Shiverbot Final Report

Problem Statement & Research Question

Our proposed Human-Robot Interaction experiment, “Shiverbot,” explored how humans’ empathy varies in response to tangible expression of biological functions. We placed a utility cart to the railing next to where the Pausch Bridge meets the entrance of Purnell and put our custom-built, knee-height robot on that. On the cart next to the robot was a small, non-functional heater with a hand crank attached to it. The independent variable we intended to control for was whether or not the robot shivered. The robot, which was either shivering or not shivering, called out to passersby, asking them to turn the crank in order to warm the robot up.

We were interested in finding out (1) how much time the user is willing to spend helping a robot solve a biological problem it cannot truly experience and (2) if the amount of time differs based on whether or not the user sees the robot shivering. If the amount of time users spent helping the shivering robot as opposed to the static robot differs by a statistically large enough margin and if our data sample possesses statistical validity and generalizability, we would have been able to infer that visual demonstration of bodily functions does impact empathy. We hoped that this experiment would provide greater depth in understanding how closely humans relate themselves to robots and whether humans would ignore the disjunction between robots and biological functions in favor of their innate empathy.

By posing the question, “How long will passersby manually turn a crank to ‘warm’ a robot if we imbue the robot with the human characteristic of shivering (versus not shivering)?”, we hoped to reach beyond the surface-level interactions taking place in order to better understand how visual confirmation of biological functions affects empathy.

Literature Review

Harrison, N. A. (2015). Feeling cold is contagious. *Temperature, 3*(1), 20-22.

Meng, X., Yoshida, N., & Yonezawa, T. (2015). Evaluations of involuntary cross-modal expressions on the skin of a communication robot. *2015 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI).*

In our experiment, we want to explore if a robot that shivers, a physiological response that only warm-blooded animals experience, can elicit empathy in humans that interact with it. We specifically want to test the level of empathy in humans through the amount of time they spend cranking a heat generator attached to our ShiverBot in the attempt to help the robot “warm up” or reach homeostasis. The main problem we are addressing is whether humans would feel so compelled by the sight of a shivering robot that they will go out of there way to help it, even though the robot does not even feel temperature or need the bodily function of shivering to regulate its internal conditions.

In a 2015 study “Feeling cold is contagious” conducted by scientists at the University of Sussex, 36 participants watched 8, 3-minute videos depicting actors with their right or left hand in visibly warm or cold water. During the participants’ observations of the videos, their right and left hand temperatures were measured using thermometers. Afterwards they were asked to rate how cold or how hot they thought the actors’ hands were. The hypothesis being tested was if another’s temperature change changes the observer’s own peripheral body temperature. The findings included that when one sees another’s hand in cold water, the temperature of the observer’s hand dropped significantly while there was little to no change in the temperature of observers who saw a hand in warm water. Furthermore, observers who had greater levels of self-reported empathy had greater differences in how cold or hot they thought the actor’s hands were compared to the actual temperature. The study attempted to explain their results saying that there was less response to the warm videos because they were less potent, there was less visual cue that the water was warm other than the steam at the beginning of the videos and the pink color of the actor’s hand compared to the blocks of ice that were clearly visible during the cold video. Lastly, Neuropsychiatrist Dr. Neil Harrison, the leader of the research, believes that the unconscious physiological changes that occurred helped the observers to empathize with the actors, leading to the broader idea that mimicking another person helps to create an internal model of their physiological state, which people can use to better understand their motivations and how they are feeling. For humans, this is significant because in the complex world that we live in, teamwork is crucial and would not be possible without the ability to rapidly empathize with each other and predict one another’s thoughts, feelings, and motivations. Overall, the neurobiological reasoning for these effects is still unknown but Harrison explained it has something to do with top-down influences on thermosensitive neurons within the preoptic area of the hypothalamus. This study greatly supports our experiment because it not only confirms that humans respond to seeing other cold individuals, and even more so than warm individuals, but it links this evidence, that humans seem to vicariously respond with their own temperatures dropping, quantifying empathy. We hope to further explore and test for this empathy with human to robot interactions despite not being able

to measure people's body temperature when they observe our robot, or at least we have not thought of a way of integrating such measurements into our experiment yet other than time on cranking our robot.

