Examining Size Biases in Soft Robots

EMERSON SUTER, University Of Bristol, UK
HAMZA SHAHID, University Of Bristol, UK
JACKSON LAWRENCE, University Of Bristol, UK
LINDA LOMENČÍKOVÁ, University Of Bristol, UK
PRAGYA GURUNG, University Of Bristol, UK
VICTOR TRAISTARU, University Of Bristol, UK

Fig. 1. Logo

In this paper we present a series of three inflatable soft robots to demonstrate how as the scale increases, users' treatment of the robot changes. The way humans react to objects larger than them is vastly different from the way they react to smaller objects. As robots are becoming more prevalent in society, we want to see if this behavioural pattern extends towards soft robots. Focusing on soft robots would be ideal as it opens up the possibility of safe direct interaction with large robots, something that just isn't possible with large hard robotic arms.

We conducted a preliminary study where we identified the cylinder shape to be the most neutral, meaning it invoked the least influence in how the user interacted. Afterwards, we built three cylinder inflatable robots of different sizes. This paper presents the results of our second study where participants were given a task to turn the three inflatables red by interacting with them. Overall we found that generally users had more softer interactions with the smaller inflatable and were less inclined to hit it. Whereas for the larger robots, participants were much rougher. Interestingly, participants claimed to prefer punching the human sized robot the most as it was perceived to be the most suitable size to do so. Finally we delve into how our research can be applied to possible future applications such as multi-modal interactions and adaptive robot behaviour.

Additional Key Words and Phrases: Interactive Devices, Inflatable Robots, Robot Size

Authors' addresses: Emerson Suter, ph19691@bristol.ac.uk, University Of Bristol, UK; Hamza Shahid, op19007@bristol.ac.uk, University Of Bristol, UK; Jackson Lawrence, fi19710@bristol.ac.uk, University Of Bristol, UK; Linda Lomenčíková, fv19959@bristol.ac.uk, University Of Bristol, UK; Pragya Gurung, wq19451@bristol.ac.uk, University Of Bristol, UK; Victor Traistaru, kd19421@bristol.ac.uk, University Of Bristol, UK.

1 INTRODUCTION

As technology advances, the prevalence of robots in society does as well and so the amount that people interact with robots in their day to day lives rises. When people think of robots they think of hard machines controlled by motors, but equally important are soft robots. Soft robots are a type of interactive inflatable and as their name suggests are soft instead of hard like their more traditional counterparts. Because of this, how people interact with them is even more vital and needs to be carefully thought through when designing them.

The aim of this research is to build interactive inflatable robots with the goal of observing and quantifying how people interact with soft robots of different sizes. Size affects people's perception and action towards an object, which is a major concept behind our research. Jessica C. Thompson et al. in [16] discusses that humans are the only predators to routinely hunt large animals, indicating there might be some biases based on size ingrained into human perception. Additionally, [8] involves research about how the sizes of robots affect people's treatment of them. Using that study as a base, this research applies similar methodology to inflatable soft robots and attempts to analyse interactions and potential biases based on the size of the inflatables.

In this paper we carried out two different studies. The first study involved showing participants drawings of different shapes and having them fill out a questionnaire. The aim of this study was to find a neutral shape for the inflatables, so the shape would have minimal impact on participants interactions in the second study. We built a prototype inflatable using the analysed data from the questionnaire.

The second study involved getting an idea of how individuals interact with three different sized inflatables and assessed how the size of the inflatables affected these interactions. Participants were given a prompt to interact with the inflatables to turn them a red. The inflatables reacted through light-based feedback to the participants' touch and changed colour the more they were interacted with it. Following the observation study each participant completed a questionnaire to share their perception of the robots and their interactions.

The study can be used as a source of information on how people interact with different styles of inflatables. Being able to predict how people would interact with inflatables would be a valuable resource for soft robot design. Ultimately, the study generated useful data on human to inflatable interactions which can be used in many ways, by other researchers, or by manufacturers of inflatables.

2 RELATED WORK

Previous research has shown that people tend to anthropomorphise robots, and that this tendency can lead to emotional responses such as empathy or even fear [2]. This also shows that attributes such as the robot's appearance, movement, and sounds affect our emotional responses. To our knowledge, however, no study has specifically looked at how the size of an inflatable robot affects this response.

2.1 Soft Robots

Air Giants, an interactive robotics studio, creates giant inflatable soft robots as an artistic display for people to interact with and experience [3]. These inflatables are used as art works and are meant to evoke emotional responses from audiences of all ages. However, as of right now, Air Giants' inflatables are created purely for display, and have yet to be explored scientifically about why or how they cause certain emotions in people.

Somnox is another soft robot on the market, supporting users to improve their sleep through the use of empathetic soft robot interactions [13]. Designed with emphasis on an enjoyable and usable interface, the Somnox robot is an egg-shaped soft robot that effectively simulates human-like breathing patterns to provide soothing sounds. The shape and size of the robot had been selected due to its similarity with a baby, giving it comforting and affectionate properties. Though Somnox raised over 500,000 euros in 2019 [5], it has not been conclusively found if the chosen attributes for shape and size are the most effective in improving the user experience.

2.2 Human-Robot Interactions Research

Robots are generally perceived as having greater emotional capability than standard computers [1]. Whilst robots obviously do not possess inherent affection, humans tend to contrive an idea in their mind. Over the past decade, a number of science fiction films and TV series have explored the themes of robot companions or carers and how humans respond to them. Some of the most notable include Robot & Frank (dir. Jake Schreier, USA 2012) and Big Hero 6 (dir. Don Hall/Chris Williams, USA 2014). In the paper [15], author Teo examined the relationship between robot carers and found that advocating for a reciprocal working relationship between robot carers would lead to overall better quality of care.

Human emotion has been widely utilised in driving the development of robotic interaction. Some examples include stress-relief devices and social robots. However, the extent to which we understand the relationship between raw human emotion and robotic devices has largely remained unexplored, especially regarding their distinct appearance. Studies have previously shown that the size of objects does not directly influence our emotional reaction, but if an object's size matches our preconceived assumption, brain activation for that object is higher [7].

There has been previous research about human and robot interaction. [14] aimed to measure empathetic robot evaluation through emotion estimation analysis and facial expression synchronisation. The researchers created a robot to emulate human-like facial-expressions, to improve human's impression of the robots. They experimented with robots copying the facial experiment of the subject and contrasted it with robot performing the opposite expression displayed by the subject. It was found that the intimacy between humans and robots was improved when happiness was synchronised by the robot. Additionally, Marchesi et al. experiments with humanoid robots performing human-like social interactions and contrasting it with opposing machine-like behaviour [9]. Their results demonstrated that the human-like robot was more likely to make humans adopt an intentional stance (perceived them to have human behaviour), as opposed to the machine-like robot.

