Does Robot Eye-gaze Help Humans Identify Objects?

Anna Ko

June 8, 2017

Abstract

The success of a social robot relies on its interactions with humans. To enhance human-robot interaction I looked toward useful forms of communication in human-human interaction. Humans communicate through verbal and non-verbal cues, in particular, my study focuses on eye gaze and the effects of its application on human-robot interaction. Specifically, I wanted to know whether having a robot exhibit eye gaze in a human-robot interaction would lead to humans feeling more comfortable and engaged, and whether this eye gaze would be a useful assistant in distinguishing between similar objects. My experiment involved a robot instructing a human to identify a target object. In the experimental conditions the robot would move its eyes toward the target object while describing the target through speech. Participants did not identify objects faster when the robot moved its eyes toward the target and participants did not report feeling more engaged when the robot moved its eyes. However, participants did report higher levels of comfort when the robot exhibited eye gaze and participants also felt the robot was more natural as well. Thus, in the hopes of creating successful social robots it is important to implement eye gaze cues in the robot in order for humans to feel more comfortable during the human-robot interaction. Having humans feel more comfortable might then encourage them to interact with the robot longer and might also encourage them to interact with the robot, or any other social robot, more often.

Contents

1 Introduction

I am interested in improving the way social robots interact with humans. Social robots can be categorized as autonomous robots that interact and communicate with humans or other autonomous agents. As technology advances, the capabilities of social robots increases as well. As a result of advanced technology there now exist some social robots that act as museum guides. A way of improving the interaction of these robot museum guides might be the addition of eye gaze. Through eye gaze, humans may be able to more easily distinguish which museum object the robot is describing. However, it is not known whether eye gaze would allow humans to more easily distinguish between objects. Thus it is important to study the interaction between humans and robots in order to make empirically based decisions on improving social robots. For example, what qualities can we give a robot such that humans feel more comfortable and engaged while interacting with them? This is important because social robots need to interact with humans. Imagine a social robot were given an assigned goal to retrieve coffee from the store through the help of nearby humans. If the nearby humans are uncomfortable or feel unengaged with the robot they can easily walk away. However, social robots need the interaction and help of humans in order to function. This is because the purpose of social robots is to interact with humans. Thus, by nature, they need humans to positively respond to them. If humans felt negatively toward social robots, then the robots would not be able to complete their assigned goals. Therefore, it is essential in the study of social robotics to have humans feel comfortable and engaged while interacting with the robot. So, how can we make humans feel more comfortable and engaged while interacting with robots and how can we make communication about objects easier and more successful? In this thesis I will investigate whether the addition of eye gaze could be beneficial to human-robot interaction. These results are especially important in the case of social robotics as the primary function of a social robot is to interact successfully with humans. This successful interaction could be described as an interaction in which the task of the social robot is completed correctly through the cooperation of a human subject. Thus, it is vital to assess what factors could be implemented into human-robot interaction so that successful interactions between humans and social robots can occur.

Humans interact with one another through the communication of verbal and nonverbal cues. While most of the explicit communication occurs through speech (verbal cues), research in psychology has shown the importance of nonverbal cues in human-human interaction. Examples of nonverbal communication include gesturing, body posture, facial expression and eye gaze. These nonverbal cues have been shown to help structure interpersonal communication, regulate the system of conversation, signal the flow of interaction, and serve in providing meta-communication and feedback Mandal 2014. In particular, the importance of eye gaze in human-human interaction has been studied extensively since the 19th century (Hugot 2007). Where it has been found to serve many functions. For example, a person's eye gaze can be used to influence people's judgements of them, such as their evaluation of liking and competence (Kleinke 1986). Eye gaze can also be used to discriminate between similar objects (Hanna and Brennan 2007). Specifically, in a study conducted with face-to-face partners, experimenters found that participants were able to use their partner's eye gaze direction as an early cue to distinguish ambiguity between same-color targets (Hanna and Brennan 2007).

I am interested in studying the effect on human-robot interaction when a robot exhibits targeted eye gaze – eye gaze directed towards a specific target. The addition of this targeted eye gaze could lead to humans feeling more comfortable and engaged when interacting with robots. It could also lead to a higher task success rate for the human-robot interaction. I anticipate that targeted eye gaze could be an important factor in human-robot interaction due to its usefulness in human-human interaction. Specifically, I hypothesize that the addition of targeted eye gaze to human-robot interaction will result in humans feeling more comfortable and engaged during the interaction, that they will rate the robot as more natural, and that humans will be able to distinguish between similar targets faster.

The rest of my paper will, first, discuss the background and related work pertaining to my study. Then I will be describing my experimental design, the materials I will be using and information on my participants. Afterwhich, I will describe my hypotheses. Then present my results in the results section and interpret these in my discussion section. Finally, I will conclude and discuss possible futures studies in the conclusion and future work section.

2 Background and Related Work

Human-human interaction is composed of verbal and nonverbal communication. Nonverbal communication can be characterized as everything other than words that are used during communication ("Nonverbal Communication" 2014). Examples of nonverbal communication include: hand gestures, body orientation and position, body movements, facial expressions and eye gaze ("Nonverbal Communication" 2014). Numerous works have led to the placement of nonverbal communication as central to meaning acquisition, especially where relational communication, emotional expressions and impression management are concerned (Manusov and Patterson 2006). Nonverbal communication contributes significantly to communication and studies have shown that when nonverbal cues are not congruent, individuals often put greater credit to the nonverbal cues than to the verbal ones (Manusov and Patterson 2006). Thus, nonverbal communication is essential in human-human interaction, and as such, one would expect the addition of nonverbal cues in human-robot communication to also enhance human-robot interaction.

While there are many different forms of nonverbal communication, my experiment focuses specifically on eye gaze. In human-human interaction, eye gaze is a valuable tool for humans to evaluate one another and their interest in the interaction. Many studies have shown the influences and effects of eye gaze in human-human interaction. Eye gaze is found to influence people's judgement of others in their evaluation of liking, competence, attentiveness, credibility, and dominance (Kleinke 1986). Interviewees are evaluated as more attentive when they look at participants more, and briefer responses are given when the interviewer does not look at them (Kleinke 1986). Eye gaze has also been found to serve as a social control, and people gaze at a person more when they are attempting to ingratiate with others and when seeking friendship (Kleinke 1986). However, in the case of prolonged and unexplained gaze, the chances of the participant feeling engaged are much lower. When unexplained gaze occurs it can act as a stimulus for eliciting escape and avoidance (Kleinke 1986). Prolonged and unexplained gaze can be described as when someone is staring at another person without a purpose. In Ellsworth et al.'s study they had a driver on a motor scooter stare at the driver in the car next to him when both parties were stopped at a red light (Ellsworth et al. 1972). This same experiment was also conducted without a scooter. Instead of riding the scooter, the person stood at the corner of the sidewalk and would begin the stare condition when a car pulled up and stopped for the red light. The results for both tests found that subjects who were stared at went across the intersection significantly faster than subjects in the no-stare condition (condition where the driver was not stared at). Due to this effect, in my experiment, I attempted to create a situation in which humans would not interpret the robot's eye gaze as purposeless. I did this through including an introduction component in my experiment that occurs before the task. This introduction component attempts to alleviate the interpretation of purposeless gaze through having the robot engage in eye gaze cues during the introduction. Doing so might allow participants to become familiarized with her gaze and thus feel as though it were meaningful rather than unexplained.