In a 2015 HRI study, "Evaluations of Involuntary Cross-modal Expressions on the Skin of a Communication Robot" 3 scientists from Kasai University Osaka developed a method of generating multiple involuntary expressions through a robot's skin (goose bumps, sweats, and shivers) using multiple thin rods under the skin, a vibration motor, and a water tank with a balloon. The hypothesis was that robots truly need emotional expression, such as like-human expressions, to realize more natural human-robot communication. And specifically emotional expression to reflect a robot's internal state in order to portray that it has emotions and psychological states. The study references many resources supporting the idea that non-verbal expressions and involuntary expressions are imperative in order to create natural communication with humans. More helpful to us was their approach to creating the three involuntary expressions. For imitating goosebumps, silicone was used to create a tactile similarity to human skin while a servomotor pulled pipes up and down to create spots on the surface. For sweating, water is pumped from a plastic bottle and injected air increases the water pressure, pushing water drops that represent sweat. Lastly for shivering, a vibration motor was used to communicate the robot's shivering. We will also be using a motor. The testing hypotheses were that each expression of goosebumps, sweating, body shivering is able to express the robot's emotion related to fear as physical physiological reactions. Their independent variable is either the robot displays the expression or not, which is exactly what we did as well. Their set up for the experiment was that participants sit in front of the robot within arm's reach and participants must evaluate using a 5 point scale of whether the robot seemed to feel fear to panic after a voice said "there is a ghost behind you." The findings were that all three were statistically significant and proved that humans can understand imitative expressions in robots. Overall, this study was rather helpful in seeing how others have thought of creating a robot that not only shivers, but has goosebumps, and even sweats. Seeing how they actually created the mechanisms confirms our idea of using a vibrating motor as well. We hope to expand and reference their studies because while this study supports the idea that humans respond to a robot with involuntary expressions, it is only limited to using shivering to portray fear and indoors, while we test it within the the context of temperature and outdoors.

Method

Conditions of the Study

We placed Shiverbot in front of the Purnell building at CMU. Previously we planned to place Shiverbot directly on the ground, but we ended up placing it upon the bed of a utility cart. The initial concept design for Shiverbot had its height at 2.5 feet, but we eventually scaled that down to about 1 foot due to design constraints explained in detail further in this paper. After placing Shiverbot on the bed of the utility cart, the robot stood at about 1.75 feet.

We expected users to squat or sit on the floor in order to be at eye level with the robot while cranking. Minus shivering, the robot did not move during the study. We decided to forgo chaining Shiverbot or the cart it rested on to a nearby railing because it felt unnecessary once we set up in the actual environment. For half of the study, we powered motors that caused Shiverbot to appear as if it was shivering.

We sat in Purnell at one of the tables in the lobby from an angle where we were able to clearly see the robot without being seen by potential users. As expected, no outsiders looked in the Purnell lobby as they passed by, whether from window glare or disinterest. The entrance to the bridge ended up being an ideal location for conducting this study, as a steady number of people constantly passed by but never so many as to create an overwhelming mob. This setting gave us more control in an environment that allowed for little consistency. Additionally, choosing a spot located between Gates, the technological center of CMU, and Purnell, where artistry and emotion take precedence over machines, allowed us to interact with a greater variety of users than if we had chosen a building unique to one department.

For Shiverbot's "face," we took an 8 inch tablet and covered the lower half so as to create a more square and balanced face versus a more extreme and rectangular one. The tablet displayed the same eye shape but had very simple mouth movements of opening and closing when it "talked". Using the wizard-of-oz way of implementation, we controlled this feature remotely using a wireless bluetooth keyboard to toggle back and forth between the two mouth movements. We also added in a third mouth position, a smile, to give users instant feedback when they initially begin cranking. In order to communicate with the users, we placed a phone and bluetooth speakers near Shiverbot. The phone was on a call to a different phone that we held while we ran the study, in

order to figure out what speech to send to the bluetooth speakers to talk to the user. We used the text-to-speech feature available on any Mac system to talk through said speakers.

Independent and Dependent Variables

The independent variable for our study was whether or not the robot was shivering. We measured how shivering versus lack of shivering affected the amount of time that people spent assisting Shiverbot (cranking the heat generator), the dependent variable.

Subjects

We originally hoped for 20 to 40 participants for our study. The pool of subjects consisted of anyone who crossed the foyer in front of Purnell and the Pausch Bridge. Since this location connects different parts of the CMU campus, we hoped that the demographics of our sample would be more reflective of CMU as a whole and not just one department in particular. Our pool of participants varied from CMU students, to CMU faculty, to visitors.

Since some individuals chose to simply ignore the robot, and because we chose CMU as our research location, we expected that our study may not have been representative of everyone who comes into contact with a robot. We understand that there was response bias and sampling bias in our methods because the subjects were obtained through self-selection, and because the sample was selected at CMU.

