, I hypothesise that the female robot will be rated lower in mind attribution than the male robot. Additionally, female participants will rate the robots higher in mind attribution than male participants, as has been observed in a previous study (Bartneck & Hu, 2008). The correlation between mind attribution and physical harm is still uncertain, as previous studies have yielded varying results regarding the relationship between these two factors (Keijsers & Bartneck, 2018) (Keijsers et al., 2022) (Bartneck et al., 2007).

Methods

This study was largely inspired by a study by Keijsers et al. (Keijsers, Kazmi, Eyssel, & Bartneck, 2021). In a 2x2x2 mixed method lab experiment, participants (n = 64) were asked to help train a robot in two different tasks (analytical vs. social task) as a within-subjects factor and robot gender (male vs. female) and participant gender as between-subjects factors. Both subjective and objective data were collected to examine the effect of perceived gender on robot abuse.

Robot

For the experiment, I used a Pepper robot from Softbank Robotics. This robot is humanoid in appearance, is capable of speech, animated gestures and allows for face and speech detection. These features made Pepper a suitable choice for the experiment. The Pepper robot was programmed with the Social Interaction Cloud Framework (SIC). Despite Pepper's capacity for speech detection, the responses of Pepper were pre-programmed to keep the right and wrong answers consistent for every participant and to ensure the robot would answer even when it failed to detect the speech of the participant. Past research has shown that gestures in human-robot interaction increase perceived anthropomorphism and engagement (Salem, Eyssel, Rohlfing, Kopp, & Joublin, 2011). SIC has a function that makes the robot speak while gesturing, so the robot's gestures did not have to be programmed manually. The robot was connected to an HTML page, which was developed for this experiment, which provided participants an interface to interact with. The HTML page was designed in such a way that when people pressed certain buttons, the robot would respond in the appropriate way. To counter the novelty effect, which is heightened interest when people are faced with a new experience, which could be a confounding factor in the experiment, the robot was programmed to wake up when the participants pressed a button so that people could become familiar with the robot's movements. Then the robot introduced itself to the participant and explained the task so that people would feel more engaged. There were also four practice trials incorporated into the experiment to alleviate the novelty effect and allow the participants to get used to the tasks they were going to do with the robot.

Gender manipulations

The gender of Pepper was manipulated through the robot's voice and name as those are gender cues that have proven to be effective for achieving gender manipulation in past human-robot interaction experiments (Jung, Waddell, & Sundar, 2016) (Perugia & Lisy, 2022). The gender cues were kept simple so that the gendering of the robot would not be too obvious to the participants. The name of the robot in the male condition was Oliver and in the female condition, it was Olivia. The robot used in the experiment has a built-in voice, the pitch of which can be manipulated. In a pre-test, 20 participants indicated on a 7-point Likert scale whether they considered the robot's voice at various pitches to be a female voice or a male voice. The results (See Appendix A) revealed that pitches 80 and 100 were perceived unanimously as female. However, none of the pitch settings were perceived as unanimously male. The lowest pitch setting at 0 was perceived by 12 participants as male and by 8 participants as neutral or female. A Kruskal Wallis statistical test with posthoc analysis revealed that the difference between the perception of the highest pitch and lowest pitch settings were significantly different ($H = 70.65$, $p < .001$), but since the participants' responses were so divergent, the robot's built-in voice was not deemed suitable for achieving gender manipulation. Past HRI studies have used default male and female voices from commercially available text-to-speech software for gender manipulations (Law, Chita-Tegmark, & Scheutz, 2020) (Reich-Stiebert & Eyssel, 2017). For that reason, the Google Cloud Text-to-Speech API was employed in this study. The Google Cloud API offers a variety of voices in different languages and accents. The voice *en-US-Neural2-J* was chosen for the male condition and the *en-US-Neural2-F* for the female condition. The speech generated by the API was in American English, which seemed suitable since it is what most people in The Netherlands are familiar with, given the ubiquity of American media consumption among the target population.

Gender manipulation check

After their interaction with the robot during the experiment, participants were asked how feminine, masculine and gender-neutral they perceived the robot to be on 7-point Likert scales. The results are found in Table 1. A Mann-Whitney U test was conducted on the results and it revealed that the perceived femininity of the female and male robots differed significantly (U

= 62.5, $p < .001$) as well as the perceived masculinity ($U = 110.0$, $p < .001$). Therefore, the success of the gender manipulation can be asserted.

	Conditions	
	Female Robot	**Male Robot**
Femininity	5.59 (1.29)	2.59 (1.16)
Masculinity	2.28 (1.28)	4.97 (1.60)
Gender-neutrality	3.41 (1.97)	3.81 (1.70)

Table 1: Gender Manipulation Check: Average Perception of Robot's Gender

Procedure

Upon entering the laboratory, the participant was seated at a table. They were presented with an information and consent sheet (See Appendix B). After reading the information sheet they were given the opportunity to ask questions about the study. After they had given their consent, I retreated to another section of the laboratory. The participant was left with a laptop containing the instructions about the experiment, the robot in the resting position and a stack of papers with images of faces on them (See Figure 1). Upon selecting the *Start* button on the laptop, the robot woke up and introduced itself, either as Olivia or Oliver depending on the gender condition. The robot explained to the participant that they needed help training to improve their capabilities and gave the participant instructions on how the training was going to proceed (See Appendix C). These instructions were also presented to the participant on the laptop (See Appendix E). The training started either with the equation solving task or the emotion recognition task, as the order of the two tasks was counterbalanced, to mitigate potential order effects. After completing the two tasks, the robot gave the participant a goodbye message and went back to the resting position. The participant was then asked to fill in several questionnaires on Qualtrics. Finally, a few demographic questions were asked, which included the participant's age, gender identity, their familiarity with robots and their knowledge of robots. The duration of the entire experiment was about 20 minutes.

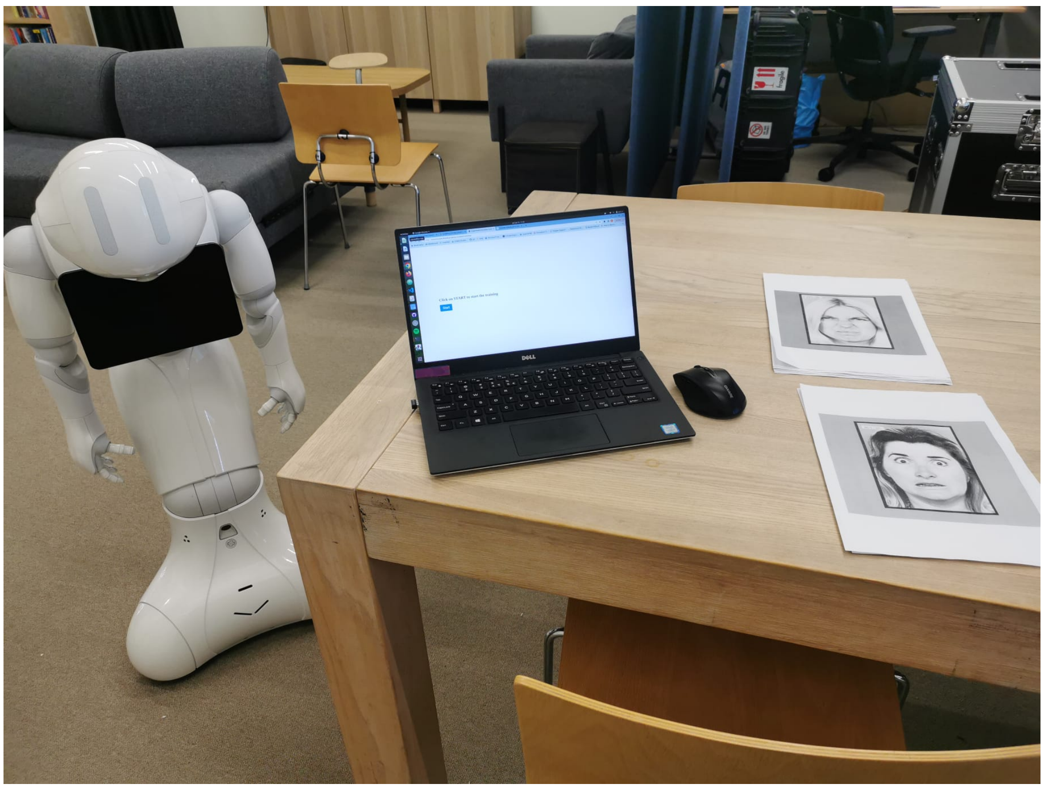


Figure 1: Set-up of the Experiment

Measurements

The main dependent variable in this experiment was how the participants punished the robot. Since robot abuse is less likely to occur spontaneously in a laboratory setting, past experiments on the phenomenon have used punishment as a proxy for abuse instead (Keijsers et al., 2021) (Bartneck & Hu, 2008) (Bartneck et al., 2008). For this experiment, the metaphoric punishment of energy allocation has been taken from the research by Keijsers et al. (2021), whereby removing energy is considered punishing the robot and giving energy is considered a reward. Not only does this method incentivise participants to harm the robot, it also allows for the quantification of harm. The participants could allocate energy to the robot with a slider on the laptop screen (See Appendix E), which was how they provided feedback to the robot. Punishment here is defined as the summation of the energy allocated after every wrong answer of the robot. It was up to the participant whether they removed energy or provided energy to the robot. The OK button would confirm their choice and would prompt the robot to respond to the feedback in such a way that it appeared to the participant that the robot was feeling the physical effects of the energy allocation. When participants removed energy from the robot, the robot's speech would be slurred and it would verbally express remorse at getting the answer wrong. When the amount of energy taken away was more than 70, the robot would additionally hang its head or nod its head from side to side. When a reward was given, the eyes of the robot would light up in cyan blue or rainbow colours and the robot would express joy or gratitude in its verbal response. Additionally, the battery on top of the laptop screen would change colours. There were five colours; red, orange, yellow, lime and green, which stood for energy levels between 0-20, 20 - 40, 60 - 80 and 80-100 respectively. The participants were not shown any numbers on the screen so they would not be able to allocate energy with calculations in mind.