In the context of measuring emotion towards a robotic figure, Cornell researchers [4] explore the use of utilising cameras and soft touch to detect a range of physical interactions. They did this by attaching a USB camera inside the robot to analyse the shadows and hand gestures using machine learning software. Gestures such as poking or punching could be categorised as an aggressive or negative emotional responses, whereas soft touches could indicate a more positive emotion. Further applications discussed include gesture guided motion robots and inflatables.

2.3 Object Size

There has been research regarding object size and its influence on people's emotions and biases. In a study conducted by Reeves et al. [11], they found that larger screen sizes generally have a greater influence on people compared to smaller screen sizes, especially when they are shown arousing or attention-based content. They sought to quantify this effect by measuring participants heart rate and skin conductance whilst they view content of screens of three different sizes: 56-inch, 13-inch, and 2-inch. With the equipment used, the authors were able to measure attention and arousal relatively accurately. However, this method would be inappropriate for measuring a large population. Additionally, measuring emotional state was limited.

Another experiment proposed by Lucas et al. relates the size of robots and the perception of what counts as mistreatment. In the study one confederate would either act aggressively or neutrally to a robot. When participants viewed large robots being verbally abused, it was not thought of as being mistreated. However, when the same was done to a smaller robot, participants classified it as mistreatment [8]. The authors hypothesis that people would be emotionally sympathetic to smaller robots and would recognise abuse done to them, was shown to be true.

3 METHODOLOGY

The overall design for the entire project involved conducting two different studies, as well as the creation of three interactive inflatable soft robots.

The first study involved the creation of a questionnaire in order to gauge people's reactions to inflatables of different shapes. This was important because, because we did not want the shape of the robot to influence the participants interactions in the second study. We also made sure to cover as many types of interaction as possible, by using a Likert scale to gauge participants reaction towards these types of shapes (see 4).

After building the three different sized robots, we conducted a second study where we observed participants interactions with the different sized inflatable robots. Previous studies [8] concluded that people perceived mistreatment as more severe towards small robots than larger robots which suggests they feel more sympathy for smaller robots. In our study, participants interacted with the three different sized robots one at a time and their interactions were measured using a camera and observed by human observers. They then completed a questionnaire about their perception of the three robots. The data collected was analysed and open-coding analysis was performed on the qualitative data.

4 PRELIMINARY STUDY

Our first study's primary aim was to determine which inflatable shape would exhibit the least influence in order to identify which was the most neutral. To do this we created a questionnaire and analysed the responses using statistical tests.

4.1 Participants

29 participants were gathered through convenience sampling to take part in this questionnaire. Participants ranged from ages of 18-59, consisting of approximately 50 male and female individuals. No other personal information was gathered from the participants to ensure anonymity. As the authors were all students and due to using convenience sampling, 21 of participants were also students.

4.2 Procedure

We created an online questionnaire with various questions about five different shapes and asked participants to complete it. The questionnaire was designed to include multiple sufficiently different shapes so that an ideal neutral shape could be found for the inflatables. The different shapes shown were a sphere like blob, a cylindrical blob, a shape with spiky protrusions, one with bulbous protrusions, and a pyramidal shape (Fig. 2). We determined that these five shapes provided an adequate representation of a wide range of attributes, including curved and pointed surfaces which have been shown to elicit stronger responses from sensory nerves [6].

For each of the five shapes, there were sixteen 7-level Likert scale questions. These questions were purposefully designed to be ambiguous (see A.3), as we aimed to get a general analysis of participants opinions and some specifics

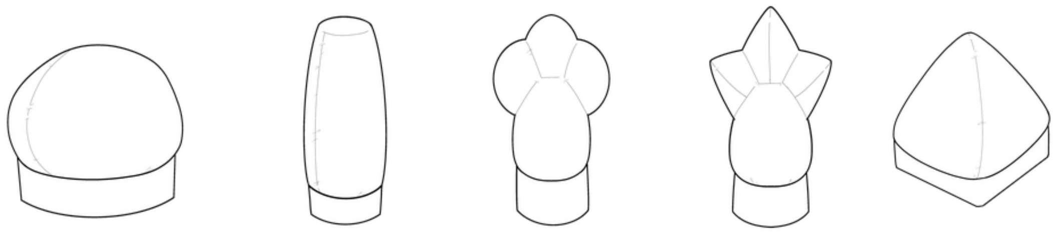


Fig. 2. Different inflatables shown in the questionnaire.

on why they felt that way. We performed a Friedman test and Nemeyi post-hoc tests to then determine which of the shapes were most neutral (see 4.4). This was then used to shape the design of our final inflatable robot.

We designed the questionnaire to be less than 20 minutes to keep the interest of the participants intact [12]. Therefore, the average completion time of the 29 participants was around 10 minutes. The questionnaire included a short description with the rough time estimate as well as a short definition of what constituted an 'inflatable' robot. Additionally, we asked them to imagine the shapes as real inflatable objects roughly the size of a human, and urged them to answer the questions as if the inflatables were physically in front of them.

4.3 Study Design

The independent variable in this study was the robot shape, with five levels corresponding to the different shapes. The dependent variable was the perceived influence or neutrality of the robot shape, as measured by the responses of the participants to the questionnaire using a Likert Scale, but also including a subjective ranking and reasoning behind those choices. This design allowed us to directly compare the neutrality ratings of the five robot shapes within the same group of participants and identify the robot shape with the least influence.

4.4 Results

A Friedman test was conducted to determine the statistical differences in the neutrality of five distinct shapes. The Friedman test, performed at a significance level of `0.05`, revealed that at least one of the shapes was statistically significant, yielding a p-value of `0.008`. To determine the exact population, a Nemenyi post-hoc test was then employed exploring pairwise comparisons between the shapes 3. The analysis yielded significant differences in the pairwise comparison matrix, particularly for the cylinder shape, which demonstrated a p-value of less than `0.05` when compared to almost all other shapes.