Dialogue

Contact Initiation
“Hey! Can you turn that crank, please?”
Get’s closer
“I wouldn’t ask unless I really needed the help, but can you please turn that crank?”

Backstory
Though we were not able to keep users cranking for long enough to develop our backstory, below are phrases that we initially planned to use for that purpose.

“Really sorry, my owner has been gone for a while, he said he’d be back soon.”
“Is it chilly out here or is it just me?”

Again, few users directly responded to Shiverbot’s attempts at conversation, but below are some of the user responses we anticipated receiving, and what we had planned replying with.

Possible User Responses:

- “Is that good?” -Sorry yes
- “What is this for?” - If you could just keep me warmer a little longer?
- “I’m sorry I have to go” - Really? [Contact Termination]
- “Do I just keep cranking?” -I’m sorry just a little bit longer please.

Elicit empathy

- How has your day been? Has anything exciting happened? Why why not?
- Are you looking forward to anything fun? If you can be anywhere right now where would you be?
- What do you like to do when you’re not working?
- What type of music are you into?
- Have you read any good books recently?
- Are you a cat person or a dog person?
- Coffee or tea?
- Do you cook?

Contact Termination

No users stayed long enough for us to engage in a goodbye more complicated than, “Thank you very much.” Below is what we had hoped to say upon terminating contact.

“Thank you so much! Sorry for troubling you! What’s your name by the way?
User says name
“Well, it was really nice to meet you, [Name]. Take care of yourself, okay?”

Plan & Procedure

Shiverbot actively tried to initiate contact with passing individuals by shouting, “Hey! Can you turn that crank please?” once we realized that the phrase, “Sorry! Excuse me?” did not cause any passersby to stop. We successfully avoided having any groups of people gather around Shiverbot simply due to lack of interest on the part of those passing by.

We tried to follow our planned script as closely as possible once a user would approach Shiverbot, but we were forced to improvise once we realized that none had the patience to wait for us to develop Shiverbot’s problem and backstory. Our possible user responses and intended replies to said responses were of no use to us, because users never verbally engaged with Shiverbot.

Contact termination always resulted in having Shiverbot shout, “Thank you!” at a user’s back as they retreated, rendering our planned personalized goodbye useless.

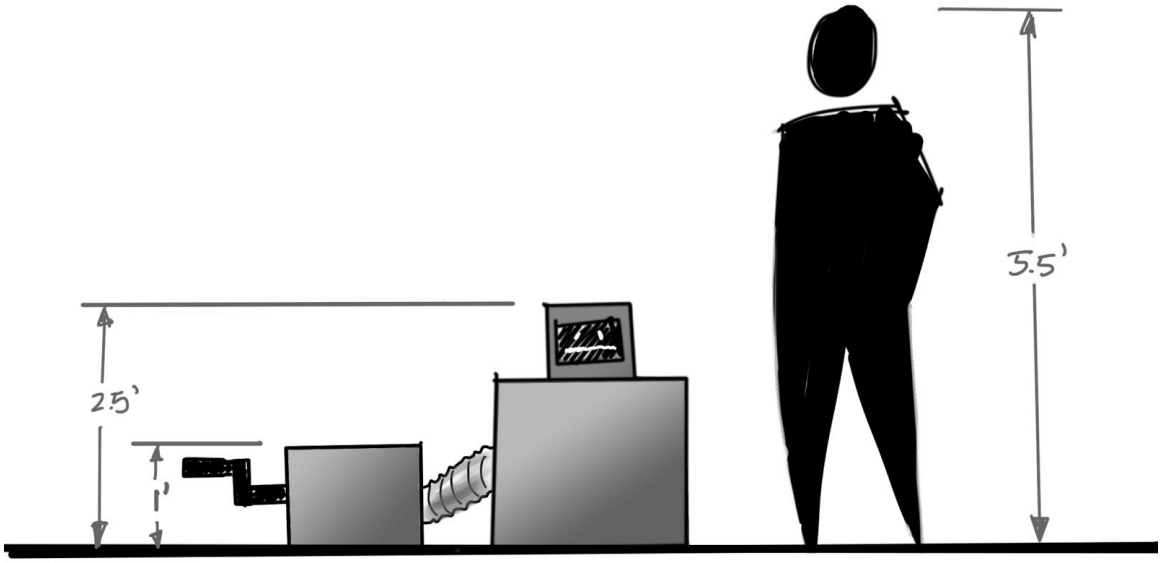
Robot Design

Initial Concept (Homework 3):

