6-2023

The Effects of Robot Eye Gaze on Task Performance

Haley Rose Kresch
Union College - Schenectady, NY

The Effects of Robot Eye Gaze on Task Performance

By

Haley R. Kresch

* * * * * * * * *

Submitted in partial fulfillment

of the requirements for

Honors in the Department of Computer Science

UNION COLLEGE

June, 2023

Abstract

KRESCH, HALEY R. The Effects of Robot Eye Gaze on Task Performance. Department of Computer Science, June, 2023.

ADVISOR: Kristina Striegnitz

Human-Robot interaction is continually becoming more and more integrated into everyday life. This study aimed to see how meaningful eye gaze from a robot with a 2D face affects human-robot interaction when completing an object identification task. An experiment was performed to see whether directional eye gaze from a robot influences the time is takes a human participant to identify the correct shape in the correct color zone based on a verbal instruction. This was measured by comparing the mean response times across three different gaze conditions. Another goal of the study was to see if using human-like eye gaze during the introduction resulted in participants learning to use gaze cues faster. This was measured by comparing the learning curves in response times in two different introduction conditions. At the conclusion of the experiment, participants filled out a survey to indicate how useful they felt the eye gaze was and to describe what they noticed about the gaze. The study found that congruent eye gaze (eye gaze that aligned with the verbal cue) led participants to respond faster. Human-like gaze in the introduction did not affect how quickly the participant learned to use the gaze cues.

Contents

List of Figures

1 Introduction

A significant amount of research has been conducted involving human-robot interaction (HRI) and perceived eye contact. Higher amounts of anthropomorphism (human-likeness) in robots have been attributed to higher levels of comfort around a robot and higher levels of receptiveness by humans to robots. The majority of studies have been focused on how humans perceive the human likeness of the robot when the robot establishes eye contact, rather than not establishing eye contact. Perceived eye contact in human-robot interactions is a growing area of study, and its implications on the future of HRI may be profound. Humans use eye contact for many reasons, including to establish trust, communicate with those around them, signal where they want their listeners to look, and portray emotion. Consistent eye contact can improve the perceived trustworthiness of a person.

Admoni et al. [1] found that meaningful eye gaze made a difference in human-robot interaction. They performed an experiment involving an object identification task with a human participant and a social robot. The robot gave both an eye gaze cue and a verbal instruction to help the participant identify the correct shape. The social robot was able to either help or mislead the human participant through directional eye gaze movements. They found in their experiment that participants identified the correct object faster when helpful directional eye gaze was present. Kompatsiari et al. [3] found that humans were more engaged when a robot establishes a mutual gaze. After a collaborative task involving human-robot interaction, participants rated feeling more engaged with a robot when the robot established mutual gaze than when they did not.

Both the Kompatsiari et al. [3] and Admoni et al. [1] experiments present interesting questions for further research. The Union College Social Robot VALERIE (as shown in Figure 1) is a 3D robot with a 2D face that is able to give verbal instructions, move, and shift eye gaze and facial expression. An example of a neutral gaze and facial expression is shown in Figure 2. The robot in the Admoni et al. [1] experiment was 3D, which lead to the question the study was based on: Can the results of the Admoni et al. [1] experiment be replicated with the Union College Social Robot VALERIE? The results of the Kompatsiari et al. [3] experiment showed that humans are more engaged when a robot tries to establish a mutual gaze. If VALERIE uses human-like gaze behavior during the introduction dialogue, does that help participants learn to use directional gaze quicker? Or will participants learn to use the eye gaze at the same speed regardless of whether they saw it in the introduction?