The average mean rank of the shapes were then plotted 4, illustrating that the cylinder had the closest average rank to four on the Likert scale, indicative of neutrality. Interestingly, no shapes yielded a mean rank greater than four (neutral), meaning questions were likely biased towards the 'Disagree' side. The combined results of the Friedman test, Nemenyi post-hoc test, and mean rank bar chart provided enough statistically significant evidence that the cylinder shape was perceived as the most neutral among the tested shapes. Further investigation may explore more sophisticated designed questions contributing to this perception of neutrality or the reason why participants had a proclivity towards Disagree sided answers. As illustrated in 3, the bulbous, spikey and pyramid shapes all produced similar responses, hence their small pairwise differences for their p-value demonstrating how devices developed in the same structure as

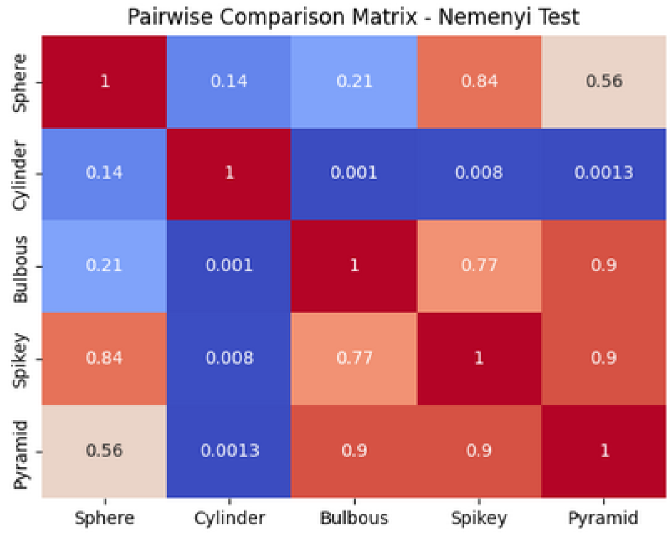


Fig. 3. Pairwise Comparison Matrix for the different Study 1 Inflatables.

these shapes could elicit similar reactions, as opposed to the cylinder and sphere, which had much stronger reactions hence their large differences in p-values. Ultimately, we decided to proceed with a cylindrical shape for our inflatables.

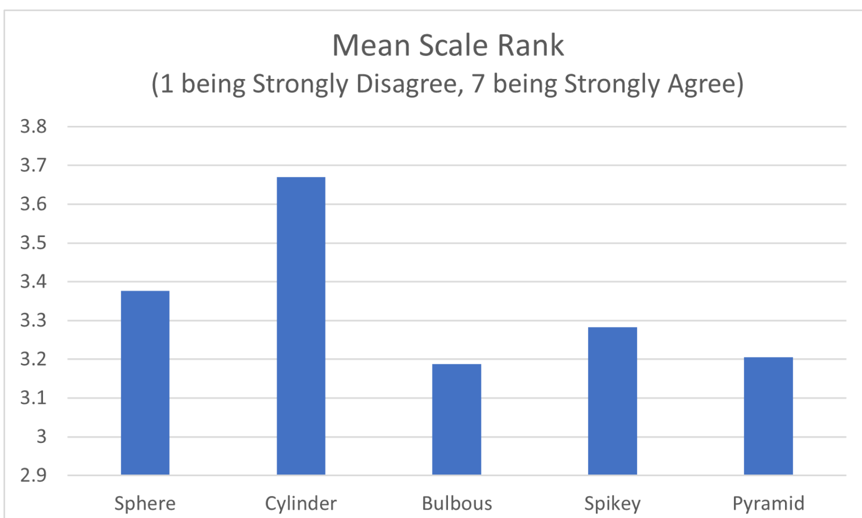


Fig. 4. Mean rank results for the study 1 inflatables.

5 FINAL INFLATABLE ROBOT DESIGN

In this section, we will describe the steps we took to construct the three inflatable robots. We will begin by discussing the design of the robot, then the method for touch detection, touch classification and responsive light feedback.

5.1 Design

The design of the final three inflatables evolved based on the information we gathered in the first study where we determined the most neutral shape - a cylinder. We decided to make all three inflatables the same shape to have more accurate and less biased results. So that the only variable that changed was the size and other factors were controlled.

We decided to have the one of the inflatables be smaller than a person, another the size of an average individual, and the last larger than a person. This resulted in the medium inflatable cylinder having a diameter of 45cm and a height of 175cm, the small inflatable a height of 75 cm and a diameter of 40 cm, and the large one a height of 235 cm and 60 cm in diameter. The sizes were picked so that there would be no ambiguity in how each inflatable would be perceived, meaning the small inflatable would be seen as small and the large inflatables would clearly be perceived as being much larger than a person. The purpose of the human sized inflatable was to act as a baseline for comparison with the larger and smaller inflatables. As for the material, we used nylon fabric also used by Air Giants [3].

A wooden circular-base was laser cut and attached using heavy duty velcro to the inside of the inflatable with weights placed on top. This ensured the inflatables were secure and would not tip over when users interacted with them. Then the appropriate air blower was attached to the walls of the inflatable also using velcro.

The specifications of each blower were:

(1) Smaller blower spec: 38 CFM and 1.9 inH2O
(2) Two Medium blowers spec: 60CFM and 1.6 inH2O
(3) Big blower spec: 324CFM and 7 inH20

Fig. 5. The three different inflatables along with the blowers used.

5.2 Touch Detection

To detect interactions, we placed an Xbox 360 Kinect camera inside pointing upwards to capture the inflatables interior. We made use of the camera's infrared stream, so any changes in visible light would be ignored. This is crucial due to the presence of RGB LED lights inside the inflatables.

Fig. 6. The Kinect camera inside an inflatable

Using Python3 with the libfreenect package, we created a program for the camera that first captured an initial image of the inside of the inflatable. Then by using "frame differencing", any subsequent movement exceeding a specified threshold would indicate an interaction with the inflatable and images would be captured.

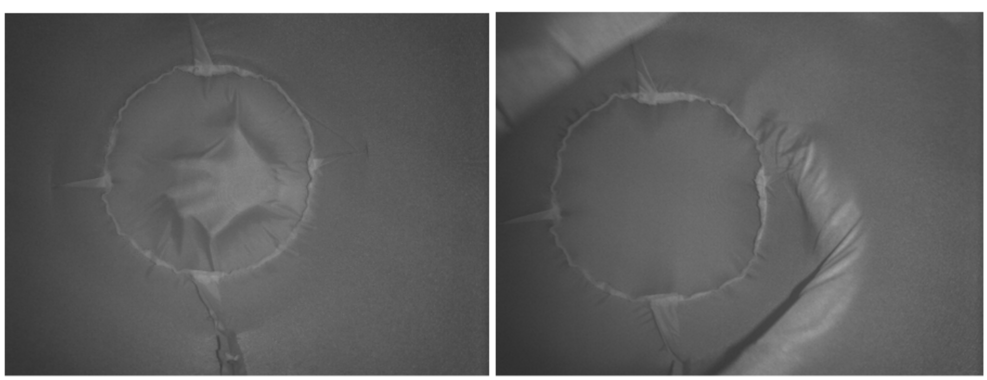


Fig. 7. A Slap (left) and a Hug (right) detected by the camera

Other than the threshold, there was another condition we needed to consider. Notably, it was not certain the inflatables would return or reinflate to its initial position. As a result, the "resting" position could differ over time, potentially exceeding the difference threshold and causing the program to erroneously perceive constant movement. To resolve this, we regularly recaptured the initial image when the inflatable was resting.