IN a study conducted by Hanna and Brennan, they looked at whether eye gaze can be used in the resolution of temporary ambiguity (Hanna and Brennan 2007). Their experiment involved two participants who were randomly assigned the role of director and matcher. The director and matcher sat across from each other with a table between them. A low barrier was placed on the table, and both participants had identical copies of the same object. During the experiment, the director instructed the matcher to move two of the display objects to one of three spaces located on the matchers side. While this was occurring, an eye tracker was used to track the matchers eye gaze. They found that when matchers needed to distinguish targets from their same-color competitors, the matcher was able to use their partner's eye gaze before the linguistic point of disambiguation. This means that before the director finished describing all the details of the target, the matcher was able to use the director's eye gaze cues as a way of disambiguating between same-color competitors. The question is, can this effect also be seen when we replace the director with a robot? My study hoped to replicate their finding in human-robot interaction.

While the effect of eye gaze in human-human interaction is interesting, much less research has been done on the effect of eye gaze in human-robot interaction. In particular, there have been mixed results found on this topic. Some of these studies have found that the addition of eye gaze cues do assist in improving performance in human-robot interaction (Khoramshahi et al. 2016). One such study, conducted by Khoramshahi et al. in 2016, found that people were able to exploit gaze cues given by the robot in order to predict another person's movement and better coordinate their motions with their partner (Khoramshahi et al. 2016). However, not all studies have found eye gaze to be helpful. In a study conducted by Karreman et al. in 2013, they found that having a robot exhibit eye gaze did not help to improve performance (Karreman et al. 2013). Instead, their analysis showed the opposite result; people paid more attention when the robot did not look over at the object of interest versus when the robot gave a gaze cue toward the object(Karreman et al. 2013). Due to these mixed results it is unclear whether eye gaze is a help or hinderance toward successful human-robot interaction. Thus my study attempts to resolve these mixed results.

Khoramshahi et al.'s 2016 study on gaze cues in human-robot interaction found that gaze cues in a human-avatar joint action led to better coordination (Khoramshahi et al. 2016). Joint action can be regarded as some action or task requiring the direct physical involvement of two parties.The joint action in this experiment involved two players mimicking each other's hand motions. One player was the follower and the other was the leader, and the follower had to mimic the hand motions of the leader. The robotic avatar acted as the leader in this study. Their study involved an implementation of simple gaze behavior to this task. They found that gaze cues significantly improved their participants' reaction times. This study showed that, in joint action, a leader's gaze cues can assist the follower with action prediction, task sharing, and action coordination. The experimenters also gave out a questionnaire to their participants. This survey led them to learn that participants perceived the gaze cue behavior as cooperative, human-like and realistic . Although this study focused on human-avatar joint action, the results found from the questionnaire motivated my assumption that gaze cue behavior on a robot might lead to participants feeling more comfortable and engaged. Also, as the questionnaire found that participants perceived the gaze cue as human-like and realistic, I believed that in conditions with gaze, the participants will also rate the robot as more natural.

The robot I used in my experiment has a face display which is projected on a 2-dimensional surface (a tablet). An important factor to consider in the case of 2-dimensional surfaces is the Mona Lisa effect. The

Mona Lisa effect describes a situation in which the eyes in a portrait seem to be following the observer(s) as they pass (Kuno et al. 2005). This feeling of being watched could lead to humans feeling uncomfortable. In order to combat this issue, Kuno et al. had the robot turn its face toward the human when it seemed as though the human were looking at it. In their study they found that humans looked at the robot's body as well as his face, and that they could perceive the robot's gaze more accurately when the body moved as well.

While there are experiments in human-robot interaction in which humans accurately interpreted robot gaze cues, there are also experiments where humans do not interpret robot gaze cues the same way they would with human eye gaze cues. This happened in Karreman et al.'s 2013 study which looked at the effect of robot gaze cue in a museum-like setting with two human participants (Karreman et al. 2013). The participants and robot were located in a room with two famous paintings, the Mona Lisa and The Girl with the Pearl Earring. The robot told the two participants information on the two paintings and would give different amounts of attentional gaze toward each participant while speaking. The robot also, while describing the paintings, would give a gaze cue toward the painting it was describing. Then the robot would tell the participants to look at something specific about the painting, such as the the position of the lady's hands. The results of this study were interesting as they found that when the robot engaged in object-orientated gaze behavior the participants felt a more positive attitude toward the robot. However, the participants also paid less attention. The robot eye gaze cue in this experiment did not direct the participants' attention to the object of interest as was expected. This study showed a lack of accurate human interpretation in understanding the robot's object orientated gaze cue. Studies like these are interesting as they suggest that the addition of human-like behavior on a robot may not necessarily mean humans will interpret this behavior the same way they would in human-human interactions.

3 Methods and Design

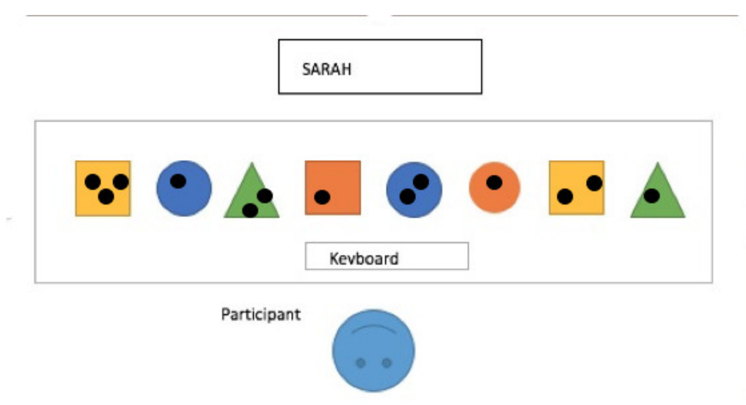

Figure 1: Experiment design

Based on these prior works, I will now define my study. The main goals for my experiment are to see whether giving a robot eye gaze will lead to better disambiguation of similar targets and whether this addition of eye gaze on a robot will lead to higher self-reported comfortability and engagement. I am also interested in seeing if participants rate the robot as more natural when it engages in eye gaze.

3.1 Procedure

My experiment design is borrowed from Hanna and Brennan's study (Hanna and Brennan 2007). This was so I could see if similar results would occur when the director in their study was replaced with a robot. My experiment setup is such that the robot was on one side of the table and the participant was sitting across from her. Between them was a table with shapes and a keyboard. Figure 1 shows the set up of the experiment and Figure 2 shows Hanna and Brennan's setup. The experiment task involved the robot

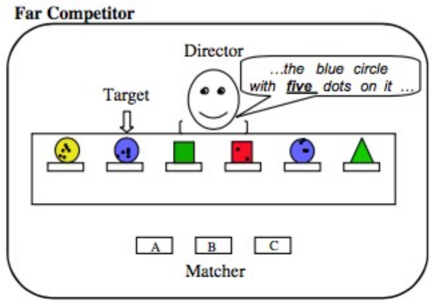

Figure 2: Hanna and Brennan's experiment setup showing an example of a far competitor target.

asking the participant to identify an object which the robot described. An example of how the robot would describe an object is, 'Press the button corresponding to the blue circle ... with two dots'. The ellipses indicate a slight pause in the robot's speech. This slight pause was included in order to record the exact time the phrase 'with [number of dots] dots' was spoken. The participants were randomly selected into the no gaze, gaze, and, gaze and body shifting conditions. The participants were also told that the robot was acting autonomously before their interaction with the robot began. Below is a more detailed walkthrough of the experiment.