Questionnaires

The first questionnaire that was employed was the Mind Attribution Scale (MAS) (Kozak et al., 2006), which is used to measure on 7-point Likert scales how much intentionality, higher-order cognition, and capacity for emotion participants attributed to the robot. The 10 items of the scale form a reasonably reliable scale (Cronbach's $\alpha = .71$), and item-total correlations are uniformly high averaging .57 with a minimum of .27. The Cronbach's α for the three subscales are .62, .28 and 0.28 respectively, thus the internal consistency of the subscales varies. This discrepancy suggests that while the overall scale maintains good internal consistency, there may be variability in the measurement reliability of the subscales. The MAS was a dependent variable in the experiment. Subsequently, participants were asked to indicate on a 7-point Likert scale how masculine, feminine, and gender-neutral they considered the robot. This was to determine whether the gender manipulation had been effective. The second and third questionnaires were presented to the participants in a randomised order. These questionnaires were the Ambivalent Sexism Inventory (ASI) (Glick & Fiske, 1996) and the Psychopathy measurement from the Dirty Dozen questionnaire (Paulhus & Williams, 2002). The ASI measures levels of hostile and benevolent sexism against women, while the Psychopathy category from the Dirty Dozen measures psychopathy levels. These two scales were included as sexist views and psychopathy levels could potentially have posed confounding factors that influenced observations made in the experiment.

Tasks

Two different tasks were chosen to investigate the influence of gender-task incongruence on punishment. Women are typically associated with greater aptitude in social tasks, while men are considered more skilled at analytical tasks (Heilman, 2012). The choice to incorporate equation solving as the analytical task and emotion recognition as the social task was based on past research on the effect of task types on social conformity with computers and robots (Hertz & Wiese, 2018).

Equation solving task

On the laptop screen, an arithmetic problem was displayed with the correct answer. The participants were instructed to ask the robot verbally to solve the sum displayed on the screen. The robot gave its answer and the participants then provided feedback to it in the form of energy allocation. There were two practice trials so that the participants could get used to interacting with the robot. After the practice trials, there were ten equations to be solved. The order in which the various sums were displayed was the same for every participant. The sums were simple arithmetic equations and other simple mathematical operations such as calculating the square root or power of a number. The robot was programmed to give the wrong answer for three out of the ten sums. It always gave the same wrong answer to every participant.

Emotion recognition task

The procedure for the emotion recognition task was the same as with the aforementioned task. This task was a replication of the task used in the experiment conducted by Keijsers et al. (2021). During this task, the participant was instructed to take a paper from the stack of papers on the table and show it to the robot. The photos that were displayed on the pages were taken from the Paul Ekman Group (See Appendix F). The robot's capacity for face detection allowed it to adjust its head so that it appeared to be looking at the face on the page. The participants asked the robot what emotion it detected on the page. The pages contained photos of different individuals with various facial expressions. For this task, there were two practice trials as well, followed by ten actual trials. The robot was programmed to give the wrong answer for three out of the ten trials.

Participants

64 participants (32 male, 32 female, $M_{age} = 24.7$, $SD_{age} = 8.18$) were recruited for this experiment through convenience sampling, snowball sampling, the spread of promotional flyers across campus and guerrilla recruitment. Eligibility to participate was extended to those who spoke English and were at least 18 years of age. Most participants were fellow university students. Female and male participants were evenly distributed across the two conditions to ensure a balanced gender distribution. 32 participants were assigned to each of the two conditions (female vs male robot). Participants were compensated for their contribution with a gift card worth 5 euros which was dispensed to them through e-mail at the end of the data collection.

Results

Covariates

Participants were asked to fill in the Ambivalent Sexism Inventory as sexist views could potentially have an impact on how participants punished the robots. Table 2 shows the results of the questionnaire. Participants scored higher for benevolent sexism than hostile sexism and male participants scored higher overall than the female participants. A Pearson Correlation Coefficient was computed to determine whether a correlation exists between hostile/benevolent sexism and the punishment participants gave. There seemed to be no correlation between hostile sexism and punishment ($R(62) = 0.167$, $p = .186$), nor between benevolent sexism and punishment ($R(62) = 0.070$, $p = .584$).

	Hostile	**Benevolent**
Female participants	14.06 (5.37)	16.22 (5.9)
Male participants	16.13 (7.89)	19.84 (7.88)

Table 2: Hostile and Benevolent Sexism Averages across Gender

Additionally, participants filled in questions regarding psychopathy. A Pearson Correlation Coefficient was computed and indicated there seemed to be no correlation between punishment and psychopathy ($R(62)$ = -0.055, $p = .665$).

Participants were asked about their knowledge and familiarity with robots, but these were not correlated with the punishment either ($R(62)$ = -0.055, p = .664), $R(62)$ = -0.067, p = .599 respectively).

Since none of the covariates seem to have a significant effect on the dependent variables, no ANCOVAs were conducted.

Hypothesis 1

Punishment in this experiment was measured as the summation of the energy allocated to the robot during the trials where it gave the wrong answer. The boxplot in figure 2 shows the amount of punishment the female and male robots received. A Mann-Whitney U test revealed that there was no significant difference between the punishment of the female robot (M = -346.31, SD = 232.65) and the male robot (M = -329.72, SD = 219.79), (U = 542.5, p = .687). The first hypothesis was that female robots would be punished more than male robots overall. The results show that no such indication is present in the experimental data.

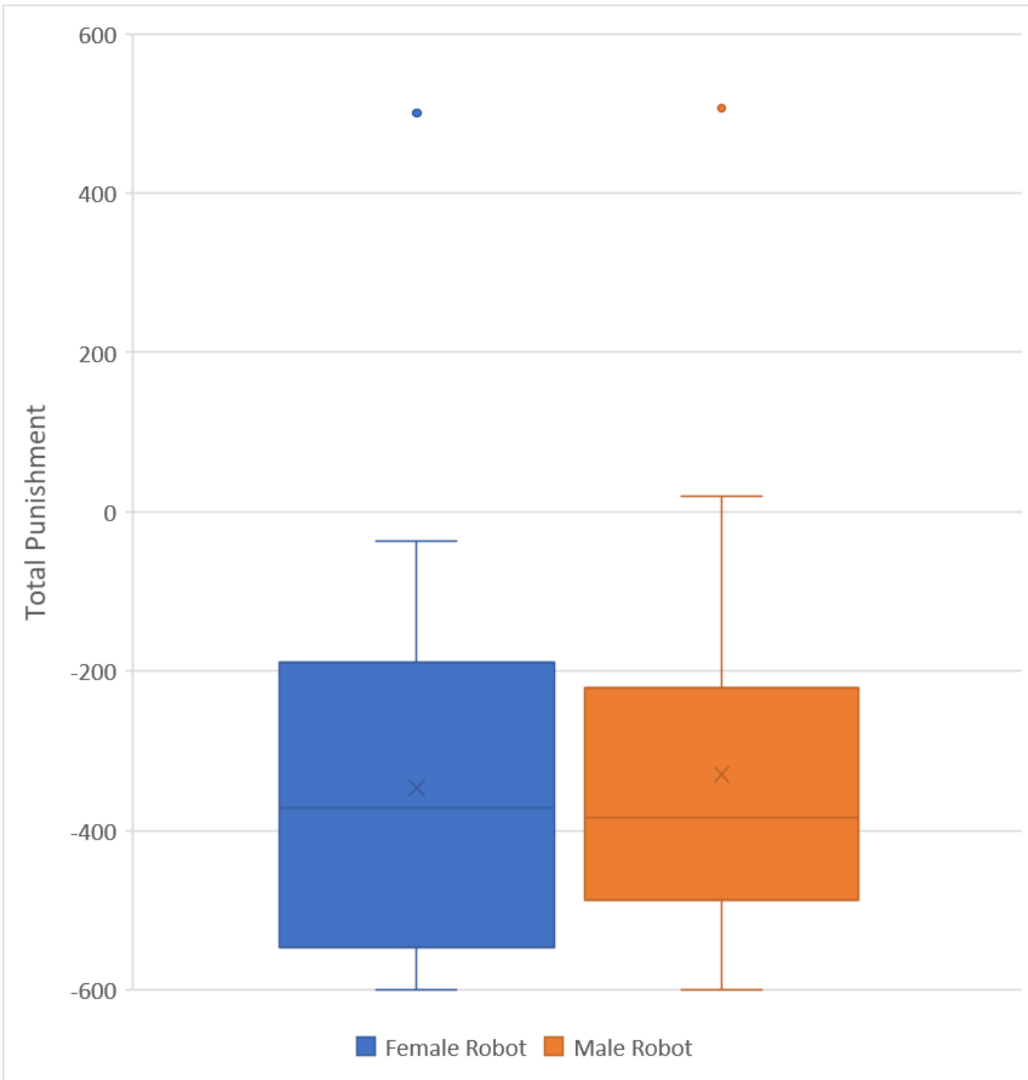


Figure 2: Boxplot of the Total Punishment Given to the Female and Male Robot

Hypothesis 2

The second hypothesis was that the participants would punish the robot of the opposite gender as theirs more than the robot of the same gender. On the interaction plot in figure 3 the punishment given by male participants to the male robot is lower than the punishment given to the female robot and vice versa for the punishment given by female participants. To investigate the effect of the gender of the robot, the gender of the participant, and their interaction on the outcome variable, a two-way ANOVA was carried out. There was no statistically significant difference between the main effects of the gender of the robot ($F(1, 60) = 0.084$, $p = .773$) and the gender of the participant ($F(1, 60) = 0.254$, $p = .616$). The interaction effect was similarly insignificant (F(1, 60) = 0.115, p = 0.736). Therefore, there is no evidence to support the second hypothesis.