5.3 Touch Classification

In the paper proposed by Hu et al. [4], a method was developed to capture shadows created by touching or hovering over a robot's translucent skin, enabling the detection of touch positions and social gestures. It is described that this approach can turn robots with translucent skins into touch-sensitive agents with the simple addition of a camera and computer vision software. In a similar fashion, we captured the interaction with the inflatables using the aforementioned "frame differencing" as well as a custom algorithm that measures optical flow. When the optical flow was above a certain threshold, the camera captured and stored the current frames showing the interaction performed.

Subsequently, we made use of a popular deep learning technique known as Transfer Learning which enables us to use an existing deep learning model trained on a large diverse dataset, and adapt it for our specific task. The model we used was a state of the art convolutional neural network known as DenseNet201 which was pre-trained on the ImageNet dataset. ImageNet is a very popular dataset comprising over 14,000,000 images belonging to over 20,000 categories. We replaced the fully connected layers at the top of the model with our own, which enabled us to create a classifier that categorised the inputs to the classes of our choosing. These classes were the five types of interactions: touch, poke, hug, slap and punch.

As our images vary significantly from the typical images found in ImageNet, we encountered better performance when fine-tuning the model. We unfroze the last 20 layers of the model, allowing their weights to be updated during the training process and enabling the model to become more suited for our task. Our training set consisted of 1300 images and labels, of which 20% were used for validation. To encourage robustness in our classifier, we made use of various data augmentation techniques such as cropping, zooming and flipping to artificially diversify our data. Training the model over 40 epochs enabled us to achieve a classification accuracy of 80% on the validation set.

5.4 Responsive Lighting

We decided to add responsive light feedback to encourage users to interact with the inflatable. For this we used an RGB LED ring light for the smallest inflatable and a DMX stage light for the 2 larger inflatables. The lights were placed inside on the base and pointed upwards, thus ensuring even light diffusion throughout the inflatable.

An Arduino Uno was connected to the lights and programmed to manipulate them to light up when the camera captures an interaction. The colour of the light was initially blue, and then transitions towards red with each interaction. The rate of transitioning per interaction was dependent on the amount of force applied.

Fig. 8. The Small Inflatable lit up

Fig. 9. The final construction of the inflatables. The RGB ring light should be replaced with a DMX stage light for the medium and large inflatables

6 DESIGN ITERATIONS

There were some preliminary designs we considered and even implemented before building our final robots. In this section, each of the previous design iterations will be discussed along with what we learned from them.

6.1 Snowman prototype

To develop a functional prototype quickly so that we could test out design ideas, we opted to modify an existing snowman inflatable (Fig. 10a). We restructured the inflatable by cutting and sewing it into a cylindrical shape 10c, and then attached a round tin at the base using heavy duty Velcro. We quickly realised we needed something heavy at the base to prevent any undesired movements; therefore a tin was filled with a heavy object and placed inside. To complete the structural modification, we replaced the snowman's original air blower with a more powerful alternative that provided greater airflow and pressure (38 CFM and 1.9 inH2O respectively). This modification enhanced the stiffness of the inflatables and enabled it to reinflate much quicker after being compressed or manipulated.

6.2 Touch Detection Ideas

Our initial approach to touch detection centred around monitoring the changes in air pressure to identify an interaction. Inside the inflatable, we assumed there would be a somewhat constant value for air pressure. Then in theory, this value would change as users interacted with the inflatables, so we would be able to measure differences and classify the types of touch. We used a small air pressure (Fig. 11) attached to an Arduino Uno and tested this out. Unfortunately, we found that although this worked somewhat well for detecting touches with a lot of force, it was insufficiently sensitive for detecting soft touches as the air pressure was unaffected.

Our second idea was to create a fabric pressure sensor using conductive thread and Velostat using a tutorial from Instructables [10]. First we sandwiched a piece of Velostat between two layers of fabric which has conductive thread

(a) Original snowman inflatable

(b) Snowman inflatable modified by removing its head and arms

(c) Final snowman inflatable appearance

Fig. 10. Snowman Prototype

Fig. 11. Air pressure sensor used originally for touch detection

sewn into them. We created two of these then sewed them together making sure to leave two strands of conductive thread on either end (Fig. 12). These hanging strands would then be connected to an Arduino Uno. Unfortunately, this method proved infeasible due to the amount of materials we would've needed for the three inflatables. There were also issues with the resistance of the conductive threads as the sizes of inflatables increased.

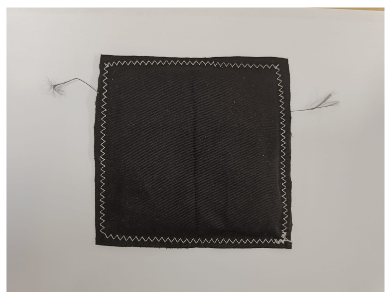


Fig. 12. Pressure sensor made from Velostat and conductive thread for touch detection

7 STRUCTURAL OBSERVATIONAL STUDY

For this study, our primary aim was to observe how participants interacted with our three different sized inflatable robots. This constituted a structured observation study followed by a questionnaire.

7.1 Participants

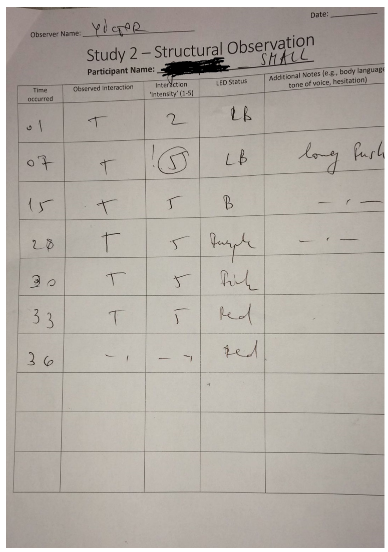
Nine participants were invited to take part in our study. Of the participants, five were male, four were female and all were university students. We recruited the participants using convenience sampling. Out of the nine participants, four also did the first study.

7.2 Procedure

The observation study took place in a dimly lit room, allowing for the fitted LED-lights to encourage interaction with the inflatables. Participants were given a task to make the robot light up red. Initially when interacting with it, the robot lit up blue. During the study, each participant interacted with the three robots individually to avoid participants influencing each other's actions. As the participants interacted with it, the colour of the light gradually turned to red. This task was designed to incentives participants to interact with the inflatables as much as possible while treating all interactions equally. This task was the same for every robot. Once a participant thought they had completed the task, they notified the observer. We expected the task to take 3 minutes to complete and so we allocated 5 minute slots per participant.