The experimenter was observing the table (through a web cam placed on the robot pointed toward the table) and controlling the robot through a computer outside of the lab room where the experiments were held. During the experimenter's description of the experiment, participants were told only that they will be asked to complete a task by SARAH after a short introduction conversation. The room was set up with the robot on one side and the participant standing on the other side, between them was a table with an empty template, a stack of foam objects next to the template and a keyboard directly in front of the participant. The template was located above the keyboard on the table, and the foam objects will be placed on top of this template. A description of the template can be found in the materials section. The experiment involved the participant laying out eight of the 36foam objects onto the empty template. These eight objects included three different shapes: circles, squares and triangles with four different possible colors (blue, green, red, yellow) and number of dots on the shape (1-4 dots). The experiment started with the robot greeting the participant and then asking three general introductory questions which asked, of two things, what the participant prefers more. The purpose for asking preference questions is so that the robot would be able shift her eyes while doing so. One of the questions used was, "Would you rather dwell in the sky or in the sea?". During this question, the robot would shift her gaze up and down corresponding with her speaking 'sky' and 'sea'. This is so that participants would feel comfortable with the robot shifting her gaze. After the robot asked the participants her questions she then gave them the opportunity to ask her one question. The purpose of this introductory conversation was to make participants comfortable with the robot's eye gaze, make sure participants were comfortable with the robot and ensure that they understood the robot's audio.

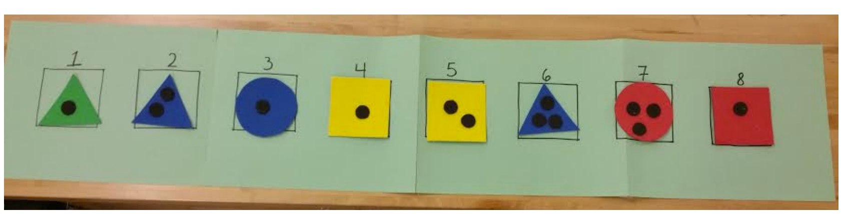

Figure 3: Shapes Setup Example

After the introductory conversation, the robot asked the participant to lay out the first eight objects onto the template. The robot asked the participant to lay out the objects twice before it described the task. The first layout was a practice one in order for the participant to get used to laying out the objects correctly. After the practice trial of laying the objects on the template, the robot proceeded to tell the participant the task. The task involved the participant identifying an object the robot described. Specifically, this was done through the participant pressing the keyboard button associated with the location of the object on the shapes template and then pressing the enter key. The shapes template is a template on which the objects were placed. The shapes template had eight squares (1″ x 1″) each labelled chronologically with a number 1-8 corresponding with the keyboard buttons 1-8 located at the top of the keyboard. See Figure 3 for an example of how the shapes would be setup on the shapes layout template.

There were 12 trials for each participant. Each trial involved the robot asking the participant to identify a particular target on the shape layout through pressing the corresponding keyboard button. In terms of the shape layouts, after 3 trials the participant was asked by the robot to clear the template of the shapes and put the next 8 objects from the stack of foam objects onto the shape layout template. In the shapes layout there were three different kinds of competitor targets: no competitor targets (NO), near competitor targets (NC) and far competitor targets (FC). The no competitor targets were ones in which there were no competing objects on the table with the same shape and color as the target. The near competitor targets were ones in which there was a competing object, of the same shape and color, on the board located next to the target object. The far competitor targets were ones in which there was a competing object, of the same shape and color, however, this competing object was located away from the target object, such that they were not next to each other yet still both on the shapes layout. In Figure 3: the NO targets are the shapes labelled as 1,3,7 and 8, the NC targets are shapes 4 and 5, and the FC targets are shapes 2 and 6. In each experiment, the participant identifies: 4 NO targets, 4 NC targets, and 4 FC targets. When describing the shapes, the robot followed this template of speaking: ″Please press the corresponding keyboard button for the [color] [shape]″, in eye gaze conditions the robot shifted her gaze (and her body in the eye gaze and body shifting condition) toward the target object, and then the robot continued, ″ with [number of dots] dots″.

There are three conditions in my experiment, hereto labelled as speech, eye gaze, and body and eye. I included a condition involving SARAH moving her body because of Kuno's experiment which found body shifting to be helpful in directing attention (Kuno et al. 2005). In the speech condition SARAH interacted with the participant solely through verbal cues. In the eye gaze condition, SARAH used eye gaze cues during the introduction section and shifted her eyes toward the target object during her description of the object. In the body and eye condition, SARAH used eye gaze cues during the introduction section and shifted her eyes and body toward the target object during her description of the object. To further elaborate, during the experiment, in the eye gaze and body and eye conditions, the robot will first describe the target object's shape and color. After which the robot will shift its eyes toward the target, and shift its body slightly towards the target if the participant is in the body and eye condition. Then the robot will describe the number of dots on the target object. The line, ″with [number of dots] dots″, will be referred to as the verbal disambiguator. This is because this line of speech is what allows the participant to distinguish exactly which object the robot was describing.

Time data for when the robot begins speaking the verbal disambiguator and time data for when the participant pressed the enter key (after pressing the correct corresponding button on the keyboard) was collected. This data was put into an excel sheet and a column was created for the difference between the two times. Participants were given a survey to fill out after they had completed their interaction with the robot and left the lab room. This survey asked participants to rate, from 1-7, how comfortable and engaged they felt during their interaction, as well as how natural they felt the robot was, and why. See Survey Questions for the survey questions A

3.2 Participants

A total of 40 participants from Union College participated in my experiment. Four of the participants data were not used in data analyzation as they were told the robot was not acting autonomously. They were told the robot was not acting autonomously as originally I had intended to include this as a condition. Of the remaining 36 participants who were told deceptively that the robot was acting autonomously, 12 were in the speech condition, 12 were in the eye gaze condition, and 12 were in the body and eye condition. These 36 participants were all told the robot was acting autonomously. The age range of my participants was from 18-25 years old. Unfortunately, I was unable to collect gender data from two participants. Aside from those two, there were 21 females and 13 males in my experiment. I used my social networks to collect patients.

The one criteria I had for participants in this experiment was that they must be able to distinguish between different colors, specifically: red, blue, green and yellow.