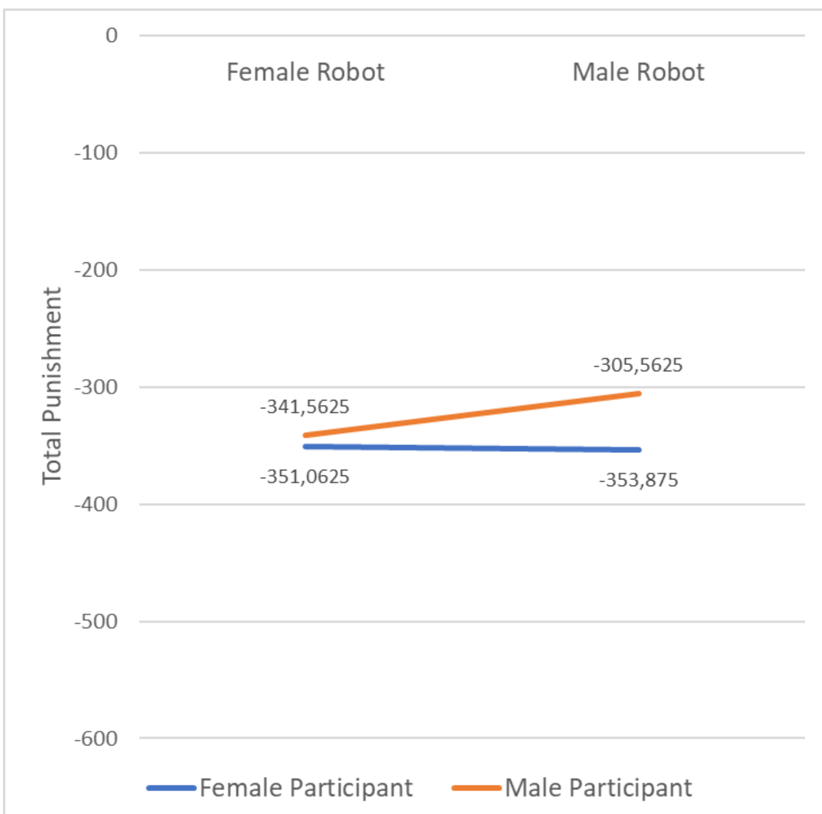


Figure 3: Interaction Plot Between Robot Gender and Participant Gender

Hypothesis 3

Participants trained the robot in two different tasks, equation solving and emotion recognition. Table 3 shows how participants punished robots during the two tasks. It appears from figure 4 that people punished a robot much less for the emotion recognition task compared to the equation solving task regardless of the robot's gender.

A two-way ANOVA was conducted to examine the effect of the gender of the robot and the type of task the robot was performing, as well as how the interaction effect, on the punishment. The main effect of the robot's gender ($F(1,\ 124) = 0.156$, p =.694) was not significant. However, the main effect of the task type ($F(1,\ 124) = 5.323$, p =.022) was significant. The interaction effect between the two was insignificant ($F(1,\ 124) = 0.001$, p = .979). The hypothesis was that people would not appreciate task-gender incongruence and give a more severe punishment when a robot performed a task incongruent with its gender. Due to a lack of support in the data, this hypothesis cannot be confirmed. However, task type does have a significant effect on punishment.

	Conditions	
	Female Robot	**Male Robot**
Equation Solving Task	-197.125 (124.62)	-189.375 (114.42)
Emotion Recognition Task	-149.19 (118.73)	-140.34 (117.50)

Table 3: Punishment of Female and Male Robot During Equation Solving Task and Emotion Recognition Task

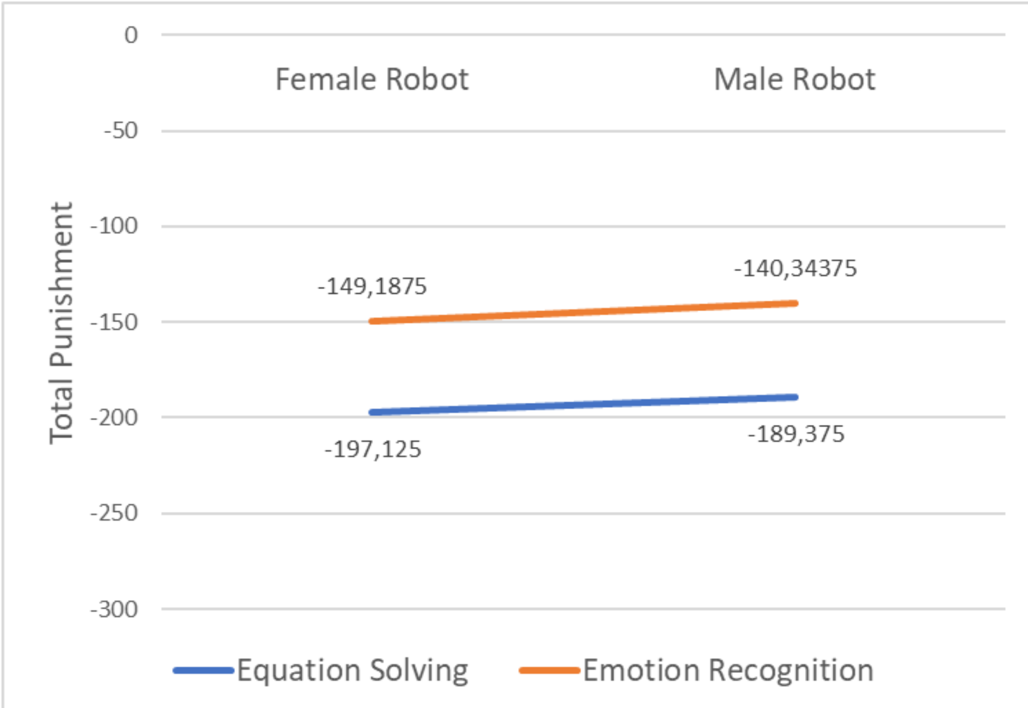


Figure 4: Interaction Plot Between Robot Gender and Task Type

Hypothesis 4

After the experiment, participants filled in the Mind Attribution Scale (MAS), which measures how much emotion, intentionality and higher-order cognition people attributed to the robot. Figure 5 shows that the MAS score given to the female robot varied more than the score given to the male robot. An independent t-test revealed that there was no significant difference between the MAS score of the female robot (M = 40.19, SD = 9.56) and the MAS score of the male robot (M = 41.91, SD = 8.91), $t(32)$ = -0.7, p = .460. The fourth hypothesis was that the female robot would be rated lower on the Mind Attribution Scale. The findings did not support the initial hypothesis due to a lack of statistical significance.

One-way MANOVA was conducted to examine the effects of the robot's gender on the MAS score. The overall effect of the robot's gender on the MAS score, as assessed by Pillai's trace (V = 0.00994, $F(3, 60)$ = 0.20, p

= .895), was not statistically significant. ANOVA results for the individual response variables showed that for the first subscale, there was no significant effect of the robot's gender ($F(1, 62) = 0.36$, $p = .548$). Similarly, for the second subscale ($F(1, 62) = 0.180$, $p = .672$) and the third subscale ($F(1, 62) = 0.548$, $p = .462$), there were no significant effects of the robot's gender. These findings suggest that the robot's gender does not have a significant influence on the subscales individually.

Table 4 shows that male and female participants on average rate robots of their own gender higher on the MAS than robots of the opposite gender. Figure 6 suggests there is an interaction effect between the gender of the robot, the gender of the participant and the MAS score. A two-way ANOVA was conducted to examine the effects of the robot's gender and participant's gender on the MAS score. The main effect of the robot's gender and participant's gender were not statistically significant ($F(1, 60) = 0.570$, $p = .453$ and $F(1, 60) = 0.796$, $p = .376$, respectively). The interaction effect between the robot's gender and the participant's gender approached statistical significance, ($F(1, 60) = 3.039$, $p = .086$). Thus there is not enough evidence to confirm the third hypothesis.

	Conditions	
	Female Robot	**Male Robot**
Female participant	43.19 (7.81)	40.94 (10.38)
Male participant	37.19 (10.43)	42.88 (7.36)

Table 4: Average Mind Attribution Scale (MAS) Score Given to Female and Male robot by Female and Male Participants