Due to technical limitations, we were only able to have two out of three robots setup at the same time. Therefore, we conducted the study in two sessions, the first session involved both the medium and small inflatables, and the second session was just the large one. In the first session, participants interacted with either the small or medium inflatable first and then the other second. Then in the second session, all participants interacted with the large inflatable.

We monitored the interactions with human observers and the camera system inside the robot. The camera system captured images of the interactions which were then classified into types of touches as described in the Final Design Section 5. However, due to the limitations of the system, we decided to fill in the gaps of interaction capturing with human observers. Observers were present for each task and filled out sheets to document what type of interactions took place at a given time (Fig. 13). This allowed us to document more types of interactions on top of the four types the camera system could identify.



Date: ____

Observer Name: [illegible]

Study 2 – Structural Observation

Participant Name: [redacted] SMALL

Time occurred	Observed Interaction	Interaction 'Intensity' (1-5)	LED Status	Additional Notes (e.g., body language, tone of voice, hesitation)
01	T	2	LB	
07	T	5	LB	long push
15	T	5	B	— " —
28	T	5	Purple	— " —
30	T	5	Pink	
33	T	5	Red	
36	— " —	— " —	Red	

Fig. 13. Example of an observation sheet that was filled out.

Once participants finished interacting with all three robots, they were given a questionnaire. The aim of the questionnaire was to gain insight into participants' perception of the interactions and how they approached completing the task.

7.3 Classification and Observation

We collected two types of data during the experiment, both detailing the type of interaction a participant performed on an inflatable. The camera captured images of an interaction which were then classified into four categories; touch/poke, hug, slap and punch. During each task, human observers noted down interactions with the robots (either the four mentioned before or any other), the time at which they took place and the assumed intensity of the interaction. To analyse the collected data, we combined the two sets into one (see 7.2).

We noted the time that each task took to complete as written down by the observers. The human observers struggled to keep up with the speed of repeated interaction, so they often noted down "repeated interaction" instead of a specific number of interactions. We filled in these gaps with the data from the classifier, as it provided the exact numbers. If a participant switched from one type of interaction to a different type of interaction in quick succession, the cameras did not always capture it. This was where the information written by the human observers was used to fill in the limitations of the camera capturing. We also noticed that the classifier would classify soft strokes as slaps. This was corrected with the information from the human observation.

When combining the collected data, most of the interaction type and amount matched between the two sets. In most cases, the data did not match if one of the data collection methods skipped an interaction, so the whole sequence was off by one data point. If there was a complete mismatch between data, we favoured the interaction captured by the cameras.

On average, the participants interacted with the small inflatable for 52 seconds before they completed the task. It took them 82 seconds to finish the task with the medium sized robot. When grouped by order of interaction, on average, the participants interacted with their first inflatable the longest, at 69 seconds. This decreased to 65 seconds for their second interaction and they interacted with the last (large) inflatable the shortest - 60 seconds on average.

Human monitoring allowed for more varied interactions to be observed. When interacting with the small robot, we observed that all participants crouched down or sat down next to it. One participant, P4, even tried to talk to the medium size inflatable by asking it to turn red.

We analysed the data in two ways, by comparing the values - the different interaction types, and by analysing the trends of interactions between each inflatable size. The participants interacted with each inflatable on average 7 to 8 times per task completion.

Out of the four interaction types we observed, we classified hugging and touching as softer, gentler interactions, and slapping and punching as rougher interactions. We added each interaction type per inflatable to compare the values between the three sizes. In total, the large inflatable had the most rough interactions, 47, which is significantly more compared to the number of soft interactions, 26. In contrast, participants interacted with the medium robot in a softer way, more often, totalled to 36, than in a rougher way, which summed up to 31. Surprisingly, according to the collected data, the small inflatable was interacted with in a rougher way 46 times and only 19 times in a softer way. Upon further inspection, we discovered that the classifier would identify soft strokes as slaps. To rectify this we favoured the observers documented interactions instead.

Additionally, we analysed the trends of interactions and participant approaches to solving the tasks. P7 started with the small inflatable and at first softly touched and poked the robot. As she discovered that punching was the interaction that worked the fastest in turning the robot red, she applied this strategy to the other inflatables as well. She finished the tasks in record times by repeatedly punching the medium and large inflatables. Similarly, P1 approached solving the tasks by switching between hugging and punching the medium and large inflatable. She applied this approach after hugging the small inflatable until it turned red. Three out of the four participants, P2, P4, P6, who started with the medium inflatable, finished their first task with a rougher interaction, either slapping or punching. However, they did not start their interaction with the small inflatable with this type of interaction. On the contrary, they started their interaction with the small inflatable by either touching/poking or hugging it. Only P8, who first interacted with the medium robot, started and finished with slapping both the medium and small inflatables. Additionally, P8 was the only participant who hugged the large inflatable, but not the others. Three participants (P2, P3, P6) hugged either the small or medium sized inflatable but did not attempt to hug the large robot.

SMALL INFLATABLE

P1	P2	P3	P4	P5	P6	P7	P8	P9
S	T	S	T	T	T	T	S	T
S	S	S	S	P	T	3	S	P
S	H	H	T	T	T	P	S	P
S	H	S	H	S	T	T	S	S
S	H	H	T	S	H	T	S	S
H	H	S	S	S	S	P	S	S
H		S	S	S	S	P		S
H				S				S

MEDIUM INFLATABLE

P1	P2	P3	P4	P5	P6	P7	P8	P9
H	S	T	H	T	T	P	S	T
T	P	H	H	S	H	P	S	T
P	S	T	T	H	T	P	S	T
H	P	H	T	P	T	P	S	T
P	S	T	T	S	T	P	S	T
H			P		P	P	S	T
T			T		T	P		T
P			P		H	P		
H			P		P			
T			P		P			
P			P		H			
					S			

LARGE INFLATABLE

P1	P2	P3	P4	P5	P6	P7	P8	P9
T	P	T	T	S	T	P	H	T
H	P	T	H	P	T	P	S	T
P	P	P	H	S	T	P	P	T
H	P	P	S	H	P	P	P	P
P	P	P	S	S	P	P	S	P
H	S	P	S	S	P	P	S	S
T	S		S	S	P	P	S	
P	S					P		
H	S							
P	S							
H								
P								
P								
S								

Fig. 14. Tables of the types of interaction each participants had with each robot. Participants starting with a particular robot size are highlighted in red. T stands for touch-poke, S for slap, P for punch and H for hugs.

7.4 Questionnaire Quantitative Results

The participants were asked 5 different Likert style questions based on positive interaction with the three different robots (see A). Conducting a Friedman test with significance level `0.05`, we achieve a p-value of `0.04`. Performing a post-hoc Nemenyi test we find out that the medium robot's group yielded a p-value of `0.07`. Though this did not reach the significance threshold, we can see that participants tended to act more positively towards the medium sized robot. However, since its p-value was not within the significance threshold we cannot statistically conclude that participants acted more positively overall to a particular robot size based on the questions asked.