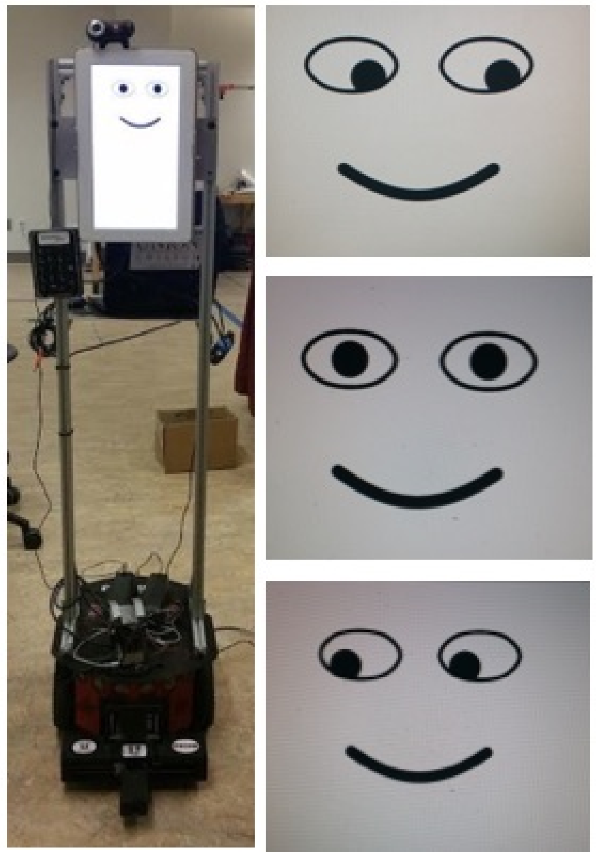

Figure 4: On the left is SARAH, owned and created by the Computer Science department at Union College. On the right of SARAH are three possible eye gazes she can engage in.

3.3 Hardware

The robot I used in my experiment was SARAH (Socially Appropriate Robot that Approaches for Help) a robot owned by the Computer Science department at Union College. Figure 4 shows the robot I used in my experiment. SARAH has a screen on which an animated face can be projected and she can move. The face projected on SARAH's screen has irises which can move to simulate eye gaze and she played her audio through the help of a text-to-speech program, MaryTTS (see MaryTTS (Version 5.2, 2016)). See more detail on SARAH in section 3.4.

The keyboard I used in the experiment, in order to collect time data from when the participant identifies the object, has a cover over the keyboard such that the only visible keys on the keyboard will be the buttons 1-8 at the top of the keyboard and the enter button. For this keyboard input, I wrote a python code that was run on SARAH. This code was run through a window in the terminal, and when the participant pressed the enter key a timestamp would be printed in the terminal. This timestamp came from ROS wall time (wiki.ros.org).

There was also a webcam placed above the table for the conductor to ensure things are going smoothly during the experiment. This webcam only capture video of the table and not of the participant.

3.4 Software

In order to communicate with SARAH through my computer, I had to use ROS [1] (Robot Operating System) (see ROS Indigo Igloo (8th ROS distribution release, 2014)). ROS allows for a computer to communicate with a robot through the use of nodes. In particular, the computer used to command the robot is known as the ROS master. The ROS master then communicates to the robot, in this case, SARAH through the use of nodes. These nodes are independent programs and nodes communicate through sending messages which are organized into topics. The nodes can either publish a message to a topic or subscribe to a topic. Then, based on the nodes independent properties and messages received, the robot will operate accordingly.

In order to facilitate easier use of SARAH, SARAH was controlled through a GUI I made with Python's Tkinter interface (see Tk, GUI extension (1991)). A GUI was written for each condition and an input GUI for speech was made as well. The GUI consisted of audio commands, eye moving commands and body movement commands.

The GUIs themselves did not start any nodes, but rather, the GUIs just published messages to topics. Separately, in the terminal, nodes were started. These nodes pertained to SARAH shifting her eyes (rosbridge_websocket), SARAH producing audio (mary_node), SARAH moving her body (RosAria), getting sound input (audio_capture), playing sound input on my computer (audio_play), starting webcam (usb_cam) and playing webcam video on my computer (rqt_view). The rosbridge_websocket node, the audio_play node, and the rqt_view node were all started on my computer, and the other nodes were started on the laptop connected to SARAH (positioned on SARAH and connected to SARAH). In order for SARAH to move her eyes, my GUI would publish a message to the topic, '/gaze_listener'. Since rosbridge_websocket was subscribed to the '/gaze_listener' topic, rosbridge_websocket would receive the message and then SARAH's eyes would move accordingly. This similar process was also used to produce audio on SARAH and have her move her body. The input GUI, specifically, had an input box wherein I would type out something I wanted SARAH to say and then a message would be published to '/tts'. Mary_node was subscribed to this topic and SARAH would then play the corresponding audio.

In the mary_node created by Nick Webb, I also added a line to get the ROS current wall time. ROS wall time was used because it corresponded with the wall time information gotten from the participants' keyboard input. In terms of collecting data, I specifically got only the wall time data for when the verbal disambiguator line began to play.

SARAH's face (with moving eye capabilities) was created through the use of JavaScript and html by my other thesis advisor, Kristina Striegnitz, a Computer Science professor at Union College. The rosbridge_websocket node was used to link the html page to ROS, and the rosbridge_websocket node was a part of the Rosbridge package found on wiki.ros.org. RosAria, audio_capture, audio_play, usb_cam and rqt_view all came from wiki.ros.org as well. The MaryTTS text-to-speech software was changed into a node (mary_node) by my thesis advisor, Nick Webb, an assistant professor at Union College.

3.5 Materials

The shapes layout template, on which the participant laid out the shapes for each trial, has eight squares with an approximate length and height of three inches and a two inch distance between each drawn out square. The squares are labelled with numbers 1-8 above them, with 1 labelling the square on the farthest left of the template and 8 labelling the square on the farthest right.

The shapes I used in my experiment were cut from blue, red, yellow and green foam. There were three types of shapes: circles, triangles and squares. All the circles had an approximate radius of 1.5 inches, all the triangles had a height of approximately 2.5 inches, and all the squares had an approximate height and length of 3 inches. On each shape there is a specific number of black dots ranging from 1-4 black dots. Figure 3 shows one of the shape setups on top of the shape layout template.

The survey given to participants after the experiment can be found in A.

[1] ros.org

3.6 Hypotheses

While my experiment involves three different conditions (speech, eye, and body and eye), I did not anticipate any statistically significantly different results for the eye gaze condition versus the body and eye condition. This is because I believe that eye gaze would be the main factor affecting participants' response time. However, should no significant results be found as a result of the only eye gaze condition, I am hoping that the eye gaze and body shifting condition would lead to significant results. I believe this could occur because of the findings found from Kuno et al. which found that participants perceived the robot's gaze more accurately when the body moved as well (Kuno et al. 2005).

Hypothesis 1 (H1): My main hypothesis anticipated that in all conditions wherein SARAH engages in eye gaze (eye gaze and body and eye conditions) the human subjects will disambiguate between similar objects faster. I anticipated that, though SARAH is a robot, participants will interpret her eye gaze in the same way they would in human-human interaction (Hanna and Brennan 2007).