This study investigated whether directional eye gaze from the Union College Social Robot VALERIE is able to increase the speed at which a human participant is able to complete an object identification task. Admoni et al. [1] found that participants learn to use congruent gaze over the course of a few trials. Kom-

Figure 1: Union College Social Robot VALERIE

patsiari et al. [3] found that human participants are more engaged when a robot uses human-like gaze. This experiment also investigated whether interacting with a robot that uses human-like gaze before the object identification task had an effect on how quickly participants learned that they can use the robot's eye gaze. Performance was measured by the time is took the participants to complete each trial of the task (explained further in Section 3), as well as the time is took them in the first 9 trials of the experiment (the block representing "the learning curve"). A post-experiment survey was conducted to ensure participants were aware of the eye gaze and gauge their feelings about it.

Based on the results found by Admoni et al. [1] and Kompatsiari et al. [3], there are two hypotheses for this experiment:

1. Participants will identify the correct object faster when directional eye gaze is helpful (compared to when it is not).

2. If the robot uses human-like gaze before the object identification task begins, participants will learn to use the helpful directional eye gaze faster than if they did not see any gaze cues in the introduction.

The remainder of the paper is organized as follows: Section 2 gives a review of previous research related to Human-Robot Interaction and directional eye gaze. Section 3 describes the methods and design of the experiment. Section 4 explains the setup of the experiment, along with how participants were collected and the technology involved in the study. Sections 5 gives a detailed analysis of the results. Section 6 discusses the implications of the results, limitations to the study, and ideas for further research. Section 7 summarizes the study and the overall findings.

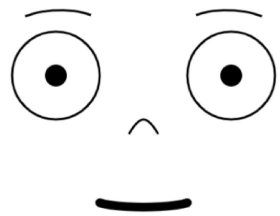


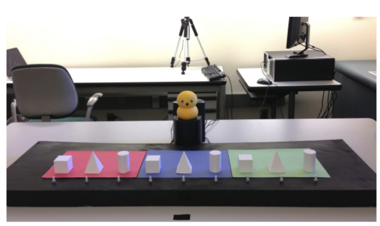
Figure 2: VALERIE Facial Expression

2 Background and Related Work

Eye contact is a distinctly human action, and therefore perceived eye contact with a robot would presumably increase the anthropomorphism of the robot. Kuchenbrandt et al. [4] researched whether humans would exhibit lower levels of EDA (electrodermal activity) when interacting with a human-like (anthropomorphic) robot versus a machine-like robot. EDA is a physiological indicator of psychological stress. They found that anticipating an HRI interaction in general increased EDA and that the choice of robot prototype did not significantly affect EDA stress reactions. Measuring stress in this experiment did not prove to be an efficient measure of how humans react to robots that are more anthropomorphic or not. In this experiment, the robot will attempt to be more anthropomorphic by establishing eye contact with the user. But rather than use stress levels as a measure of how this eye contact is perceived by the human, this experiment measured whether human participants completed a task faster with helpful directional eye gaze from VALERIE.

Task-related robot eye gaze has been shown to improve the efficiency of collaborative tasks between humans and robots [2]. There is also a noticeable crossover between the area of the brain that processes theory of mind and the area of the brain that processes eye gaze. Pierno et al. [5] found that "observing someone signaling the presence of an object with referential gaze elicits the same neural response as observing someone physically reaching to grasp that object". This indicates that eye gaze is a very useful tool in indicating objects, potentially even as useful as physically pointing them out.

This experiment follows in the footsteps of the research done by Admoni et al. [1]. In their work, they focused on the effects of conflicting gaze cues in a robot's multi-modal communication. They particularly studied when a robot's verbal instructions did not align with its gaze. Their experiment surrounded an object identification game with a robot. The robot verbally instructed the human participant to select the given shape in the given color section, while the robot directed its eye gaze at the target section, a different



(a) Robot agent condition

Figure 3: Admoni et al. (2014) Experiment Setup

color section, or directly forward. Figure 3 shows the experiment setup, with the participant sitting across from the robot. Their experiment included two conditions - a Human Agent condition and a Robot agent condition. Both the human and robot conditions were given the same instructions and changed their eye gaze based on the given trial. Both the robot and human conditions with the congruent gaze lead the participant to identify the shape quicker than with other types of eye gaze. This Capstone experiment aimed to see whether the Union College social robot VALERIE could elicit similar responses. VALERIE can move the position of its pupils in the eye gaze, simulating an eye gaze that a human participant could pick up on. A neutral eye gaze is shown in Figure 2.