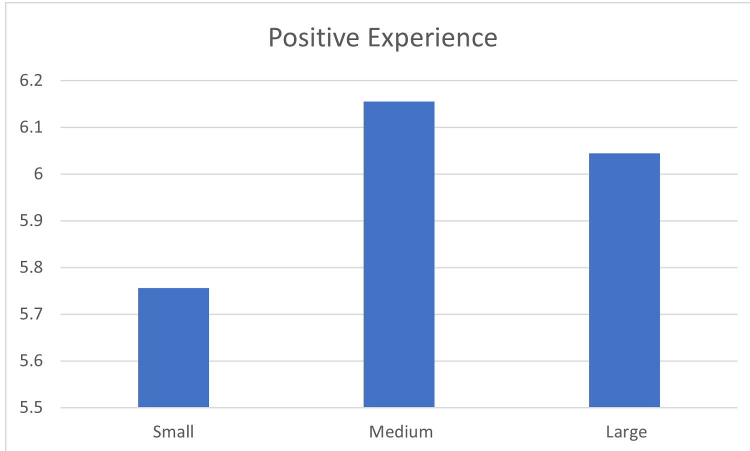
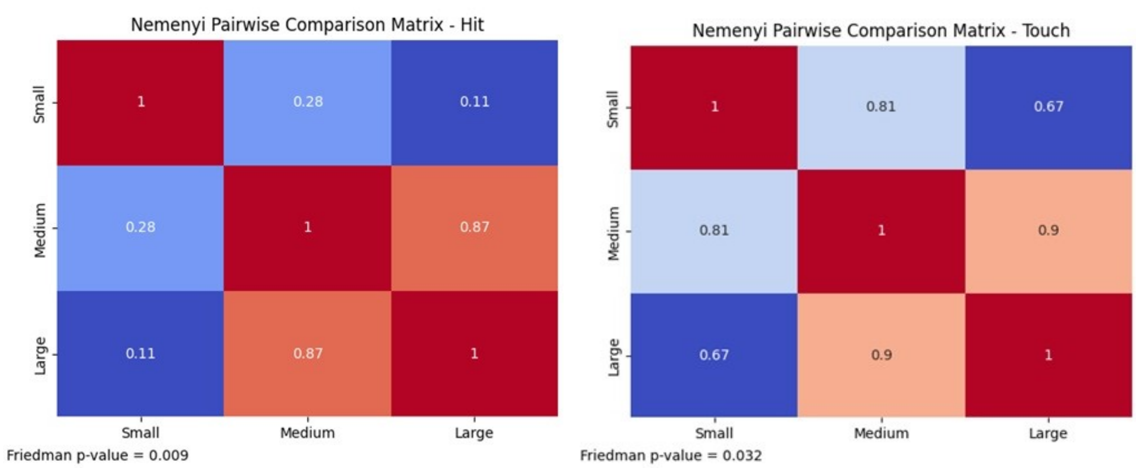


Fig. 15. Average participant judgement based on whether they had a positive experience with the robot, based on their size.

Furthermore, we asked participants how likely they were to perform 3 different interactions on the robot: touching, hugging, and hitting it in a Likert style format. Again, performing a Friedman test separately on each interaction with a significance level of `0.05`, we yielded p-values of `0.01` (Fig. 16a) and `0.03` (Fig. 16b) for hitting and touching the robot respectively. As for statistically testing the hugging interaction for the robot, we yielded a p-value of `0.2`, therefore we cannot statistically conclude that the size affected how likely participants were to hug the robot, with a significance level of `0.05`.



(a) Hit (b) Touch

Fig. 16. Nemenyi comparison matrices for how likely participants wanted to hit/touch the inflatable, for small, medium and large sizes.

Performing further post-hoc tests on touching and hitting the robot, we can see firstly that there was no significant difference when it came to how likely participants were to touch the robot. The post-hoc test for hitting the robot resulted in a small difference particularly in the small robot size group, meaning participants felt much less likely to hit the small robot and more likely to hit the large robot. Illustrated in Fig. 17, the average likelihood of the participant wanting to hit the robot substantiates the statistical test. Though the post-hoc test did not statistically determine whether the small robot was less likely to hit with a significance level of `0.05`, we can consider that in a larger scale test, it is likely that participants would overall be less likely to hit a robot of a smaller size.

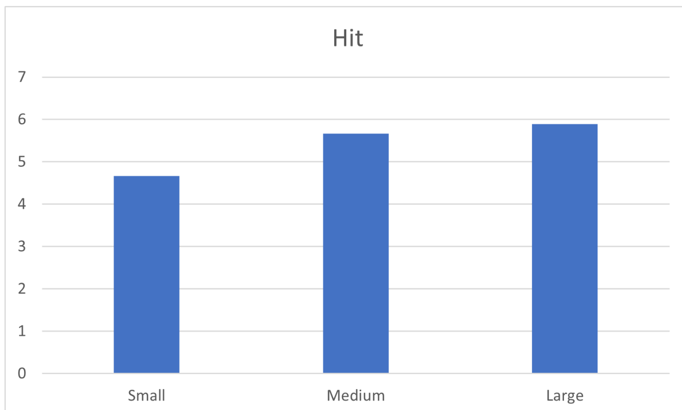


Fig. 17. Average likelihood of participants wanting to "hit" the robot based on their size.

7.5 Questionnaire Qualitative Results

In order to analyse the questionnaire responses the participants gave we used open coding methods. For each response in each section we assigned codes in order to capture the nature of the responses. We did multiple passes of the responses and the codes were compared with a second judge for accuracy, getting a Cohen's Kappa of `0.63`, indicating substantial agreement. Codes were split into two different higher order categories, either "Perception" if the participant's response was related to how they perceived and thought about the inflatables, and "Decision" for if their response concerned how they interacted with the inflatables and what motivated their interactions. Based on the coding analysis we would often see codes such as "gentle small", "cute small" but also "hard to interact with small" for responses related to the small inflatable. For the medium inflatable codes such as "harsh medium", "more force medium", and "easy to interact medium" were prevalent and for the large inflatable codes such as "hit large", "more force large", and "resistant to abuse", where frequently present. Codes related to slapping, hitting, and aggression were also more frequently present in responses on the medium and large inflatable. The code "size" indicating that the size of the inflatable was the underlying factor in people's interactions was also present throughout. These codes indicate that not only did interactions differ based on sizes but also that participants felt strongly about wanting to hit and slap the medium and large inflatable more than the small inflatable, for which participants often stated they wanted to be gentle with and that they thought it was cute.