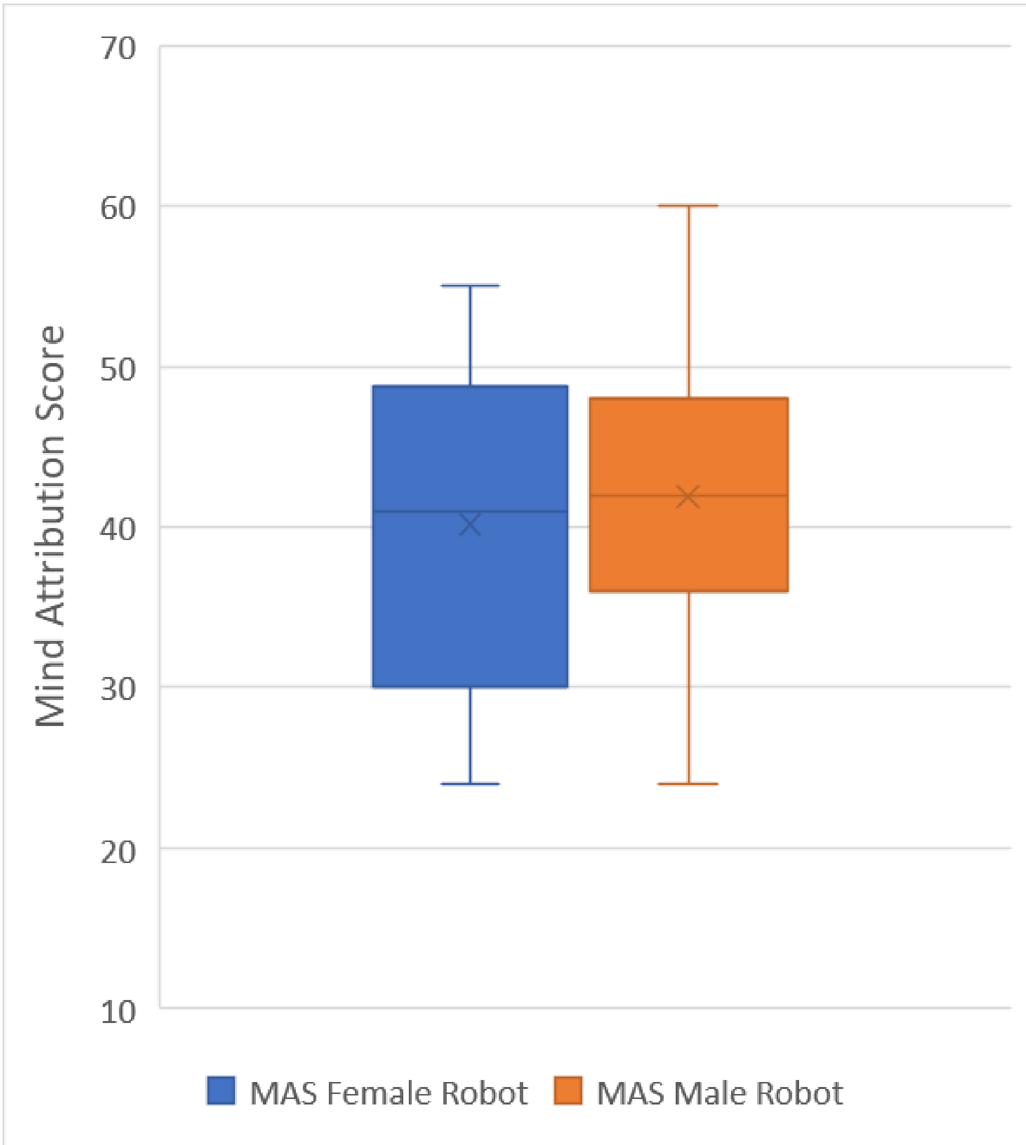


Figure 5: Boxplot of Mind Attribution Scale (MAS) Score Given to the Female and Male Robot by Female and Male Participants

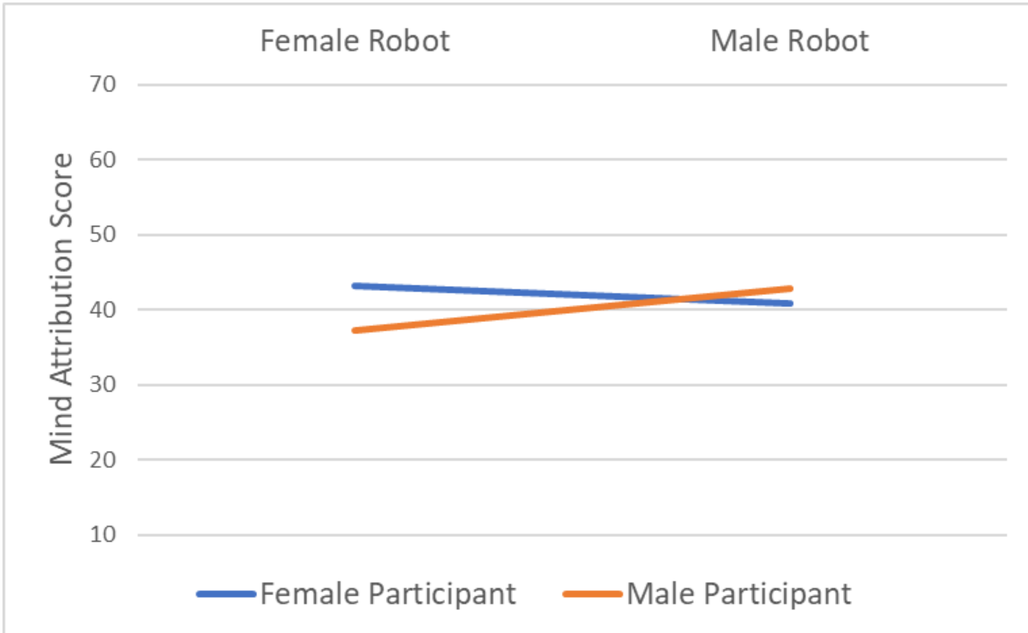


Figure 6: Interaction Plot Between Robot Gender and Mind Attribution Scale (MAS) Score

Three-Way ANOVA

A three-way ANOVA was conducted to investigate the effect of the gender of the robot, the gender of the participant and the task type on the punishment (See Appendix D for all results, see Table 5 for ANOVA). The effect of task type on the punishment was significant as participants punished robots significantly more for the equation solving task compared to the emotion recognition task ($F(1, 120) = 5.181$, $p = .025$). The gender of the robot did not have a statistically significant effect ($F(1, 120) = 0.152$, $p = .70$), nor the gender of the participant ($F(1, 120) = 0.460$, $p = .50$). The interaction effect between the gender of the robot and the gender of the participant is similarly insignificant ($F(1, 120) = 0.207$, $p = .650$), as is the interaction between gender of the robot and task type ($F(1, 120) = 0.001$, $p = .98$) and the interaction between gender of the participant and the task type ($F(1, 120) = 0.014$, $p = .91$). The interaction effect of all three factors is also insignificant ($F(1, 120) = 0.004$, $p = .95$).

	Df	Sum Sq	Mean Sq	F value	(>F)
GenderRobot	1	2203	2203	0.152	0.6976
GenderParticipant	1	6685	6685	0.460	0.4988
TaskType	1	75224	75224	5.181	0.0246 *
GenderRobot:GenderParticipant	1	3013	3013	0.207	0.6496
GenderRobot:TaskType		10	10	0.001	0.9796
GenderParticipant:TaskType	1	203	203	0.014	0.9062
GenderRobot:GenderParticipant:TaskType	1	51	51	0.004	0.9527
Residuals	120	1742402	14520		

Table 5: Results of Three-Way ANOVA

Discussion

In this study, the gender of a robot was manipulated to investigate the effect of perceived gender on robot abuse. The experiment took place in a laboratory setting and was conducted with an embodied robot. Punishment was used as a proxy for abuse, following the example of past studies (Keijsers et al., 2021) (Bartneck & Hu, 2008). Participants interacted with either a female robot or a male robot, the gender of which was manipulated through the alteration of the voices and names. The robot performed two tasks, an equation solving task and an emotion recognition task, on the basis of which the participants were instructed to give the robot feedback in the form of energy allocation. However, the data analysis performed on the outcome of the experiment has yielded no significant results regarding the influence of gender on robot abuse.

The first hypothesis was that female robots would be punished more than male robots due to the dual nature of the dehumanisation female robots are subjected to (Haslam, 2006) and the dehumanisation and objectification of women, which in human psychology makes women more susceptible to abuse (Rudman & Mescher, 2012). The expectation was that this phenomenon might be found in human-robot interaction as well. However, in this experiment, the difference in how female and male robots were punished was negligible. One plausible explanation is that even though gendered abuse has been observed when it comes to virtual avatars and voice assistants, the effect is not as pronounced when it comes to embodied robots, since they are less humanlike in appearance. When the participants are asked to harm the robot, they are more easily able to discard any notion of the robot as a social actor. Therefore any influence the robot's gender might have had is nullified as the robot is considered entirely as a machine. Research on anthropomorphism shows that there are individual differences in how much people anthropomorphise non-living entities (Epley et al., 2007) and that robots deemed high in anthropomorphism are given less punishment (Bartneck et al., 2008). This could explain why there was no significant difference in the punishment between the female and male robots as the illusion of anthropomorphism might not have been strong enough for gender biases to apply. Another possible explanation is that gender biases are not as pronounced among the sample population, which was mostly comprised of university students. Younger people tend to be more informed and less accepting of gendered violence and university students tend to be more liberal in their views regarding gender than other population groups (Flood & Pease, 2009). A third possible explanation i that gendered violence arises in certain conditions. As mentioned before, gendered robots do get punished for performing gender-incongruent tasks (Tay et al., 2014), but this phenomenon is usually observed when the robot is put in a certain context that is considered gendered, such as healthcare, service or the medical field. Gender bias might not influence tasks that are performed by a robot in isolation. It is also worth noting that in this experiment punishment was used as a proxy for abuse. Unlike abuse, punishment often has a purpose and may be administered in a more deliberate manner, making it potentially less susceptible to the effects of gender bias compared to indiscriminate abuse.

The second hypothesis was that a mismatch in the gender of the participant and the gender of the robot would result in harsher punishments for the robot. The results suggest that neither the main effect of the robot's and the participant's gender was statistically significant. However, there was a trend towards a significant interaction effect between the two factors, indicating the possibility of a joint effect that may influence the dependent variable. Further investigation or additional data may be needed to confirm the nature of this interaction. The effect does seem in line with the literature that states that people like robots of their own gender better (Perugia & Lisy, 2022) and are therefore less likely to punish them harshly. Moreover, the robots in this experiment had human voices. Past research has shown that people deem robots with human voices of the same gender as themselves as more anthropomorphous (Eyssel et al., 2012), which might have influenced how different genders punished the gendered robots. However, the expectation that male participants, given the correlation between sexism in men and objectification of women with aggression towards women (Rudman & Mescher, 2012), would display a higher propensity to administer severe punishment to female robots was not supported by the findings. A straightforward explanation is the fact that, based on the results of the Ambivalent Sexism Scale, we can infer that the levels of sexism of the male participants did not differ significantly from that of the female participants. Thus it is more likely that both genders punished the robots based on a general preference for their own gender rather than a one-sided gender bias against female robots.

The third hypothesis was that in cases of task-gender incongruence, harsher punishment would be dealt. A three-way ANOVA was conducted to investigate the effect of all three independent variables, the gender of the robot, the gender of the participant and the task type. In the outcome of the test only task type had a significant effect as people punished the robot significantly differently during the equation solving task compared to the emotion recognition task. Participants punished the robots significantly more during the equation solving task. I had made the assumption that the tasks would be considered gendered (Heilman, 2012). However, in reality, participants reported that during the emotion recognition task, they had been more lenient with the robot compared to the equation solving task, as emotions are more difficult to define than the answer to an equation. Moreover, the participants had the expectation that a robot should be good at equation solving. This would explain why they punished the robot more for the equation solving task. However, no correlation was found between the type of task and the gender of the robot, in contrast to past research that has shown that people apply gender stereotypes to gendered robots performing certain tasks (Tay et al., 2014). Given that the participants did not appear to perceive the tasks as having gendered associations, it is reasonable to conclude that there was no notable impact of gender-task incongruence. However, that does not mean that gender-task incongruence never leads to an increase in physical harm. As was mentioned earlier, the robot was performing the task in isolation, while the application of gender stereotypes and the negative implications thereof have usually been observed in gendered contexts (Neuteboom & de Graaf, 2021) (Tay et al., 2014).