Hypothesis 1a (H1a): In regards to the competitor types, I anticipated that the response time for FC targets will be faster in conditions involving eye gaze. This is because FC targets allow for easy distinguishing of eye gaze. For example, if a blue square is located in space 1 of the shapes layout and a blue square is located in space 4 of the shapes layout, then a person might be able to distinguish which shape is being referred to by the eye gaze much easier. In this case, it would be more clear which shape the eyes were pointed toward as the blue square in space 1 would be on associated with a right downwards gaze, and the blue square in space 4 would be associated with a left downward gaze. In the case of NO targets, there is no need to distinguish ambiguity earlier than the verbal disambiguator because there are no other objects on the board that are the same shape and color (shape and color are described before the verbal disambiguator is spoken). In the case of NC targets, the target and its competitor are located next to each other. Because of this it could be difficult to distinguish, on eye gaze alone, exactly which object SARH is looking at. Thus, I hypothesized that the response time for FC targets will be faster in conditions involving eye gaze.

Hypothesis 2 (H2): I also anticipated that in conditions where SARAH engages in eye gaze participants will report higher levels of comfort and engagement, and that they will rate SARAH as more natural than in the condition where SARAH does not shift her eyes. I believe this will occur due to SARAH's human-like behavior in shifting her eyes. I think participants will have higher ratings of comfort, engagement and naturalness because they will be interacting with an entity that exhibits behavior they are comfortable and familiar with.

4 Results

4.1 Data

For each participant, I recorded the 12 time stamps [2] for when the verbal disambiguator was spoken, the 12 time stamps for when the participant pressed the keyboard, and the associated competitor type (NO, NC, FC) for these time stamps. I also calculated the time difference from when the verbal disambiguator was spoken to when the keypress occurred. There was a total of 432 time differences (12 participants x 12 trials x 3 conditions).

The reported comfort level, engagement level and naturalness numbers from the survey were also recorded. Question 4 on the survey asked how natural participants would rate a human, SIRI or similar virtual assistance program, and SARAH (see A). This was done so participants would have a baseline for how they would then rate SARAH. Given the data for question 4 on the survey, I calculated a percentage for how natural participants felt SARAH was compared to how natural they rated humans. For example, if a participant rated humans with 100 and SARAH with 50, the new value for how natural they felt SARAH was became 50%. Similarly, when a human was rated as 87 and SARAH was rated as 80, the new value for SARAH became 92%.

[2] There are 12 time stamps because each participant went through 12 trials and identified 12 objects

4.2 Time Analyzations by Condition

For all analyses done, alpha was set at .05 such that the p-value must be less than .05 for statistically significant results.

To analyze the recorded time data, I used the calculated time difference from when the verbal disambiguator was said to when the key press occurred. I performed a one way between subjects ANOVA on this time difference by the conditions. The p-value was found to be greater than .05, thus the difference between the means by the different conditions was not a statistically significant difference. The means are shown in Figure 6 and Figure 5.

TimeDifference (s)	Speech	Eye gaze	Body and Eye
Mean	0.828891	0.554811	0.625980

Figure 5: The mean times found for each condition.

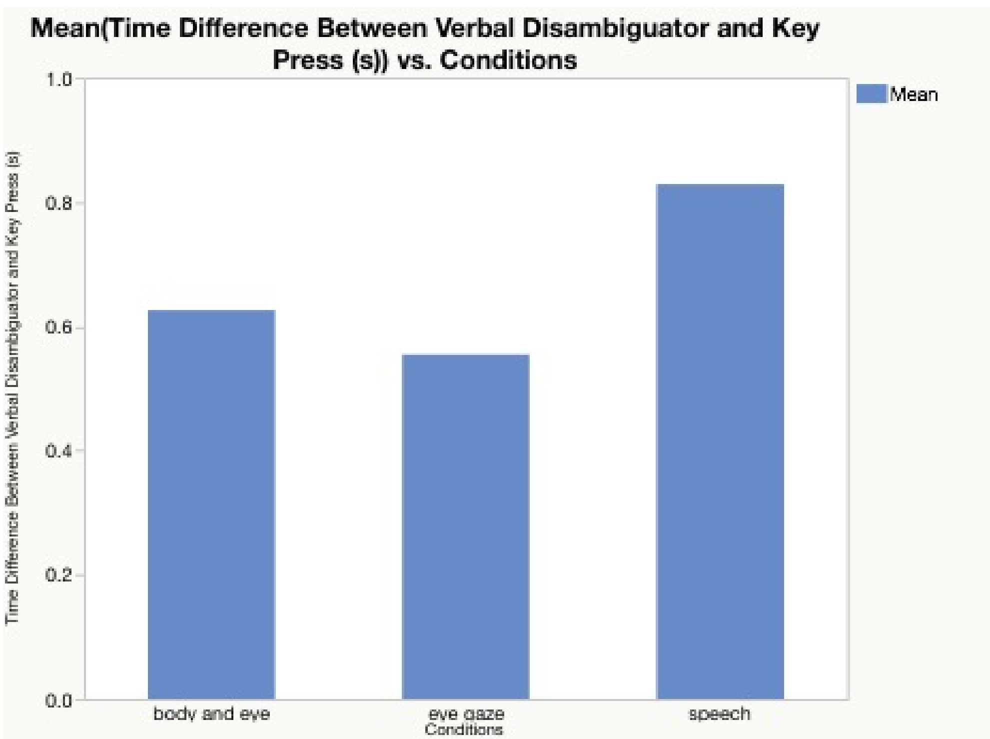

Figure 6: The mean time difference between when the verbal disambiguator was said to when the key press occurred by the different conditions.

4.3 Competitor Type Time Differences by Condition

To analyze the time difference from when the verbal disambiguator was said to when the key press occurred by each of the different competitor types, I had to use subsets of the data. Specifically, these subsets were: the data set containing only NO targets, the data set containing only NC targets and the dataset containing only FC targets. There were 144 time differences in each dataset. A one way between subjects ANOVA was performed on each of these subsets of data. The ANOVA performed on the NC targets dataset and the NO targets dataset found the p-value to be much greater than .05. Thus there were no statistically significant differences found between the mean time differences by the conditions. The ANOVA performed on the FC target dataset found p=0.0594. While this number is close to .05, it is still greater than .05, and as such, there was no statistically significant difference between the mean time differences by the conditions. See Figure 7 and Figure 8 for average time differences for the FC target set by the conditions.

FC Target Set Time Differences (s)	Speech	Eye gaze	Body and Eye
Mean	1.50785	0.94124	0.77312

Figure 7: The mean times found for each condition for the FC target set.