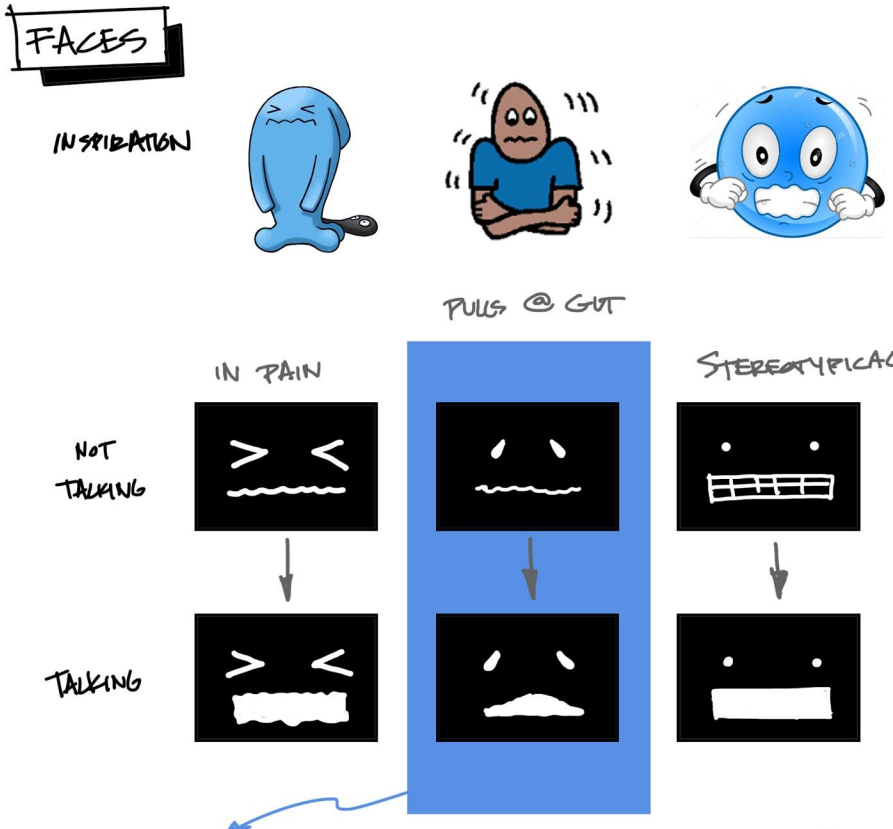
Our initial concept was to create a much larger robot (at least 2 ft in height) because of where we wanted to situate it (Purnell to Gates). We felt that a larger sized robot would be able to successfully attract attention and garner more interaction in such a busy intersection. To manufacture this robot, we chose to laser cut ⅛” acrylic to create the two internal cube shell structures (head and body) as opposed to cardboard because we thought we needed stronger material to support a motor, the tablet, and the heat duct. Next we would adhere thin sheet metal to create a higher fidelity robot and to help give the impression that it was cold if anyone touched the metal. Lastly for the heater, we thought of creating it in the same manner as the robot with a laser cut shell and sheet metal exterior, suggesting that it was part of the robot.

Rachel Nakamura, Adella Guo, Alan Qiu
Guobotics

SIZE COMPARISON



FACES



Evolving from Homework 3:

We ended up making several changes to our initial design. From our class presentation on Autonomy Design, we received feedback from our class that helped us to create a better backstory for the robot. Instead of just having a large robot on the ground at Purnell with no context, we decided to add a moving cart with cardboard boxes and a poster advertising a robotics open house. That way, our robot will seem less like it was randomly placed in the environment.

The second feedback we received was the concept of making the heater seem less of a hack job and more of an actual heater that was separate from the robot's body. This was because reducing the heater to just a sheet metal box makes it less convincing that the action of cranking helps the robot. We decided to purchase a small 9 inch tall heater from Amazon and took out the internal parts to attach the heat duct.