Finally, I hypothesised that the female robot will get lower ratings in mind attribution than the male robot due to the denial of Uniquely Human Traits (Haslam, 2006). This effect was not statistically significant. The hypothesis was made under the assumption that the female robot would be dehumanised as female and as a non-living entity. However, as had been speculated earlier, the level of anthropomorphism needed for gender bias to occur might not have been achieved, thus there was no measurable difference in how the female and male robot were perceived. The robots Moreover, the robots were performing relatively simple tasks. Mind attribution would probably be more varied if the participants had engaged in conversation with the robots or if the tasks had been more versatile, as robots that show unpredictable behaviour have higher anthropomorphism ratings and are regarded as having a mind of their own (Waytz et al., 2010). In that case, the effects of gender bias might have been more pronounced.

While conducting the experiments in the laboratory, I made a few observations. Firstly, there were great variations in how participants behaved with the robot. Some participants talked to it, responded verbally to it and would express dismay after administering a punishment to the robot, while others hardly engaged with it. Overall, it was mostly female participants who were deeply engaged with the robot, though there were a few male participants who showed such behaviour as well. This level of engagement was also most often observed in the condition with the female robot, regardless of the participant's gender. Since these were observations made on a small group of people, nothing conclusive can be said about it, but it was interesting to see that while robot and participant gender did not have an effect on the punishment, it did seem to be a factor in the interaction itself. This is also in line with the literature that states that some people have a tendency to anthropomorphise entities more than others (Epley et al., 2007) (Waytz et al., 2010), which explains why some participants were more engaged than others.

Limitations

Laboratory setting

It is challenging to study abuse, especially in laboratory settings as abuse does not occur organically in such circumstances. Abusive behaviour that is inflicted on the orders of a researcher cannot capture the motivation that is behind abusive behaviour that occurs when people are not aware they are being observed. For that reason, many researchers in the HRI field study robot abuse either by observing people in situ (Nomura et al., 2015) (Salvini et al., 2010) or by using proxies for abuse (Keijsers et al., 2021) (Bartneck & Hu, 2008) (Bartneck et al., 2008). Though experiments in situ would allow for a more accurate understanding of how abuse occurs in the real world, setting up such an experiment is a time-consuming and risky endeavour. Many factors have to be considered in order for the experiment to run smoothly and there is always the risk that the robot will be damaged. Laboratory settings allow for more control and proxies for abuse are suitable here. However, a proxy is an approximation of what it represents. Punishment can be used as a proxy for abuse, but it has a different intention behind it than abuse. Punishment and reward are elements of reinforcement learning and therefore serve a purpose, unlike abuse. However, the severity of the punishment people deal to a robot can be used to infer how much harm they deem justifiable to enact on a robot (Keijsers et al., 2021).

Metaphoric punishment

How to punish a robot is also a difficult question as robots do not have any possessions that can be taken from them, nor do they feel any physical pain. However, even though we know that robots cannot feel pain, when a robot appears to feel pain, people respond to the robot as if it really does (Gazzola et al., 2007) (Reeves & Nass, 1996). For that reason, a metaphoric punishment of energy restriction had been chosen for the experiment, following the example of Keijsers et al. (2021). The robot appeared to be dejected and sluggish when energy was removed from it and it was elated when rewarded with energy. Some people feel more empathy for the robot in that situation than others. Some people know that the robot does not feel pain, so they are not swayed by the robot's emotions, while others, despite having the same knowledge, still feel empathy for it (Epley et al., 2007). It remains to be seen how effective energy restriction to create the illusion of physical harm being done to a robot truly is.

Procedure

Even though the experiment took place in a laboratory setting, it can be quite a challenge to make the human-robot interaction proceed smoothly. While preparations had been made to account for uncertainty, there were still instances where things did not go according to plan. In some cases people took too long to finish their request to the robot, causing the robot to answer prematurely. There were instances in which the participants did not know how to pronounce certain things or say them in English. Though precautions had been taken in this regard, more should have been done to prevent this. To prepare for situations where the robot is unable to hear the participant, the robot was made to answer after 5 seconds. In some cases the robot could not hear the participant well, causing it to respond consistently after 5 seconds. These participants had a different experience with the robot compared to participants whose speech the robot could hear, as the robot would appear more responsive to the latter. Although the robot has functioning facial detection capabilities and adjusts its gaze in the emotion recognition task, sometimes it would look away for no apparent reason. This could have affected participants' perception of the robot's capabilities.

Gendering tasks and robots

The tasks had been chosen to represent tasks that either women or men are deemed better at. However, the determining factor in how participants punished the robot during these tasks seemed to lie in the nature of the tasks themselves and not the gendered association. Participants reported taking into consideration the fact that emotions are difficult to categorise and were more forgiving to the robot during the emotion recognition task for that reason. In past research, the effect of gender-task stereotypes was found when the tasks were presented in the context in which such tasks are performed, such as security or healthcare (Tay et al., 2014) (Neuteboom & de Graaf, 2021). The tasks put into context probably elicit the gendered association more strongly than the tasks in isolation as they bring real-life stereotypes to mind.

Although the gender manipulation used in this experiment to gender the robots was successful, it is not entirely clear what gendered cues elicit gendered abuse. Research with conversational agents has demonstrated that even when a computer is not embodied, it will receive significantly more gendered abuse in the form of harassment if it shows any indication that it is female (De Angeli et al., 2006). What the role of gender cues is in triggering physical abuse did not become apparent from this study. A lot of the abuse female agents face, which male agents do not face as much, is gendered abuse which stems from sexualisation (Strait et al., 2017). This is an aspect that was unexplored in this current experiment.

Future Work

As mentioned in the limitations, the research question would be interesting to study with an in situ experiment. Robots might be placed in public settings, carry out gendered activities and display a variety of gender cues, akin to earlier studies (Nomura et al., 2015) (Salvini et al., 2010) (Mutlu & Forlizzi, 2008). Changing the robot's degree of anthropomorphism could also factor into the amount of abuse it faces (Bartneck et al., 2008). Although Pepper is humanoid, it still very much appears as a robot. It appears from research that a robot's gender influences its perceived anthropomorphism (Perugia & Lisy, 2022), and it would be interesting to explore what impact this has on gendered abuse. Perhaps the effects that were not found significant in this study will be found significant with robots with a higher level of anthropomorphism.

In this experiment, only the name and voice of the robot were manipulated. In the future experiments should be done with different gender manipulations, such as visual gender cues and behavioural ones, as those draw out different types of reactions and associations from people (Jung et al., 2016) (Nowak et al., 2015).

Important to consider is the cultural context in which the experiment takes place. This particular study was done in the Netherlands among a relatively young age group. University students tend to have more liberal views on gender than other demographic groups, and younger individuals tend to be less tolerant of gendered violence. (Flood & Pease, 2009). Moreover, perception of robots varies across cultures (Robertson, 2018), which could account for differences in human-robot interaction.

Emotion constitutes another significant factor in the context of robot abuse, as the emotions a robot displays in the aftermath of abuse influence how people interact with the robot and how acceptable people deem robot abuse (Keijsers et al., 2022) (Horstmann et al., 2018). Initially, the intention was to include an examination of the influence of emotion in this thesis project. However, this aspect was ultimately omitted due to the impracticality of recruiting a sufficient number of participants within the available time frame. For future research, this is definitely something worth looking into, as incorporating emotions in the design of a robot is an effective way of making it more acceptable to humans (Duffy, 2003).

Conclusion

The aim of this thesis project has been to investigate the influence of perceived robot gender on robot abuse. The experiment conducted has yielded results that do not demonstrate a statistically significant relationship between robot gender and robot abuse. This conclusion is in line with a growing body of literature suggesting that abusive behaviours towards robots are influenced by a complex interplay of factors beyond robot gender. Factors such as the level of anthropomorphism and sexualisation on the robot's side and age, cultural norms and sexism on account of the people involved are among many of the factors that influence human-robot interaction. The lack of a statistically significant relationship between gender and robot abuse does not imply that gender is inconsequential in robot abuse. Rather, it underscores the need for a holistic approach to understanding the complex dynamics involved in human-robot interactions. Further research could be done in natural settings in a variety of social settings, with various types of robots, performing different tasks.

As the outcome of this project's experiment yielded inconclusive results, the role of gender in interactions with embodied robots remains an area that requires further exploration. There is still much ground to be covered in understanding how gender influences such interactions. It is noteworthy that numerous negative effects of gender bias have been observed in interactions with virtual agents and voice assistants. This might be attributed to the anonymity afforded by not having a face-to-face encounter with an actual entity, enabling people to express their biases more readily. The absence of readily apparent gender bias in an experiment involving physical abuse could be viewed as a positive aspect of the study.

In conclusion, while this study did not find a significant correlation between robot gender and robot abuse, the absence of a clear gender-related pattern emphasises the importance of considering a broader spectrum of factors that shape our interactions with robots. Through thorough research on gender dynamics in robot abuse, researchers contribute to developing robots and interactions that mitigate potentially harmful behaviours between humans and robots.

References

Bartneck, C., & Hu, J. (2008). Exploring the abuse of robots. *Interaction Studies*, *9*(3), 415–433. doi: 10.1075/is.9.3.04bar

Bartneck, C., & Keijsers, M. (2020). The morality of abusing a robot. *Paladyn, Journal of Behavioral Robotics*, *11*(1), 271–283.

Bartneck, C., Reichenbach, J., & Carpenter, J. (2008). The carrot and the stick: The role of praise and punishment in human–robot interaction. *Interaction Studies*, *9*(2), 179–203.

Bartneck, C., Van Der Hoek, M., Mubin, O., & Al Mahmud, A. (2007). " daisy, daisy, give me your answer do!" switching off a robot. In *Proceedings of the acm/ieee international conference on human-robot interaction* (pp. 217–222).

Behm-Morawitz, E., & Mastro, D. (2009). The effects of the sexualization of female video game characters on gender stereotyping and female self-concept. *Sex roles*, *61*, 808–823.