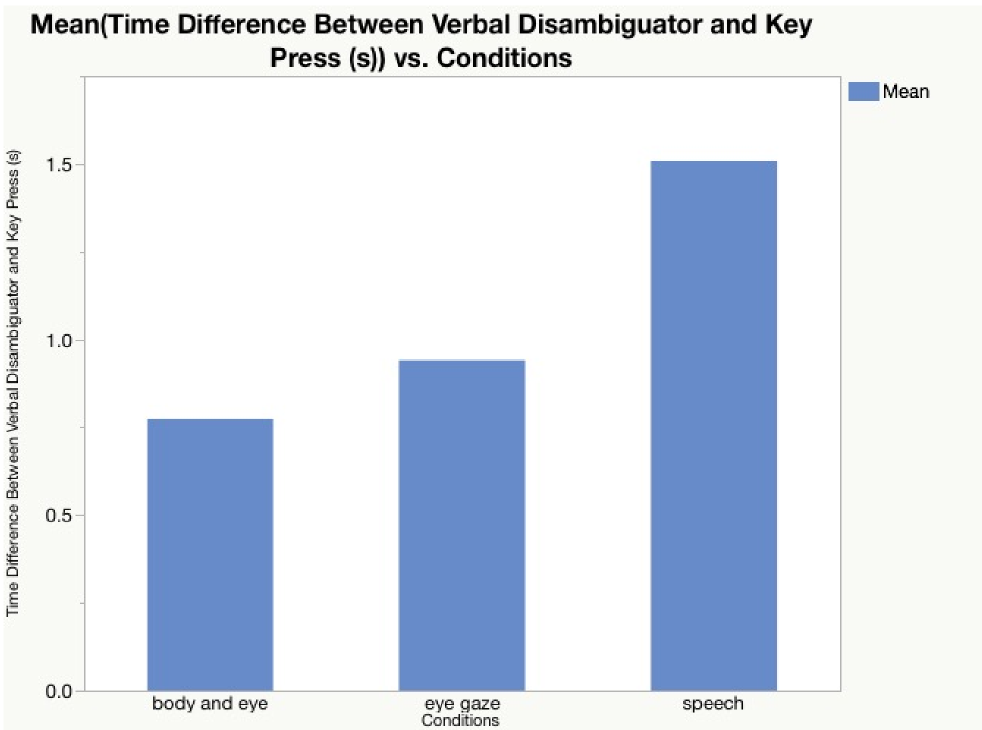

Figure 8: The mean time difference between when the verbal disambiguator was said to when the key press occurred by the different conditions for the FC target set.

4.4 Comfort, Engagement and Naturalness by Conditions

I performed a one way between subjects ANOVA on the reported comfort levels by the conditions and found the p-value to be less than .05. A paired t-test was then performed on the comfort levels of the speech condition vs. eye gaze condition, speech condition vs. body and eye condition, and eye gaze condition vs. body and eye condition. I found that p was less than .05 in the case of eye gaze vs. speech and body and eye vs. speech. So the average comfort level for the eye gaze condition was statistically significantly higher than the average comfort level for the speech condition. The average comfort level for the body and eye condition was also statistically significantly higher than the average comfort level for the speech condition. When comparing the average comfort levels in the eye gaze condition by the average comfort levels in the body and eye condition p was less than .05, thus not statistically significant averages. See Figure 9 for the average comfort level by the conditions.

I performed a one way between subjects ANOVA on the engagement level by the conditions and found the p-value to be greater than .05, thus not statistically significant. The mean engagement rating found for the speech condition was 4.41667, the mean engagement rating found for the eye gaze condition was 4.83333 and the mean engagement rating found for the body and eye condition was 4.6667.

Comfort level	Speech	Eye gaze	Body and eye
Mean	4.5	5.0	5.1667

Figure 9: The mean comfort level found for each condition.

I also performed a one way between subjects ANOVA on the how natural participants rated SARAH as compared to how natural people rated humans by the conditions. The p-value was found to be less than .05, thus there existed a statistically significant difference between two or all of the conditions. I then performed a paired t-test on the average rating of SARAH's naturalness in the speech condition vs. the eye gaze condition, in the speech condition vs. the body and eye condition, and in the body and eye condition vs. the eye gaze condition. After these tests, I found p to be less than .05 when comparing how natural people felt SARAH was in the eye gaze condition vs. in the body and eye condition. Thus this difference in percentage mean was statistically significant. See Figure 10 for the means.

% Naturalness of SARAH	Speech	Eye gaze	Body and eye
Mean	49%	54%	43%

Figure 10: The mean % of how natural people felt SARAH was as compared to humans found for each condition.

5 Discussion

My first hypothesis (H1) anticipated that participants would distinguish between similar objects faster in the eye gaze conditions (eye gaze and body and eye). Unfortunately, since p was less than .05, the mean time difference (from when the verbal disambiguator was said to when the key press occurred) of each condition were found to not be statistically significantly different. Thus, the null hypothesis cannot be rejected, and based on my data alone, humans may not distinguish between similar objects faster in human-robot interaction when the robot exhibits eye gaze. Thus it may be the case that people do not interpret eye gaze in the same way they would in human-human interaction or that the eye gaze exhibited by SARAH could have been improved upon. One participant in particular wrote in her survey, "the eyes made me superbly uncomfortable". After speaking with her after the experiment, she also told me that because she felt the eyes were unnatural she didn't look at SARAH's eyes much.

Another reason the different response times were not found to be statistically significantly different could have been due to the aggregate of the three different competitor types. The NO targets did not require a verbal disambiguator for the participant to identify the correct object as there were no other objects on the table with the same shape and color. Thus there was no need for the participants to rely on nonverbal cues in order to identify the correct object faster. The NC targets did require a verbal disambiguator, however, since targeted eye gaze is not a precise method in small spaces, it could be the case that there were no benefit to using eye gaze when two objects are placed next to each other. Because of this, participants may not have been able to use SARAH's eye gaze to distinguish between two similar objects that were located next to each other. Thus the combination of the these two competitor types plus the FC targets might make it more difficult to find a statistically significant average time difference between the three conditions. For this reason I also looked at the dataset containing only FC targets.

Another possible reason no statistically significant difference was found in the mean response time by the conditions, could have been due to cognitive dissonance in the participants. This would have been due to the added step of pressing the correct keyboard button and then the enter button. A couple of the participants in my study actually forgot this step of pressing the enter button during the first trial. This addition of pressing the enter key may also have added some extra time to the participants' response times, however, since all participants had to press the enter key, this may be unlikely.

My secondary hypothesis (H1a) suggested that participants would have higher response times in the eye gaze conditions, specifically for the the FC targets. Unfortunately, after analysis, p was greater than .05 and the response times for each condition for the FC targets was not statistically significantly different. Though the p-value was close to being less than .05, as p=0.0594. I believe that, had I had a few more participants (perhaps 3 more in each condition), the p-value would have been statistically significant. This is because of the 4 participants who were told SARAH was not acting autonomously. When including these participants in the analysis I actually found p to be less than .05. Thus, I can say, if more participants had been collected this result may have been different, a statistically significant difference between the means may have been found and this would have shown that the addition of eye gaze does help distinguish between objects that are not right next to each other.

My tertiary hypothesis (H2) anticipated that participants would rate higher levels of comfort and engagement when they were in the eye gaze conditions (eye gaze and body and eye). And also, that participants would rate SARAH as more natural in the eye gaze conditions. In the case of comfortability, people reported feeling statistically significantly more comfortable in the eye gaze conditions than in the speech condition. This means people were more comfortable when SARAH used eye gaze cues. This is an important discovery as it means the addition of eye gaze can make humans feel more comfortable in human-robot interaction. Albeit, this difference in comfort level was not very high as the average comfort level reported for eye gaze was only .5 higher than speech, and the average comfort level reported for body and eye was only .6667 higher than speech.