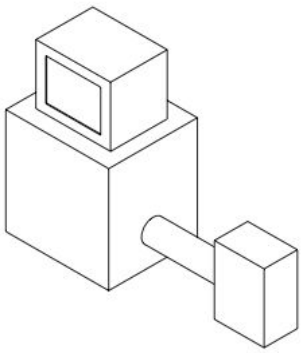
The most substantial change we made to our design was decreasing its size. There were several reasons for this. The first was when we compared the size of the 9 inch heater to the initial proposed size of our robot (pictured below). The robot was just too big in comparison to the heater. Then when we received the crank, it was much smaller than we thought it would be. Although it fit well with the small heater, it was too small compared to the robot and seemed like it wouldn't help warm the robot at all. Lastly, the main reason for making our robot smaller was because we could not figure out how to work the AC motor that we bought. Although the motor was a large enough size to make our large robot shiver, we did not know how to actually make the motor work. After consulting with an ECE friend, we opted for 4 small DC motors and decided to make the robot as small as possible. We decided to go with an 5x8 inch Dell tablet as opposed to a 7.5x9.7 inch iPad. This allowed us to substantially reduce the size of the robot. This helped us in the long run because not only was it easier to built because we did not have to use as much material but it also contributed to the cute, helpless charisma that we wanted our Shiver Bot to have.

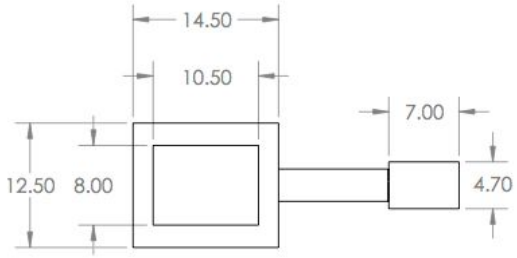
The final change we made to our design was adding a blushing happy face. We felt compelled to add this because when we began testing the feedback after someone cranks, the voice replies of gratitude were just too slow. We needed an immediate gratification for the user to feel like cranking actually did something to the robot.

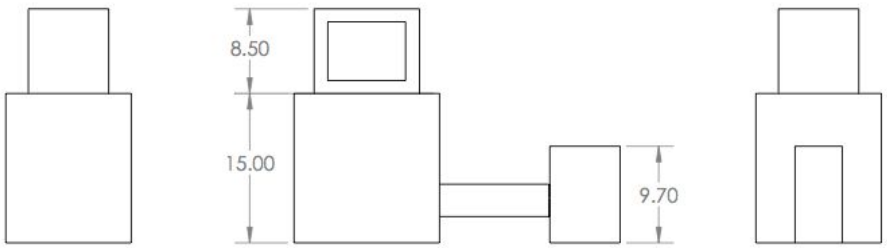
Final Design

Size Change

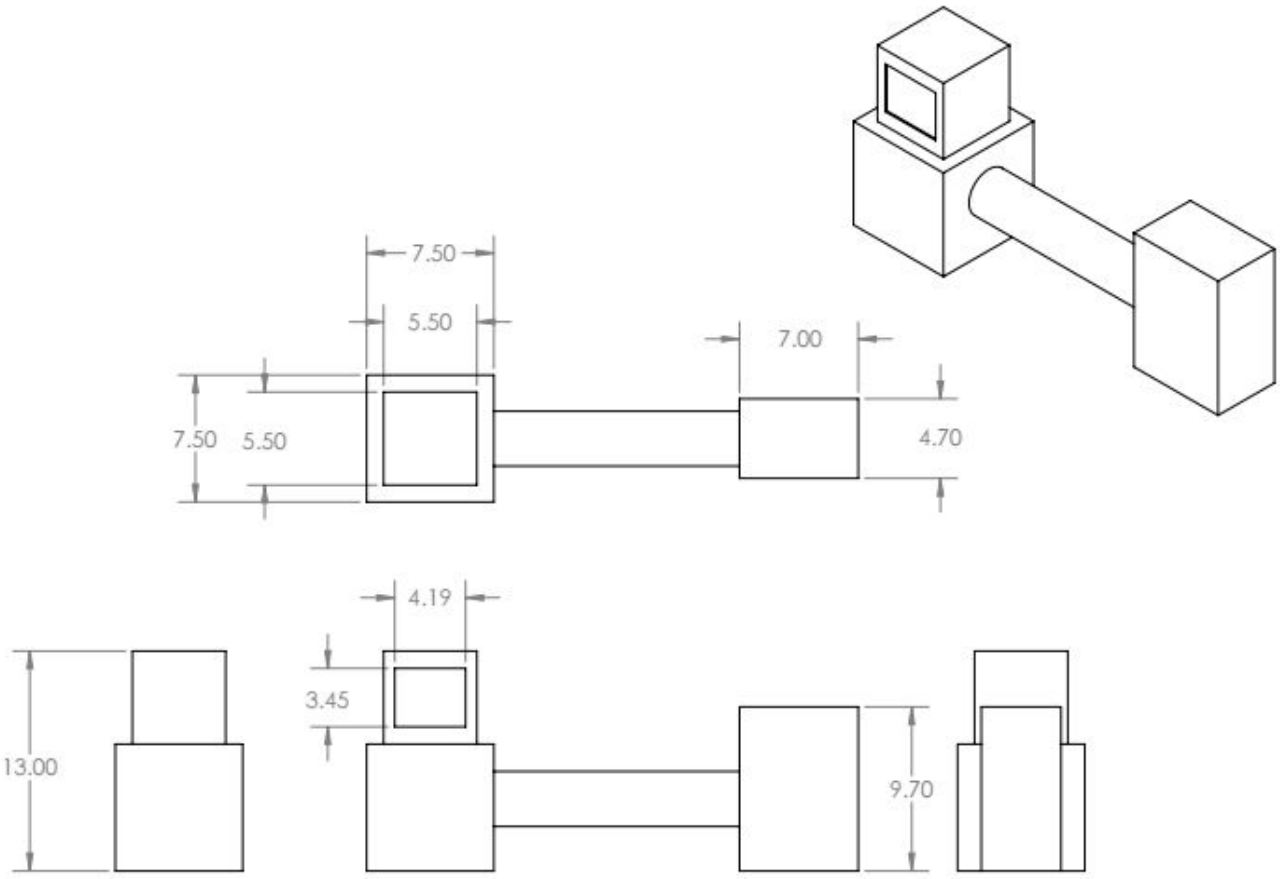
Initial Size Design:



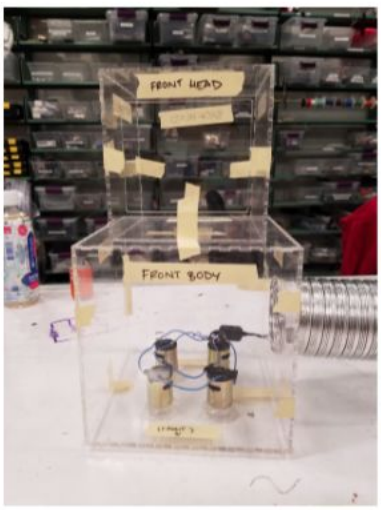




Final Size:



Shell



Expressions

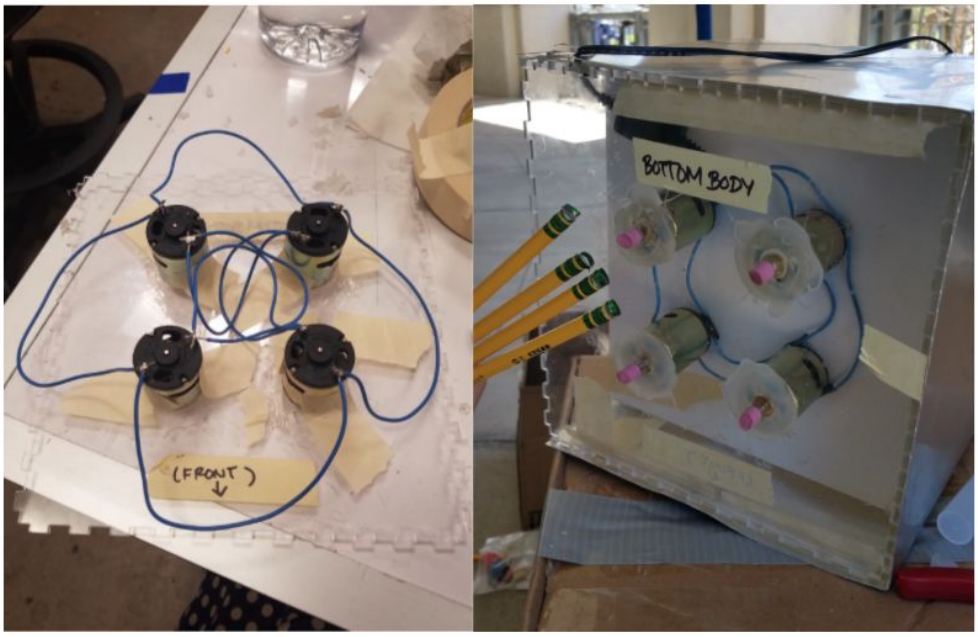
Default Face

Talking

After Someone Cranks

Crank

Motor



Measurements & Observations

We measured the amount of time that the user spent with Shiverbot, from initial contact to contact termination, through estimation with a stopwatch. Time spent cranking was also estimated using a stopwatch.