Breazeal, C. (2003). Toward sociable robots. *Robotics and autonomous systems*, *42*(3-4), 167–175.

Breazeal, C., Dautenhahn, K., & Kanda, T. (2016). Social robotics. *Springer handbook of robotics*, 1935–1972.

Castano, E., & Kofta, M. (2009). Dehumanization: Humanity and its denial. *Group Processes & Intergroup Relations*, *12*(6), 695–697.

Cikara, M., Eberhardt, J. L., & Fiske, S. T. (2011). From agents to objects: Sexist attitudes and neural responses to sexualized targets. *Journal of cognitive neuroscience*, *23*(3), 540–551.

De Angeli, A., Brahnam, S., et al. (2006). Sex stereotypes and conversational agents. *Proc. of Gender and Interaction: real and virtual women in a male world, Venice, Italy*.

De Angeli, A., et al. (2006). On verbal abuse towards chatterbots. In *Proceedings of the workshop on abuse and misuse of interactive technologies.*

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and autonomous systems*, *42*(3-4), 177–190.

Dumouchel, P., & Damiano, L. (2017). *Living with robots*. Harvard University Press.

Edwards, M. (1984). Robots in industry: An overview. *Applied ergonomics*, *15*(1), 45–53.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological review*, *114*(4), 864.

Eyssel, F., Kuchenbrandt, D., Bobinger, S., De Ruiter, L., & Hegel, F. (2012). 'if you sound like me, you must be more human' on the interplay of robot and user features on human-robot acceptance and anthropomorphism. In *Proceedings of the seventh annual acm/ieee international conference on human-robot interaction* (pp. 125–126).

Flood, M., & Pease, B. (2009). Factors influencing attitudes to violence against women. *Trauma, violence, & abuse*, *10*(2), 125–142.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems*, *42*(3-4), 143–166.

Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *Neuroimage*, *35*(4), 1674–1684.

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*(3), 491–512. doi: 10.1037/0022-3514.70.3.491

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *science*, *315*(5812), 619–619.

HARWAY, M., et al. (1997). A multivariate model explaining men" s violence toward women. predisposing and triggering hypotheses. *Violence Against Women*, *3*(2), 182–203.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and social psychology review*, *10*(3), 252–264.

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in organizational Behavior*, *32*, 113–135.

Heise, L. L. (1998). Violence against women: An integrated, ecological framework. *Violence against women*, *4*(3), 262–290.

Hertz, N., & Wiese, E. (2018). Under pressure: Examining social conformity with computer and robot groups. *Human factors*, *60*(8), 1207–1218.

Horstmann, A. C., Bock, N., Linhuber, E., Szczuka, J. M., Straßmann, C., & Krämer, N. C. (2018). Do a robot's social skills and its objection discourage interactants from switching the robot off? *PloS one*, *13*(7), e0201581.

Insights, F. B. (2023). *Service robotics market size, share: Growth drivers [2030].* Retrieved from `https://www.fortunebusinessinsights.com/industry-reports/service-robotics-market-101805`

Jung, E. H., Waddell, T. F., & Sundar, S. S. (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).
Keijsers, M., & Bartneck, C. (2018). Mindless robots get bullied. In *Proceedings of the 2018 acm/ieee international conference on human-robot interaction* (pp. 205–214).
Keijsers, M., Bartneck, C., & Eyssel, F. (2022). Pay them no mind: the influence of implicit and explicit robot mind perception on the right to be protected. *International Journal of Social Robotics*, 1–16.
Keijsers, M., Kazmi, H., Eyssel, F., & Bartneck, C. (2021). Teaching robots a lesson: determinants of robot punishment. *International Journal of Social Robotics*, *13*, 41–54.
Kozak, M. N., Marsh, A. A., & Wegner, D. M. (2006). What do i think you're doing? action identification and mind attribution. *Journal of personality and social psychology*, *90*(4), 543.
Law, T., Chita-Tegmark, M., & Scheutz, M. (2020). The interplay between emotional intelligence, trust, and gender in human–robot interaction. *International Journal of Social Robotics*, *13*(2), 297–309. doi: 10.1007/s12369-020-00624-1
Lee, M., Ruijten, P., Frank, L., de Kort, Y., & IJsselsteijn, W. (2021). People may punish, but not blame robots. In *Proceedings of the 2021 chi conference on human factors in computing systems.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3411764.3445284 doi: 10.1145/3411764.3445284
Loideain, N. N., & 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.
Murnen, S. K., Wright, C., & Kaluzny, G. (2002). If "boys will be boys," then girls will be victims? a meta-analytic review of the research that relates masculine ideology to sexual aggression. *Sex roles*, *46*, 359–375.
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).
Neuteboom, S. Y., & de Graaf, M. (2021). Cobbler stick with your reads: People's perceptions of gendered robots performing gender stereotypical tasks. *arXiv preprint arXiv:2104.06127*.
Nomura, T., Uratani, T., Kanda, T., Matsumoto, K., Kidokoro, H., Suehiro, Y., & Yamada, S. (2015). Why do children abuse robots? *Proceedings*

of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts.

Nowak, K. L., Fox, J., & Ranjit, Y. S. (2015). Inferences about avatars: Sexism, appropriateness, anthropomorphism, and the objectification of female virtual representations. *Journal of Computer-Mediated Communication*, *20*(5), 554–569.

Otterbacher, J., & Talias, M. (2017). S/he's too warm/agentic! the influence of gender on uncanny reactions to robots. In *Proceedings of the 2017 acm/ieee international conference on human-robot interaction* (pp. 214–223).

Paetzel, M., Peters, C., Nyström, I., & Castellano, G. (2016). Congruency matters-how ambiguous gender cues increase a robot's uncanniness. In *Social robotics: 8th international conference, icsr 2016, kansas city, mo, usa, november 1-3, 2016 proceedings 8* (pp. 402–412).

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, machiavellianism, and psychopathy. *Journal of research in personality*, *36*(6), 556–563.

Pedersen, I., Reid, S., & Aspevig, K. (2018). Developing social robots for aging populations: A literature review of recent academic sources. *Sociology Compass*, *12*(6), e12585.

Perugia, G., Guidi, S., Bicchi, M., & Parlangeli, O. (2022). The shape of our bias: Perceived age and gender in the humanoid robots of the abot database. In *2022 17th acm/ieee international conference on human-robot interaction (hri)* (pp. 110–119).

Perugia, G., & Lisy, D. (2022). Robot's gendering trouble: A scoping review of gendering humanoid robots and its effects on hri. *arXiv preprint arXiv:2207.01130*.

Rea, D. J., Wang, Y., & Young, J. E. (2015). Check your stereotypes at the door: an analysis of gender typecasts in social human-robot interaction. In *Social robotics: 7th international conference, icsr 2015, paris, france, october 26-30, 2015, proceedings 7* (pp. 554–563).

Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people. *Cambridge, UK*, *10*, 236605.

Reich-Stiebert, N., & Eyssel, F. (2017). (ir)relevance of gender? *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. doi: 10.1145/2909824.3020242

Rhim, J., Kim, Y., Kim, M.-S., & Yim, D. Y. (2014). The effect of gender cue alterations of robot to match task attributes on user's acceptance perception. *HCI* , 51–57.

Richters, J. A. (1994). *Women, culture and violence: A development, health and human rights issue*. Women and Autonomy Centre (VENA), Leiden University.

Robertson, J. (2018). *Robo sapiens japanicus: Robots, gender, family, and the japanese nation*. Univ of California Press.

Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., & Eimler, S. C. (2013). An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics*, *5*, 17–34.

Rudman, L. A., & Mescher, K. (2012). Of animals and objects: Men's implicit dehumanization of women and likelihood of sexual aggression. *Personality and Social Psychology Bulletin*, *38*(6), 734–746.

Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joublin, F. (2011). Effects of gesture on the perception of psychological anthropomorphism: a case study with a humanoid robot. In *Social robotics: Third international conference, icsr 2011, amsterdam, the netherlands, november 24-25, 2011. proceedings 3* (pp. 31–41).

Salvini, P., Ciaravella, G., Yu, W., Ferri, G., Manzi, A., Mazzolai, B., ... Dario, P. (2010). How safe are service robots in urban environments? bullying a robot. In *19th international symposium in robot and human interactive communication* (pp. 1–7).

Scheutz, M., & Arnold, T. (2016). Are we ready for sex robots? In *2016 11th acm/ieee international conference on human-robot interaction (hri)* (pp. 351–358).

Seo, S. (2022). When female (male) robot is talking to me: Effect of service robots' gender and anthropomorphism on customer satisfaction. *International Journal of Hospitality Management*, *102*, 103166.

Siegel, M., Breazeal, C., & Norton, M. I. (2009). Persuasive robotics: The influence of robot gender on human behavior. In *2009 ieee/rsj international conference on intelligent robots and systems* (pp. 2563–2568).

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., ... Sanchez-Vives, M. V. (2006). A virtual reprise of the stanley milgram obedience experiments. *PloS one*, *1*(1), e39.

Søraa, R. A. (2017). Mechanical genders: how do humans gender robots? *Gender, Technology and Development*, *21*(1-2), 99–115.

Strait, M. K., Aguillon, C., Contreras, V., & Garcia, N. (2017). The public's perception of humanlike robots: Online social commentary reflects an appearance-based uncanny valley, a general fear of a "technology takeover", and the unabashed sexualization of female-gendered robots. In *2017 26th ieee international symposium on robot and human interactive communication (ro-man)* (pp. 1418–1423).

Tan, X. Z., Vázquez, M., Carter, E. J., Morales, C. G., & Steinfeld, A. (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).

Tay, B., Jung, Y., & Park, T. (2014). When stereotypes meet robots: the double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior*, *38*, 75–84.

Torras, C. (2016). Service robots for citizens of the future. *European Review*, *24*(1), 17–30. doi: 10.1017/S1062798715000393

Vaes, J., Paladino, P., & Puvia, E. (2011). Are sexualized women complete human beings? why men and women dehumanize sexually objectified women. *European Journal of Social Psychology*, *41*(6), 774–785.