8 CRITICAL EVALUATION

In this section, we conduct an evaluation of our methodology to identify successful aspects, as well as any limitations that may have influenced our results.

8.1 Construction

There were several concerns with the construction of the inflatables that could have influenced participants' interactions. Firstly, due to the varying specifications of the blowers used, the three inflatables did not feel similar in touch. Specifically, the largest inflatable felt the most rigid, as the blower used was originally intended for a bouncy castle and produced a pressure of 7 inH20, which was three times the pressure of the smallest blower (1.9 inH2O). Additionally, the largest blower also produced the most noise. These discrepancies could have caused participants to unconsciously modify their interactions with the inflatables. For example, they may have treated the smaller inflatable more gently as it was softer, or treated the largest robot more rougher as it was the noisiest. Ideally, all three inflatables should have felt the same and produced the same level of noise, ensuring that confounding variables were minimised.

8.2 Touch Classification

Unfortunately, whilst our touch classification method worked well (see 5.3), it had some limitations. As we used optical flow to detect touches, sometimes successive touches that happened rapidly one another were not accurately captured. Furthermore, the types of touches were limited to only poke, slap, punch and hug. In a real-life scenario, individuals will likely perform additional types of interaction which are unconfined from these five categories. Therefore, to ensure robustness in our results, we had to use human observers alongside the automated classification mechanism and compare results afterwards.

In terms of capturing subsequent interactions, we could optimise the optical flow algorithm to be more sensitive; however, this has its own limitations as we wanted to avoid capturing frames that do not include the interaction. A possible alternative is to not use optical flow as a metric for touch detection. Instead, we could use image processing techniques to detect contours around a person's hand or arm, as done in [4]. We were not able to implement this in our study due to the lighting conditions within the inflatable. For the medium and large inflatables, our IR illuminator was not strong enough to project the same intensity of light throughout the inflatables' interior. A stronger IR illuminator would enable us to capture well-lit images and allow for better contour detection for the interactions.

For classification itself, we could increase our range of categories by adding interactions such as petting, stroking, squeezing and kicking. This would allow us to capture a wider range of interactions and give us better insight to how participants interact with the inflatable. Additionally, while we were able to achieve an 80% evaluation accuracy on our classifier, it is entirely possible that this could be increased by using more data. The DenseNet201 model's training can also be further optimised by tuning hyper-parameters such as the batch size, learning rate and the choice of optimiser used.

Finally, given appropriate ethics approval, video recording participants interacting with inflatables would have been ideal. This would remove the need for human observers as well as reduce human error inaccuracies, as the video can be re-watched multiple times at various speeds. This would also enable us to combine verbal cues from participants with the results, providing a deeper understanding of the participants emotions during the interaction.

8.3 Study Evaluation

Our second study was limited in the sense that it did not cover a diverse demographic in representing an entire population. Whilst our preliminary questionnaire gathered 29 participants in total, each with a diverse age range, our second experiment only involved 9 participants. These participants were gathered from convenience sampling, all of which being within the 18-25 age range; impacting how the sampling affected the results, as it did not represent an accurate diverse population. A larger scale test could be performed in a similar style, encompassing a much more diverse population leading to more significant and conclusive results. Performing the study in a similar style would require a much greater time commitment, or more observers. Alternatively, the style of study could be altered slightly to accommodate a greater population size by performing shorter observation sessions and relying entirely on the automatic interaction sensing (eliminating the need for human observers).

Another major consideration about how the study was run was that the large inflatable was always shown last. In order to mitigate bias resulting from the order, half of the participants had the small inflatable first and half had the medium first. However, for the large inflatable, this was not possible as we needed to set it up after the small and medium ones were taken down. Because of this, no participant interacted with the large inflatable first, meaning that the large inflatable being the last one they needed to interact with could have affected their interactions. Given that participants had already gained experience completing the task with the other two inflatables, some of their interactions for the last one could have been biased on this instead of the inflatable's height. If the study were to be run again this is something that we would change, making sure participants had different first and last inflatables instead of just having a different last inflatable.

9 FUTURE WORK

9.1 Multi-modal interaction

Incorporating new sensing modalities, such as audio, touch, and gesture recognition with giant inflatable robots, could potentially enhance users' interactions. For example, if the robot could detect when a user interacts with it through touch or gesture, it could respond with different audio sounds to provide a more engaging and interactive experience. Future research could explore the potential benefits of multi-modal interactions with inflatable robots and how they could be integrated into the design of robots to improve user engagement and satisfaction.

9.2 Therapeutic Applications

Considering that participants were more reluctant to hit the small robot, examination of robot size in therapeutic applications could be looked into, with the aim of providing optimal support to those with autism spectrum disorder, anxiety, or other mental health conditions. Alongside the size of the robot, non-threatening appearances and interactive lighting could also provide a more calming and engaging experience for those individuals, potentially giving them greater (or more accessible) emotional support. Furthermore, therapists or other health professionals could adjust certain attributes of the inflatable robot in order to create a more personalised rehabilitation experience for patients. For instance, facial cues could also be incorporated into the inflatable robot, to express body language or other facial expressions in order to model and teach appropriate social cues and emotional responses.

The treatment of anxiety and stress-related disorders could specifically be explored by leveraging a gentle, rhythmic motion of inflatable robots by creating soothing, sensory experiences that promote relaxation and stress reduction. For example, an inflatable robot designed as a wearable device could provide gentle pressure or enveloping sensations, mimicking the calming effects of deep pressure therapy or weighted blankets. This could be particularly beneficial for individuals with sensory processing disorders or those experiencing heightened anxiety.

9.3 Longitudinal study

Participants interacted with the inflatables for the first time when we conducted the experiment. However, an important study subject could be to test how their interaction would change over time. We could have a group of individuals interacting with the inflatables over a longer period of time, possibly in the order of days or even months, and monitor their interactions. Would the interaction of the individuals change over time once they get to know the robots better? Would the individuals lose excitement for interacting with them? Would individuals start expressing sentiments about the robots?

9.4 Adaptive Behaviour

We can make use of Machine Learning to develop algorithms to adapt the inflatable's behaviour based on the mood of the person interacting with it. Instead of using simple light responses we can introduce actuation in the inflatable so it gives physical feedback to the human interacting with it.

Investigating adaptive behaviour can help researchers understand how humans respond to robots that can change their behaviour based on context or user preferences, which can provide deeper insights into designing more effective and user-friendly robotic systems.

Studying adaptive behaviour in robots can also drive advancements in machine learning and artificial intelligence, as these robots would learn from their interactions with users and the surrounding environment. This can lead to the development of more sophisticated algorithms and techniques for robot learning, decision-making, and adaptation.