Kompatsiari et al. [3] found that human participants are more engaged when a robot establishes mutual eye gaze. Their study was focused on the importance of mutual gaze in human-robot interaction. They conducted an interactive experiment between a humanoid robot called "iCub" and human participants. Participants were engaged in a "target identification task" where the iCub either established a mutual gaze (looking at the eyes of the participant) or a neutral gaze (looking down). The experiment started with the robot establishing one of the two types of gaze, then peering at one of the screens positioned around it. The target letter was displayed on the screen (either a V or T). Participants identified the letter as quickly as they could by clicking on a mouse. They found that mutual gaze led to higher ratings of engagement with the robot. In this Capstone experiment, an introduction scenario with dialogue was included that allowed VALERIE to establish either a human-like gaze or no gaze before the beginning of the experiment. The idea behind the human-like gaze in the introduction was to see if participants would learn to use the gaze cues in the trials quicker because they had already seen the gaze cues.

3 Methods and Design

Due to technical limitations, VALERIE was not used during the study. VALERIE's face (as shown in Figure 2) was displayed on a monitor placed on a cart to simulate VALERIE's height. The limitations of not using VALERIE will be discussed further in Section 6. The cart was tall enough so participants could stand while

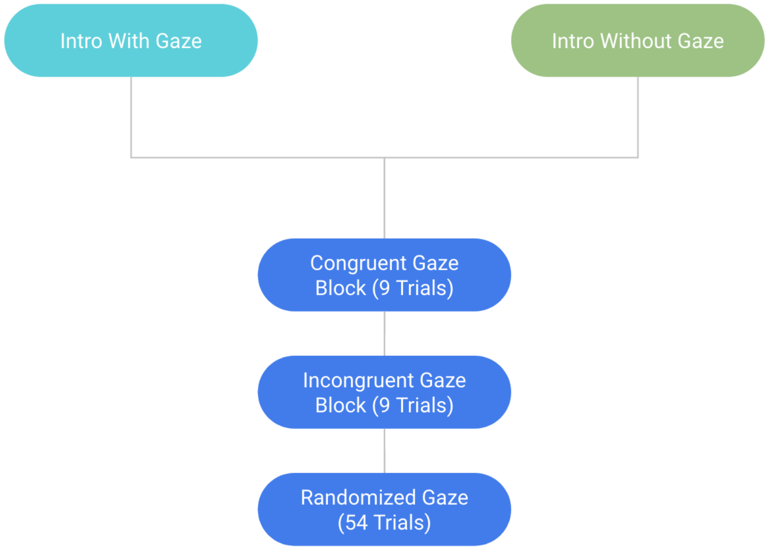


Figure 4: Order of Events in Experiment

engaging in the task, mimicking what they would have done if VALERIE was operational. Participants entered a room and were instructed to stand in front of the setup shown in Figure 5. The robot began the introduction dialogue when the participant stood in front of it.