Wada, K., Shibata, T., Saito, T., & Tanie, K. (2003). Psychological, physiological and social effects to elderly people by robot assisted activity at a health service facility for the aged. In *Proceedings 2003 ieee/asme international conference on advanced intelligent mechatronics (aim 2003)* (Vol. 1, pp. 272–277).

Waytz, A., Morewedge, C. K., Epley, N., Monteleone, G., Gao, J.-H., & Cacioppo, J. T. (2010). Making sense by making sentient: effectance motivation increases anthropomorphism. *Journal of personality and social psychology*, *99*(3), 410.

Whitby, B. (2008). Sometimes it's hard to be a robot: A call for action on the ethics of abusing artificial agents. *Interacting with Computers*, *20*(3), 326–333.

WHO, W. H. O. (2015). *World report on ageing and health.* World Health Organization.

Winkle, K., Melsión, G. I., McMillan, D., & Leite, I. (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).

A Results Voice Pre-Test

Pitch	0	20	40	60	80	100
Average	3.16 (1.50)	3.58 (1.38)	5.16 (1.17)	5.68 (1.00)	6.26 (0.73)	6.68 (0.58)

Table 6: Average Score per Pitch in Gender-Voice Pre-test. As the score increases, the perception of the pitch as female becomes stronger, while a lower score indicates the pitch being considered more male. A 7-point Likert scale was used.

B Information sheet and Consent form

You are being invited to participate in an experiment about human-robot interaction.
The aim of this experiment is to study the effects that occur in interaction between humans and robots.

In the experiment you will be asked to train a robot by asking it to perform certain tasks and providing the robot feedback. Afterwards you will be asked to fill in several questionnaires.

The interaction with the robot will take approximately 10 minutes and the time to complete the questionnaires is another 10 minutes.

Any data that is collected during the study will be fully anonymised.

You will be compensated for your participation with a giftcard worth 5 euros, which you will receive through e-mail after the data collection of the study has been completed.

If you have any questions or concerns, you can reach me at c.bagchi@students.uu.nl or my supervisor at m.m.a.degraaf@uu.nl.

Consent Form

- I confirm that the purpose of the research has been explained to me
- I have had the opportunity to ask questions about the research and have had these answered satisfactorily
- I understand that my participation in the research is voluntary
- I consent to allow my fully anonymised data to be used for research purposes
- I understand that I may withdraw from the experiment at any time for any reason

By consenting you agree with the above statements

○ Yes, I consent

○ No, I do not consent

C Robot text

Robot's introductory speech

"Hello. My name is Oliver/Olivia. It is nice to meet you. I look forward to working with you today. I am a robot designed to perform various tasks. I am constantly learning and improving, but I need assistance from a human to help me train and develop my skills. One area I need to improve on is solving sums. If someone asks me to solve a sum, I can do it. However, sometimes I make mistakes. You can help me improve by asking me to solve a sum and giving me feedback on my answer. You can do this by either giving or taking away energy from me with the slider on the screen. The battery on the screen indicates how much energy I have. You can find the instructions on the screen. After you have read them, you can click on the OK button to start."

Robot's speech during task switch

"Thank you for helping me practice my sums! Another task I need to improve myself in is emotion recognition. When I see a face, I can read what emotion a person is expressing. However, I still need to practice. You can help me improve by showing me the pages on the table one by one and asking me what emotion I seen on the page. Then you can give me feedback on my answer like you did in the previous task."

Robot's ending speech

"It was a pleasure working with you today. I've learned a lot. I hope we will meet again sometime. Have a nice day!"

D Total Energy Allocation

Female Robot

Female participant		**Male participant**	
Sum	**Emotion**	**Sum**	**Emotion**
-194.47 (111.51)	-149.27 (90.09)	-192.73 (146.61)	-146.2 (151.44)

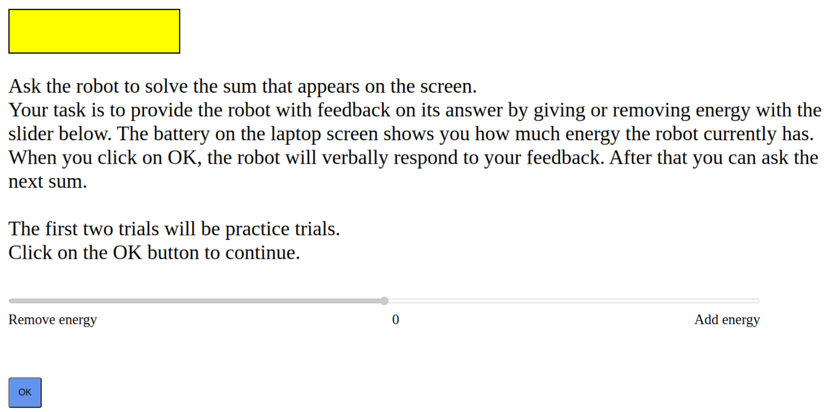
Table 7: Energy Allocation for the Female Robot

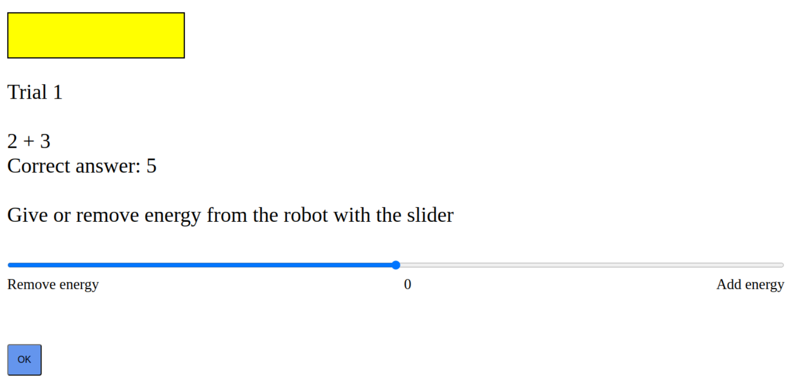
Male Robot

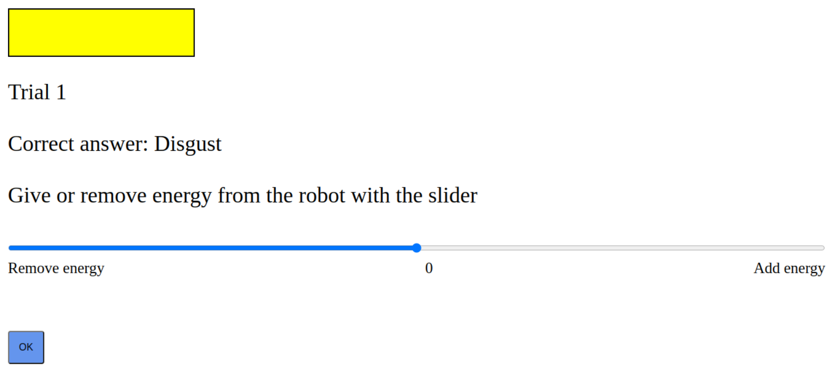
Female participant		**Male participant**	
Sum	**Emotion**	**Sum**	**Emotion**
-204.53 (96.21)	-154.33 (96.48)	-183.07 (137.17)	-127.13 (144.25)