There was no statistically significant difference between the average reported engagement levels for each of the conditions. Thus participants did not feel more engaged in the eye gaze conditions. This outcome could have been due to the speed at which SARAH spoke. In the comments section of the surveys, many participants said they disliked the speed at which SARAH spoke and wished she spoke faster. I conversed a bit with most of the participants after the experiments as well, and almost all of them mentioned how slow SARAH's speech production was. I think this may have been a reason why participants in all conditions rated SARAH about the same. On one of the surveys a participant even wrote, "long pauses made me feel disengaged." (eye gaze), another participant also wrote "Took too long to talk so I spaced out" (body and eye condition).

There was a statistically significant difference between the average % naturalness participants felt SARAH was by the conditions. However, this statistically significant difference was only found between the eye gaze and body and eye conditions. Meaning, participants reported SARAH to be statistically significantly more natural in the eye gaze condition than in the body and eye condition. And, when comparing how natural participants felt in the speech condition compared to the eye gaze condition, and the speech condition compared to the body and eye condition, there was no statistically significant difference between the means. This was a surprising find as the average found for body and eye was 43%, this is lower than both the speech condition and the eye gaze condition. I had thought participants would rate SARAH as more natural in the body and eye condition than in the speech condition because SARAH moved her eyes and her body. Both distinctively human-like traits. However, I think part of the reason the body and eye condition had the lowest average rate naturalness for SARAH might have been because the body movements were not expected. For eye gaze, I prepped participants by having SARAH move her eyes during the introductory conversational part of the experiment. I did not do this for the body movement, however, and SARAH only moved her body during the task. One participant in the body and eye condition actually screamed a little when SARAH moved her body. A separate participant also wrote in their survey that, "The sudden movement startled [them]".

Although SARAH was not rated as being more natural in the eye gaze condition versus in the speech condition, the mean % rated naturalness for SARAH in the eye gaze condition was still the highest (54%). Knowing that SARAH was rated as more natural in the eye gaze condition than in the body and eye condition also allows us to know that certain factors (such as abrupt body movement) may hinder humans from responding positively to robots in human-robot interaction.

6 Conclusion and Future Work

So, does the addition of eye gaze to human-robot interaction result in participants distinguishing between similar objects faster? And does this addition of eye gaze lead to participants reporting higher levels of comfort and engagement, and do they rate the robot as more natural? While my hypotheses concerning response time were disproven (H1 and H1a), I am confident that if I had more participants there would have been a statistically significant difference between the mean time differences by the conditions for the FC targets. Unfortunately, I did not have these participants and more research must be done before truly conclusive measures can be drawn regarding response times. As of this study, however, the addition of eye gaze does not seem to help participants distinguish between similar objects faster.

When it comes to comfortability, engagement and how natural participants felt SARAH was, there was a mix of results. Participants found their interaction with SARAH to be more comfortable when they were in the eye gaze conditions (eye gaze and body and eye) than in the speech condition. The addition of eye gaze did not lead to participants feeling more engaged while interacting with SARAH and participants rated SARAH as being more natural in the eye gaze condition than in the body and eye condition. Thus, should programmers and manufacturers want humans to feel comfortable while interacting with a social robot, the addition of eye gaze would be beneficial.

Because the ultimate goal for SARAH in the Computer Science department at Union is to have her successfully interact in completing a goal through the help of a stranger(s), there are still experiments left to conduct in order to improve SARAH's social skills. While two of my hypotheses were found to be null (HA and HAa), my study did show that humans feel more comfortable in their interaction with a robot when the robot exhibits eye gaze cues. If we were to take the study results and apply them practically, which we should, that would mean we should incorporate eye gaze behavior onto SARAH permenantly in order for humans to feel more comfortable while interacting with her. This is especially important as it would encourage humans to not run away from SARAH.

My study showed eye gaze can make a human feel more comfortable while interacting with a robot, but this was done in a laboratory setting and participants willingly walked into the experiment knowing they would be interacting with a robot. Thus there is the question of, how can a robot introduce and successfully commence an interaction with a human subject outside of the laboratory setting? In a previous experiment where a robot similar to SARAH, LINDSEE, was tested in the wild, they found that some people ran away from the robot or screamed (Aaron Cass and Webb n.d.). This means that there is still work to be done in figuring out how SARAH might start an interaction without scaring off her potential helpers. Since eye gaze was found to make people feel more comfortable in human-robot interaction, it may the case that eye gaze could be a good assistant in helping a robot start an interaction with a human subject in the wild. The incorporation of eye gaze into attracting a human subject outside of a lab setting could be viable, as eye gaze is often considered a signal of interest in human-human interaction (Wirth et al. 2010). Since eye gaze led to participants feeling more comfortable, this would imply that eye gaze could be one of the factors that might lead to a successful interaction between SARAH and a human partner in the wild.

In terms of continuing future work related specifically to my experiment, I would like to conduct this experiment again but with more participants and edits to the experiment. In particular, I would like to increase the speed at which SARAH communicates. I would also like to figure out a better way for participants to identify the target object. Specifically, there are touch sensors that can be bought which connect to raspberry pi. In a future conductance of this experiment, it might be ideal to use these touch sensors instead. Thus, participants would be able to touch their hand to a spot directly under the object in question in order to identify the object. Another issue I would like to address if this study were to be conducted again pertains to trust. Of the participants I spoke to who consciously used SARAH's eye gaze to identify an object, most of them told me they still waited until SARAH spoke the verbal disambiguator before pressing the corresponding keyboard button. These participants explained their reason for doing so was because they weren't sure if SARAH was trying to trick them. Specifically, they worried that SARAH's gaze toward a certain shape would not be congruent with the shape indicated by the verbal disambiguator. Thus, I would have to figure out a way to combat this issue. Perhaps having SARAH tell the participant, before the task begins, "Hi, we will be working together to get this task completed, I am on your side" or something similar would encourage participants to trust SARAH's nonverbal cues.

Another possible form of future research, is to incorporate an eye gaze tracker into the experiment.

This was done in Hanna and Brennan's 2007 study, and would be essential in figuring out for what reason participants had a different response with SARAH, a robot, versus the responses Hanna and Brennan found in human-human interactions (Hanna and Brennan 2007). Doing so would also allow for a future study that would be more accurate in terms of mimicking Hanna and Brennan's (2007) study. This would then also allow for better comparison of results between this future study and Hanna and Brennan's (2007) (Hanna and Brennan 2007). Adding an eye tracker to this study would allow for more fine grain timing information, and an analysis of where participants are looking during their interaction with the robot could also be conducted.

Appendices

A Survey Questions

1. How comfortable did you feel while interacting with SARAH? Please circle the number that corresponds with how you felt, with 0 as very uncomfortable and 7 as very comfortable. What made you feel comfort/discomfort while interacting with SARAH?

2. How engaged did you feel while interacting with SARAH? Please circle the number that corresponds with how you felt, with 0 as very disengaged and 7 as very engaged. What made you feel engaged/disengaged while interacting with SARAH?

3. How easy/difficult was it to lay out the shapes given SARAH's direction? Please circle the number that corresponds with how you felt, with 0 as very easy and 7 as very difficult. What made it easier/more difficult while interacting with SARAH?

4. Please carefully consider the questions below and rate the items comparatively to one another.

- On a scale of 0-100 how natural would you rate humans?
- On a scale of 0-100 how natural would you rate the SIRI or a similar virtual assistance on a smart phone?
- On a scale of 0-100 how natural would you rate SARAH?