In addition to recording the age and gender of the participant, one of our team members would take notes on a recording sheet, which we created digitally on Google forms, and code the notes afterwards. The note taker recorded the user's emotional state, reactions to the robot, actions taken, and overall speech patterns. Though we would have liked to record the conversation topics had with the user, our inability to hold two-sided conversations with users eliminated conversation topic as a variable.

Quantitative Measurements

- Time spent with Shiverbot (shivering and nonshivering)
- Time spent cranking (shivering and nonshivering)
- Age (estimate)

Qualitative Measurements

- Gender
- Emotional state (curious, scared, helpful, etc.)
- Actions taken (help Shiverbot, leave immediately, etc.)
- Conversation topic

Study Recording Sheet

Age (estimate)

Your answer

Gender

- ○ Male
- ○ Female
- ○ Other : _____

Emotional State

- ○ Curious
- ○ Helpful
- ○ Evasive
- ○ Other : _____

Actions Taken

- ☐ Approach Shiverbot
- ☐ Use Crank
- ☐ Talk to Shiverbot
- ☐ Other: _____

Time spent with Shiverbot

Your answer

Time spent cranking

Your answer

Additional Notes

Results

After measurements, we grouped the data into when Shiverbot is shivering and when Shiverbot is not shivering. We tested whether or not shivering had a significant impact on the amount of time the participant spend cranking. Since the independent variables are in 2 groups (shivering and nonshivering) and the dependent variable is ratio data, we used the T-Test to test for significance. The test is statistically significant when the p-value is less than 0.05.

Hypothesis
A shivering robot will cause the participant to spend more time cranking the generator than a non-shivering robot.

Null Hypothesis
A shivering robot will cause the participant to spend less or equal time cranking the generator than a non-shivering robot.

We conducted additional statistical analysis between the participant's age, gender, and the willingness to help Shiverbot to see if there were any lurking variables or unanticipated correlations that we did not account for.

Collected Data

We were able to collect 15 data points of people interacting with Shiverbot. Of the 15, four of the data points were recorded when Shiverbot was shivering. Most of the users were students around the age of 20. We had 8 male users and 7 female users.

Shivering	Age	Gender	Emotional State	Actions	Time Spent	Time Spent (Cranking)	Crowd
N	20	M	Helpful	Talk	8	3	
N	20	M	Helpful	Crank	7	5	
N	20	F	Helpful	Crank	10	7	
N	20	F	Helpful	Crank	5	2	

N	30	M		Fix Crank	30	20	
N	20	M	Helpful	Crank	5	3	6
N	20	M	Helpful	Crank	10	5	
N	20	F	Curious, Helpful	Crank	30	5	
N	20	M	Helpful	Crank	15	10	
N		M	Curious, Helpful	Crank	20	7	
Y	20	M	Curious, Helpful	Crank	30	3	2
Y	20	F	Helpful	Crank, Take Picture	30	20	
Y	20	F	Aww	Crank, Take Picture	8	5	3
Y	20	F		Crank	10	6	

Analysis

In total, people spent an average of 16 seconds interacting with Shiverbot, and an average of 7 seconds using the crank.

- Total Time Spent (Mean): 16 s
- Time Spent (Median): 10 s
- **Total Time Cranking (Mean): 7 s**
- **Total Time Cranking (Median): 5 s**

When we break down the data to shivering and nonshivering, we are surprised to find that both the mean time of interaction for total time spent and time spent cranking are higher when Shiverbot is shivering.

Time Spent Total (s)

	Non-Shivering	Shivering
Mean	14	19.50
SD	9.59	12.15
SEM	3.03	6.08
N	10	4

Time Spent Cranking (s)

	Non-Shivering	Shivering
Mean	6.7	8.5
SD	5.23	7.77
SEM	1.65	3.88
N	10	4

We used conducted two unpaired T-Tests to test for statistical significance, one for total time spent and one for total time cranking. However, both of the p-values are much greater than 0.05. We do not have sufficient evidence to reject the null hypothesis that people spend an equal amount of time cranking regardless of whether or not the robot is shivering.

Total Time Spent

P value and statistical significance:

The two-tailed P value equals 0.3841

By conventional criteria, this difference is considered to be not statistically significant.

Confidence interval:

The mean of Non-Shiver minus Shiver equals -5.50

95% confidence interval of this difference: From -18.77 to 7.77

Time Spent Cranking

P value and statistical significance:

The two-tailed P value equals 0.6193

By conventional criteria, this difference is considered to be not statistically significant.