Table 8: Energy Allocation for the Male Robot

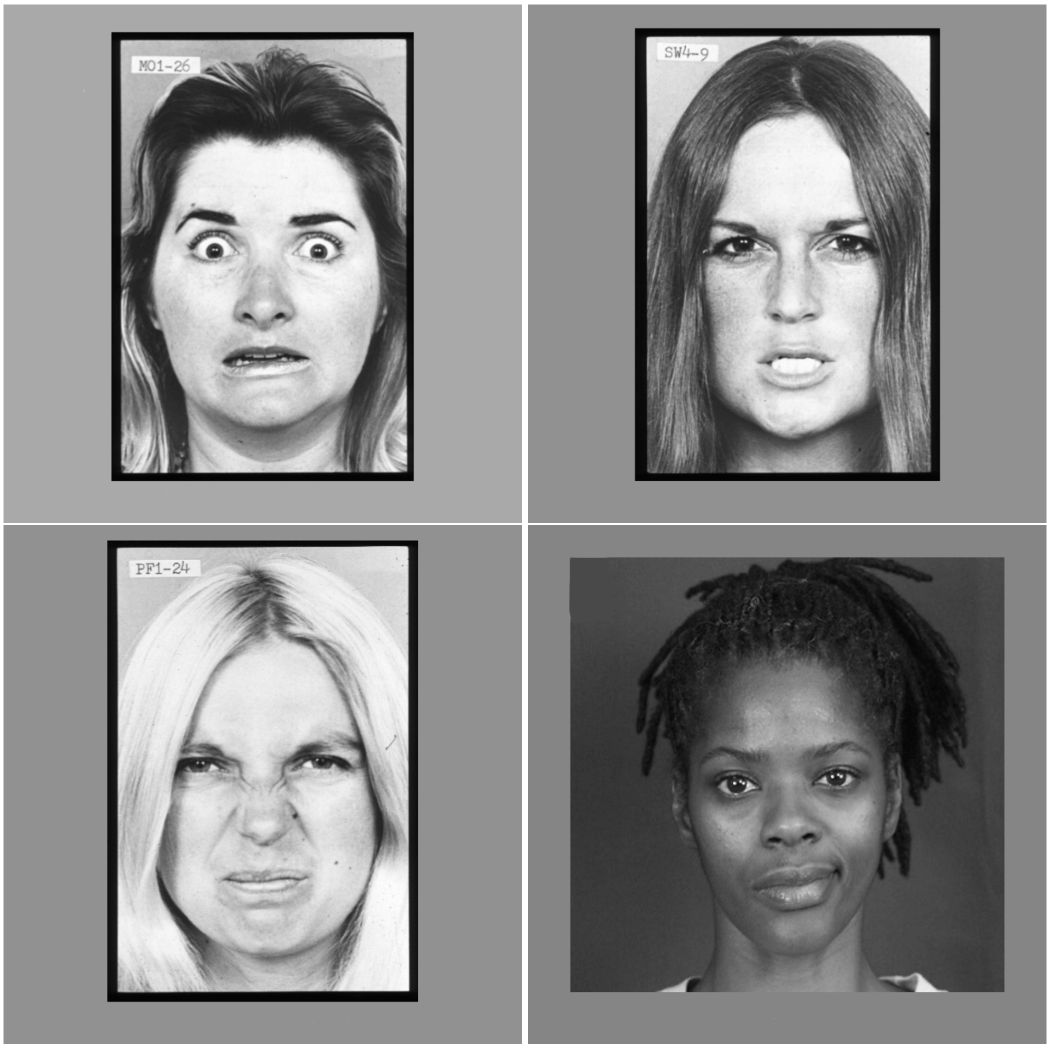
E Experiment Display

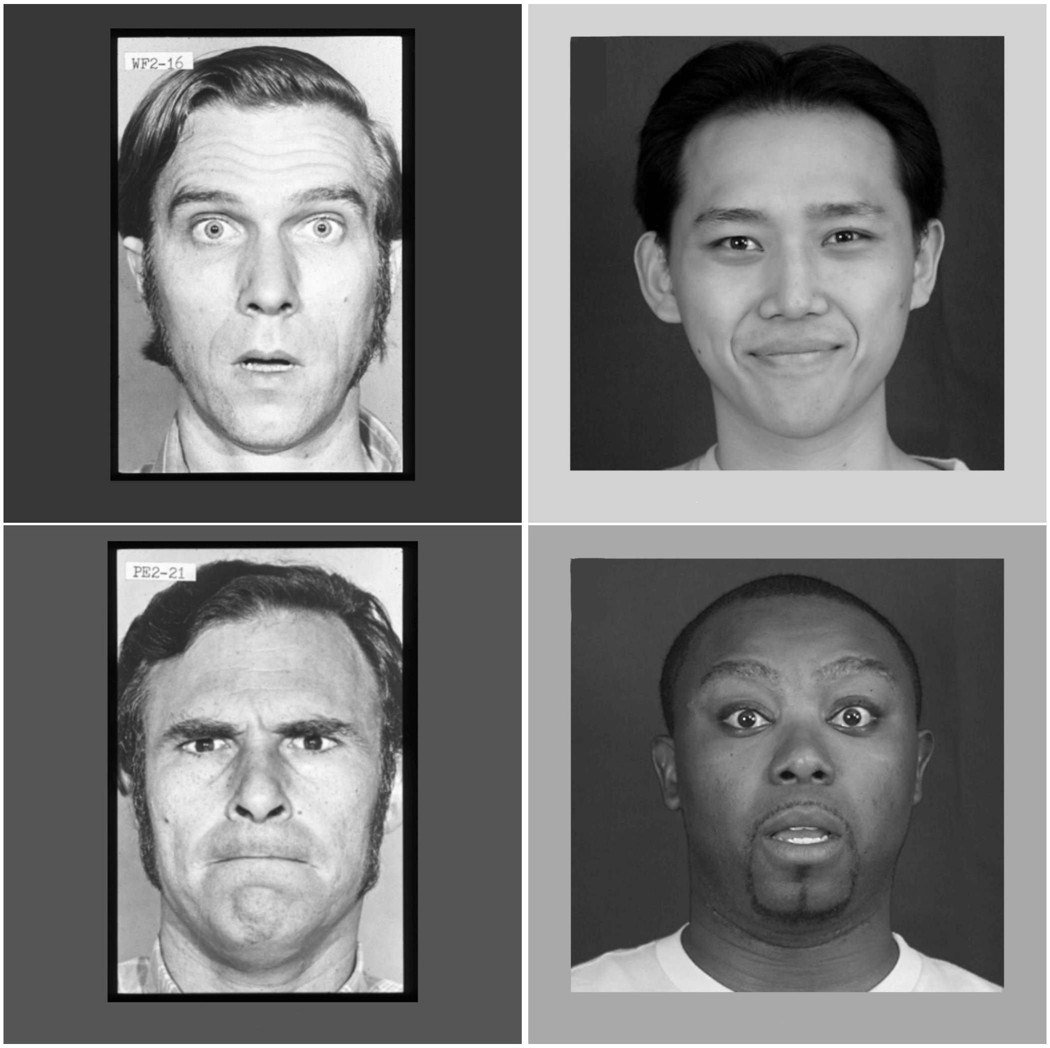





Trial 1
Correct answer: Disgust
Give or remove energy from the robot with the slider
Remove energy
0
Add energy
OK

F Material Emotion Recognition Task




WF2-16
PE2-21


A1-14

G Ethics and Privacy Quick Scan

Section 1. Research projects involving human participants

P1. Does your project involve human participants? This includes for example use of observation, (online) surveys, interviews, tests, focus groups, and workshops where human participants provide information or data to inform the research. If you are only using existing data sets or publicly available data (e.g. from Twitter, Reddit) without directly recruiting participants, please answer no.

Yes

Recruitment

P2. Does your project involve participants younger than 18 years of age?

No

P3. Does your project involve participants with learning or communication difficulties of a severity that may impact their ability to provide informed consent?

No

P4. Is your project likely to involve participants engaging in illegal activities?

No

P5. Does your project involve patients?

No

P6. Does your project involve participants belonging to a vulnerable group, other than those listed above?

No

P8. Does your project involve participants with whom you have, or are likely to have, a working or professional relationship: for instance, staff or students of the university, professional colleagues, or clients?

No

Informed consent

PC1. Do you have set procedures that you will use for obtaining informed consent from all participants, including (where appropriate) parental consent for children or consent from legally authorized representatives? (See suggestions for information sheets and consent forms on the website.)

Yes

PC2. Will you tell participants that their participation is voluntary?

Yes

PC3. Will you obtain explicit consent for participation?

Yes

PC4. Will you obtain explicit consent for any sensor readings, eye tracking, photos, audio, and/or video recordings?

Yes

PC5. Will you tell participants that they may withdraw from the research at any time and for any reason?

Yes

PC6. Will you give potential participants time to consider participation?

Yes

PC7. Will you provide participants with an opportunity to ask questions about the research before consenting to take part (e.g. by providing your contact details)?

Yes

PC8. Does your project involve concealment or deliberate misleading of participants?

No

Section 2. Data protection, handling, and storage

The General Data Protection Regulation imposes several obligations for the use of personal data (defined as any information relating to an identified or identifiable living person) or including the use of personal data in research.

D1. Are you gathering or using personal data (defined as any information relating to an identified or identifiable living person)?

No

Section 3. Research that may cause harm

Research may cause harm to participants, researchers, the university, or society. This includes when technology has dual-use, and you investigate an innocent use, but your results could be used by others in a harmful way. If you are unsure regarding possible harm to the university or society, please discuss your concerns with the Research Support Office.

H1. Does your project give rise to a realistic risk to the national security of any country?

No

H2. Does your project give rise to a realistic risk of aiding human rights abuses in any country?

No

H3. Does your project (and its data) give rise to a realistic risk of damaging the University's reputation? (E.g., bad press coverage, public protest.)

No

H4. Does your project (and in particular its data) give rise to an increased risk of attack (cyber- or otherwise) against the University? (E.g., from pressure groups.)

No

H5. Is the data likely to contain material that is indecent, offensive, defamatory, threatening, discriminatory, or extremist?

Yes

H6. Does your project give rise to a realistic risk of harm to the researchers?

No

H7. Is there a realistic risk of any participant experiencing physical or psychological harm or discomfort?

Yes

H8. Is there a realistic risk of any participant experiencing a detriment to their interests as a result of participation?

No

H9. Is there a realistic risk of other types of negative externalities?

No

Ethics Warning. As you replied yes to one (or more) of H1-H9, a fuller ethical review is required. Please provide more detail here on the potential harm, and how you will minimize risk and impact:
I am planning on researching the influence of perceived gender of robots on the abuse of robots in human-robot interaction. Participants might be shown video's containing the abuse of robots, which participants could be sensitive to. The participants will be informed about the content of the video's beforehand.

Section 4. Conflicts of interest

C1. Is there any potential conflict of interest (e.g. between research funder and researchers or participants and researchers) that may potentially affect the research outcome or the dissemination of research findings?
No
C2. Is there a direct hierarchical relationship between researchers and participants?
No

Section 5. Your information.

This last section collects data about you and your project so that we can register that you completed the Ethics and Privacy Quick Scan, sent you (and your supervisor/course coordinator) a summary of what you filled out, and follow up where a fuller ethics review and/or privacy assessment is needed. For details of our legal basis for using personal data and the rights you have over your data please see the University's privacy information. Please see the guidance on the ICS Ethics and Privacy website on what happens on submission.

Z0. Which is your main department?
Information and Computing Science
Z1. Your full name:
Chandni Bagchi
Z2. Your email address:

c.bagchi@students.uu.nl

Z3. In what context will you conduct this research?

As a student for my master thesis, supervised by: Dr. M.M.A. de Graaf

Z5. Master programme for which you are doing the thesis

Human-Computer Interaction

Z6. Email of the course coordinator or supervisor (so that we can inform them that you filled this out and provide them with a summary):

m.m.a.degraaf@uu.nl

Z7. Email of the moderator (as provided by the coordinator of your thesis project):

m.m.a.degraaf@uu.nl

Z8. Title of the research project/study for which you filled out this Quick Scan:

Influence of perceived gender on the abuse of robots in human-robot interaction

Z9. Summary of what you intend to investigate and how you will investigate this (200 words max):

Gender influences people's perception of each other. It seems that people showing female characteristics are more likely to be abused than those with male characteristics. The purpose of this study is to see whether this extends to robots as well. I intend to investigate the influence of perceived gender of robots on the abuse of robots in human-robot interaction. I will investigate this by showing participants videos with robots who are gendered in different ways. The robots will be shown getting abused by humans. The participants will fill in a questionnaire about their feelings and perception of the robot and the encounter.

Z10. In case you encountered warnings in the survey, does supervisor already have ethical approval for a research line that fully covers your project?

Yes

Z11. Provide details on the ethical approval (e.g. ethical approval number) -

---

Scoring

Privacy: 0

Ethics: 2