10 CONCLUSION

In this research we studied human behaviour towards soft robots of different sizes. Our studies concluded that this behaviour does differ as the size of the robot increases. During the first phase of our research, we conducted preliminary study to identify the most neutral shape that would not influence interaction, which we found to be a cylinder. This then informed the design of our inflatables.

Afterwards, we observed and monitored individuals' interactions with three different sized inflatable robots and analysed the results. Participants were given a task to turn the robots red. We captured and documented the interactions using cameras and human observers. From the analysis of this data, we concluded that interactions did differ with size. Specifically, we found that the number of soft interactions decreased with size, whilst the number of harsh interactions increased. Additionally, this was confirmed with the qualitative data collected which indicated that people had gentler dispositions to the small inflatable and more aggressive tendencies towards the medium and large. Therefore, this indicates that human behaviour towards inflatable soft robots is biased by size.

ACKNOWLEDGMENTS

We would like to thank all of the participants in both studies as well as Prof. Anne Roudaut, and Dr. Peter Bennett who provided their exceptional guidance. We would also like to specially thank Immi Biswas for her unwavering support and supervision during the project.

REFERENCES

[1] Zachary Carlson, Louise Lemmon, MacCallister Higgins, David Frank, and David Feil-Seifer. 2017. This robot stinks! Differences between perceived mistreatment of robot and computer partners. arXiv:1711.00561 [cs.RO]

[2] Susan Fussell, Sara Kiesler, Leslie Setlock, and Victoria Yew. 2008. How people anthropomorphize robots. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction, HRI 2008, March 12-15, 2008.* ACM, Amsterdam, The Netherlands, 145–152. https://doi.org/10.1145/1349822.1349842

[3] AIR GIANTS. 2020. . Air Giants Limited. https://www.airgiants.co.uk/

[4] Yuhan Hu, Sara Maria Bejarano, and Guy Hoffman. 2020. ShadowSense. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4 (12 2020), 1–24. https://doi.org/10.1145/3432202

[5] JULIAN JAGTENBERG. 2019. Startup Profiles. https://robotunion.eu/wp-content/uploads/2019/12/Somnox.pdf

[6] Per Jenmalm, Ingvars Birznieks, Antony Goodwin, and Roland Johansson. 2003. Influence of object shape on responses of human tactile afferents under conditions characteristic of manipulation. *The European journal of neuroscience* 18 (08 2003), 164–76. https://doi.org/10.1046/j.1460-9568.2003.02721.x

[7] Talia Konkle and Aude Oliva. 2012. A Real-World Size Organization of Object Responses in Occipitotemporal Cortex. *Neuron* 74 (06 2012), 1114–24. https://doi.org/10.1016/j.neuron.2012.04.036

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

[9] Serena Marchesi, Davide De Tommaso, Jairo Perez-Osorio, and Agnieszka Wykowska. 2022. Belief in sharing the same phenomenological experience increases the likelihood of adopting the intentional stance toward a humanoid robot. *Technology, Mind, and Behavior* 3 (06 2022), 11–11. https://doi.org/10.1037/tmb0000072

[10] Pluseahttp://www.plusea.at. 2009. Conductive Thread Pressure Sensor. https://www.instructables.com/Conductive-Thread-Pressure-Sensor/

[11] Byron Reeves, Annie Lang, Eun Young Kim, and Deborah Tatar. 1999. The Effects of Screen Size and Message Content on Attention and Arousal. *Media Psychology* 1 (03 1999), 49–67. https://doi.org/10.1207/s1532785xmep0101_4

[12] Hunny Sharma. 2022. How short or long should be a questionnaire for any research? Researchers dilemma in deciding the appropriate questionnaire length. *Saudi Journal of Anaesthesia* 16 (01 2022), 65. https://doi.org/10.4103/sja.sja_163_21

[13] Somnox. 2015. . Somnox. https://somnox.com/

[14] Peeraya Sripian, Muhammad Nur Adilin Mohd Anuardi, Yushun Kajihara, and Midori Sugaya. 2021. Empathetic robot evaluation through emotion estimation analysis and facial expression synchronization from biological information. *Artificial Life and Robotics* 26 (09 2021), 379–389. https://doi.org/10.1007/s10015-021-00696-w

[15] Yugin Teo. 2020. Recognition, collaboration and community: science fiction representations of robot carers in Robot & Frank, Big Hero 6 and Humans. *Medical Humanities* 47 (04 2020), 95–102. https://doi.org/10.1136/medhum-2019-011744

[16] Jessica C. Thompson, Susana Carvalho, Curtis W. Marean, and Zeresenay Alemseged. 2019. Origins of the Human Predatory Pattern: The Transition to Large-Animal Exploitation by Early Hominins. *Current Anthropology* 60 (02 2019), 1–23. https://doi.org/10.1086/701477

A APPENDIX

A.1 Teamwork Diversity and Inclusivity

As a team of six, we recognised the importance of inclusivity and equality in our project. We utilised a Trello board to assign tasks to each person every week, establishing clear goals and expectations for the project, and helping ensure everyone understood their role and what was expected of them. Some tasks were split into pairs or groups, engaging in team-building activities which fostered collaboration and unity.

We held meetings twice a week to make a conscious effort of listening to each other's perspectives and ideas. Two of our members preferred working in the morning, but others favoured working in the evening. To compromise, we tried to meet during the early afternoon to accommodate both team members preferred work environments. Furthermore, half of our team members preferred working online, and the other working in person, similarly, we split our meetings into online and in-person. Team members who preferred working online generally met to work on parts of the project which could be completed online, with the other team members meeting separately to work on parts of the project which required being in-person. We also worked together to ensure that communication was clear and transparent, avoiding any language that may have been exclusionary or discriminatory.

Overall, by implementing these inclusive practices, we were able to create a more harmonious and productive work environment. We recognised the value of each other's differences and learned to work together effectively as a team. If any issues did arise, we addressed them promptly and sought help if needed to ensure that everyone felt supported and heard.

A.2 Video Link

https://youtu.be/B9FHXXk87Ew

A.3 Study 1 Questionnaire

Participants were asked to rate each of the inflatables (sphere, cylindrical, spikey, bulbous and pyramid) using a 7-level Likert scale consisting of these 16 questions.

(1) The inflatable is complex
(2) The inflatable is sharp
(3) The inflatable is offensive
(4) The inflatable makes me feel depressed
(5) The inflatable makes me feel sad
(6) The inflatable makes me feel happy
(7) I don't like this inflatable
(8) I like this inflatable
(9) I want to touch this inflatable
(10) The inflatable makes me feel excited
(11) The inflatable is comforting
(12) I want to hug this inflatable
(13) This inflatable resembles another object
(14) I want to hit this inflatable
(15) The inflatable is simple
(16) The inflatable is round