Confidence interval:
The mean of Non-Shiver minus Shiver equals -1.80
95% confidence interval of this difference: From -9.49 to 5.89

Discussion

For the sake of brevity, we have listed some of the more specific, unique results in bullet-form below. The results with greater significance we explain in greater detail further below.

- Large number of passersby didn't interact at all
- Polite, verbose requests (e.g. "Hi! Would you mind turning my crank?") didn't work at all
    - Had to just yell phrases such as "Hey, you!" and "Can you turn this crank, please?" in order to initially get people's attention
- Initial voice chosen difficult to understand; did not enunciate clearly enough
- Speaker too echo-y in box; had to place outside of box
- Mic on Samsung Galaxy S4 not strong enough; had to use Samsung Galaxy S7 Edge
- Had to move closer to speaker to keep bluetooth connection through glass once the speaker battery decreased too much
- Had to give continual auditory feedback in order to keep current user cranking
- Weight of heat duct inhibited shivering
- Accidentally superglued motors trying to fix erasers at some point
- Didn't create a conversation
    - Didn't make it clear that the cranking would warm the robot
    - Should have put ears or something on robot to indicate ability to listen

Results

There were a large number of passersby that we did not count as users because they did not interact with Shiverbot on any level. This may be a result of conducting the study on CMU's campus, as those who frequent CMU tend to be desensitized to the abundance of robots here.

Although we found that the mean for total time spent and time spent cranking are both higher for when Shiverbot is shivering, we were not able to reject the null hypothesis because of the lack of sufficient evidence. Because of various constraints around the project, we were not able to gather as much data as we had originally wanted. With more data, we can see the possibility of rejecting the null hypothesis.

In the end, we were not able to get the users to have a conversation with Shiverbot. Most users did not stay long enough to be engaged on a deeper level with Shiverbot. To save time, we also did not convey specifically that turning the crank would warm the robot. Shiverbot's expressions and voice feedback may also be delayed for the user because of manual input.

Possible Biases

As with a majority of Human Robot Interaction studies, our choice to conduct the study outside of a sterile, carefully controlled lab meant that a majority of the environmental factors affecting our study were beyond our control. Conducting the study on CMU's campus could have been just one possible source of bias, as the novelty of seeing robots in everyday life has likely worn off on those who frequent CMU.

We could neither control nor maintain specific weather throughout the duration of the study, and we could not eliminate variation among the various users who chose to interact with Shiverbot. Furthermore, due to outside commitments, there was no way for us to test at every time of day, meaning our data could have differed had we conducted the study over a 24-hour period. Time of year and day of week, for similar reasons, are also possible sources of bias.

In terms of human error, we as researchers could not ensure that we used precisely the same phrases each time we spoke to users, nor could we ensure that the response time for speaking to users was the same every time.

The biggest source of bias, however, was our reduced sample size. If we had tested a larger portion of the population, we would have been able to find more significant trends in the data collected. As such, we can only hypothesize as to whether or not the variation in our data is statistically significant or not.

Future Improvements

For future testing, we would choose to attach unbalanced weights to one DC motor in order to create the appearance of shivering instead of the method we used in our actual study. This method would require much less maintenance than replacing the uneven erasers on the bottom of the robot every time they wore down too much for the robot to shake.

As suggested in feedback from our presentation of the study's results, users would likely feel more compelled to verbally interact with Shiverbot if we had an external artifact

attached to the robot that could visually imply that Shiverbot is capable of receiving auditory input and not just manufacturing auditory output. The mere attachment of prosthetic ears to either side of Shiverbot's "head," for example, could encourage more users to hold two-sided conversations with the robot.

A stronger microphone would have allowed for us to more clearly understand any comments made by users. As such, we were only able to pick out occasional phrases or words. The best we could do was to figure out when users were and were not speaking and respond based on users' pauses. This was another obstacle preventing us from creating more authentic and "human" conversation.

To ensure that our robot can reliably shiver, we would explore lighter material than acrylic such as styrene or cardboard. We initially decided against cardboard because we were afraid that our robot would not be sturdy enough or look convincing enough. But we now recognize that the sheet metal can cover bad craft and lighter material allows the motor to work more effectively and achieve the shivering effect.

Due to the unforeseen number of tasks that needed to be completed simultaneously while conducting the study, we would have been greatly aided by the addition of more researchers to help us record data while interacting with users. We would delegate at least one person to verbally interact with users, one to control open-and-closing of the animated mouth, one to record data on the written forms, one to time the length of each interaction, and one to time how long each user spends cranking.