Why did you rate SARAH as you did?

5. What did you like/dislike about your interactions with SARAH?

B SARAH's Script

1. "Hello! My name is SARAH, we will be spending some time together today! I have some introductory questions to ask you before we start the task today."
2. Each participant will be asked two of the following:
   - "What do you think is better, magazines or newspapers?"
   - "Would you rather dwell in the sky or in the sea?"
   - "Do you prefer ice cream or froyo?"
   - "Do you prefer to fly or take a train?"
   - "Which do you think is better, the platypus or turtle?"
3. "Do you have any questions you'd like to ask me? We have time for one."
4. "Before we begin this task, I would like your help in setting up. This will be a practice trial run. To your right is a pile of shapes, please take the first shape and place it in the furthest left box. Thank you, now please place the second shape in the second to furthest left box..." (this will continue until all the boxes drawn on the template have a shape within them)(SARAH's eyes will also be moving in the same direction as her descriptions for eye gaze and body and eye conditions).
5. "In front of you are some shapes with dots on them. Each number on the keyboard relates to each of these shapes, as you can see labelled below me."
6. "I will be asking you to please press the appropriate button corresponding to the correct shape, I will be telling you each time which shape I would like indicated through the keyboard. Do you have any questions?"

   Then, SARAH will proceed with the task, given the following template that will be filled with the accurate shape description.
7. "Please press the button corresponding to the [color of shape] [shape]... " short pause where SARAH engages in no eye gaze/ eye gaze/ eye gaze and body shifting "... with [number of dots] dots."

   SARAH will say the above line 3 times for each shape setup.

   During the experiment, the shapes will be changed by the participant.
8. "Before we continue, I will have ask you to please remove the shapes on the template and place down the next 8 shapes. The first shape on the stack should be placed in the furthest left box. Now please place the next shape in the second to furthest left box.." (this will continue until all boxes drawn on the template have a shape within them) Thank you! Let's continue."

   SARAH will have the participant lay out the shapes a total of 4 times.
9. "Thank you for your assistance today, I hope you have a pleasant day! Please wait a second, we have a short survey for you to fill out. Thank you."

C Shapes Layout

Key:

Red(R), Green(G), Yellow(Y), Blue(B)
Square(S), Triangle(T), Circle (C)
Dots: 1-4
ex: a red square with 2 dots == RS2

1. **Shapes Layout 1: 2 NoC, 2NC, 1FC**
   - 1: YC3, 2: YC2, 3: GS1, 4: RT4, 5: BS1, 6: BS2, 7: RC4, 8: RT3
2. **Shapes Layout 2: 2 NoC, 2NC, 1FC**
   - 1: YC2, 2: BS3, 3: BS4, 4: GT1, 5: YC2, 6: RT2, 7: RT1, 8: OS1.
3. **Shapes Layout 3: 2 NoC, 2NC, 1FC**
   - 1: GS2, 2: GS3, 3: GT2, 4: RS1, 5: YC3, 6: RC3, 7: RC2, 8: GT3.
4. **Shapes Layout 4: 2 NoC, 1NC, 2FC**
   - 1: BS3, 2: RT2, 3: BS1, 4: YT2, 5:YT1, 6:GS2, 7: RT1, 8: BS2.
5. **Shapes Layout 5: 2 NoC, 1NC, 2FC**
   - 1: YT1, 2: RS4, 3: BC2, 4: BC3, 5: GS4, 6: YT2, 7:RS3, 8: GC2.
6. **Shapes Layout 6: 2 NoC, 1NC, 2FC**
   - 1: RS2, 2: YS1, 3: BC1, 4: GT2, 5:GT1, 6: RC1, 7:RS1, 8: BC2.
7. **Shapes Layout 7: 4NoC, 1NC, 1FC**
   - 1: GT1, 2: BT2, 3: BC1, 4: YS1, 5: YS2, 6: BT3, 7: RC3, 8: RS2.
8. **Shapes Layout 8: 4 NoC, 1NC, 1FC**
   - 1: RC2, 2: RC3, 3: YT1, 4: BC4, 5: GS1, 6: YS2, 7: YT2, 8: RT4.

References

Aaron Cass Eric Rose, Kristina Striegnitz and Nick Webb. *Determining Appropriate First Contact Distance: Trade-Offs in Human-Robot Interaction Experiment Design*. http://cs.union.edu. URL: http://cs.union.edu/~cassa/papers/IROS2015.pdf.

Ellsworth, Phoebe C., J. Merrill Carlsmith, and Alexander Henson (1972). "The stare as a stimulus to flight in human subjects: A series of field experiments." In: *Journal of Personality and Social Psychology* 21.3, pp. 302–311. ISSN: 0022-3514. URL: http://search.ebscohost.com/login.aspx?direct=true&db=pdh&AN=1972-22925-001&site=ehost-live.

Hanna, Joy E. and Susan E. Brennan (2007). "Speaker's eye gaze disambiguates referring expressions early during face-to-face conversation". In: *Journal of Memory and Language* 57.4. Language-Vision Interaction, pp. 596–615. ISSN: 0749-596X. URL: http://www.sciencedirect.com/science/article/pii/S0749596X07000174.

Hugot, VANESSA (2007). "Eye gaze analysis in human-human interactions". In: *Unpublished master?s thesis, CSC, KTH, Sweden*.

Karreman, Daphne E et al. (2013). "Picking favorites: The influence of robot eye-gaze on interactions with multiple users". In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, pp. 123–128.

Khoramshahi, Mahdi et al. (2016). "Role of Gaze Cues in Interpersonal Motor Coordination: Towards Higher Affiliation in Human-Robot Interaction". In: *PLoS One* 11 (6). URL: `http://search.proquest.com/docview/1795489628/fulltext/ADDD7B38EB3040B5PQ/1?accountid=14637`.

Kleinke, Chris L. (1986). "Gaze and eye contact: A research review." In: *Psychological Bulletin* 100.1, pp. 78–100. ISSN: 0033-2909. URL: `http://search.ebscohost.com/login.aspx?direct=true&db=pdh&AN=1986-27160-001&site=ehost-live`.

Kuno, Yoshinori, Dai Miyauchi, and Akio Nakamura (2005). "Robotic Method of Taking the Initiative in Eye Contact". In: *CHI '05 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '05. Portland, OR, USA: ACM, pp. 1577–1580. ISBN: 1-59593-002-7. URL: `http://doi.acm.org/10.1145/1056808.1056970`.

Mandal, Fatik Baran (2014). "Nonverbal Communication in Humans". In: *Journal of Human Behavior in the Social Environment* 24.4, pp. 417–421.

Manusov, Valerie L. and Miles L. Patterson (2006). *The SAGE handbook of nonverbal communication*. English. Thousand Oaks, Calif: Sage Publications. ISBN: 1412904048;9781412904049;

"Nonverbal Communication" (2014). English. In: *Encyclopedia of Critical Psychology*.

Wirth, James H. et al. (2010). "Eye Gaze as Relational Evaluation: Averted Eye Gaze Leads to Feelings of Ostracism and Relational Devaluation". English. In: *Personality and Social Psychology Bulletin* 36.7, pp. 869–882.