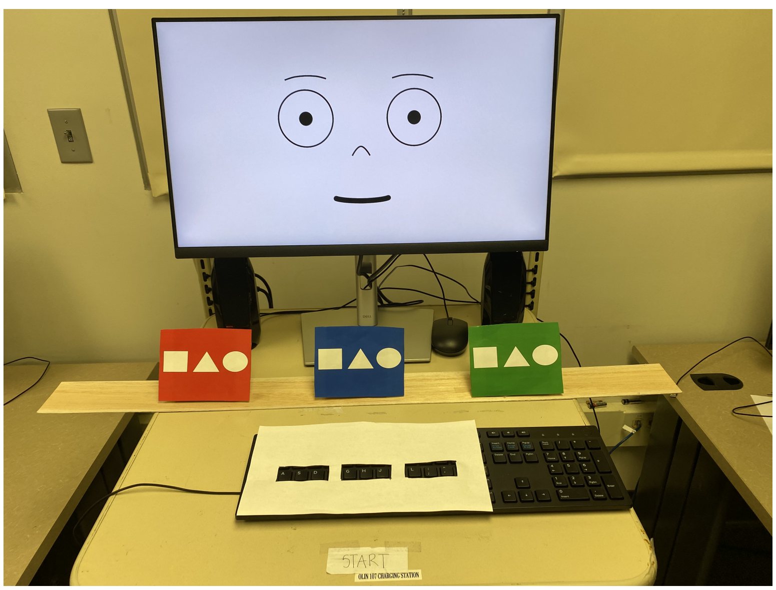


Figure 5: Experiment Setup

The gaze conditions for the introduction included:

- **Human-Like:** Imitating human-like behavior, including looking straight ahead and looking to objects

as they are being spoken about

- **No Gaze:** Gaze remains straight ahead, no movement throughout the opening dialogue (as shown in Figure 2)

To start, participants were randomly assigned to an introduction condition. Regardless of the introduction condition, the verbal instructions were the same. The robot began by introducing itself. The initial introduction was as follows:

"Hello, my name is Valerie and I am the Union College Social Robot. Today I will be asking you to choose the shape in front of you based on my verbal instruction."

The robot then proceeded to explain the correspondence between keyboard keys and shapes. For example:

"To select the red square, press A".

The keys lined up as much as possible with the given shape (see Figure 5). In the introduction with gaze, the robot looked at the shapes as it talked about them. When it talked about the red square, it looked toward the red color section (and the same for blue and green). After explaining to the participant which key they will press for each shape, it reestablished a neutral gaze (looking straight ahead) and instructed the participant:

"Please select the shape as quickly and accurately as possible. In between trials, please return your hand to the marked start position in front of the keyboard to signal you are ready for the next instruction."

After it explained the task, the robot asked the participant two questions:

"Have you ever interacted with a robot before?"
"What movies have you watched with robots in them?"

The point of the questions was for the robot to establish a "human-like" gaze. In the introduction with gaze cues, the robot looked off to the side (the pupils moved up and left) to indicate it was thinking and then moved back to neutral when it concluded asking the question. Because humans don't simply stare at each other, the idea of moving the eye gaze while asking the question was to attempt to mimic a human looking off to the side as they think about what they are saying. It also served as another way of establishing eye gaze before the experiment began. The hypothesis was that seeing a "human-like" gaze during the introduction could lead participants to learn to use the gaze cues faster during the task.

In the introduction without gaze, the pupils did not move at all. There was no indication to the participant that the gaze could or would move. They received exactly the same instructions and questions as the introduction with gaze condition.

The object identification task began immediately after the introduction. VALERIE used 3 different gaze cues throughout the object identification task.

Gaze Types for Object Identification Task:

- **Congruent Eye Gaze:** Speech and instructions aligned with normal human interaction. The robot looked toward the color section being spoken about
- **Incongruent Gaze:** The robot intentionally misled the participant by looking at a different color section than what they were speaking about
- **No Gaze:** No gaze cues were used in this condition. Gaze remained straight ahead

The flow chart in Figure 4 demonstrates the order of the experiment. After the introduction scenario, the participant went through a block of congruent trials and a block of incongruent trials. Both blocks were randomized separately, and each participant saw the same randomized order. Each block consisted of 9 trials. After both block trials, the participants went through 54 randomized trials, with an even mix of congruent, incongruent, and no gaze cues. The random trials evenly covered all of the shapes in each color section. The randomization was the same for each participant, and the trials were randomized using the Python shuffle function.

In each trial, the gaze cue went slightly before the verbal cue (as shown in Figure 6).

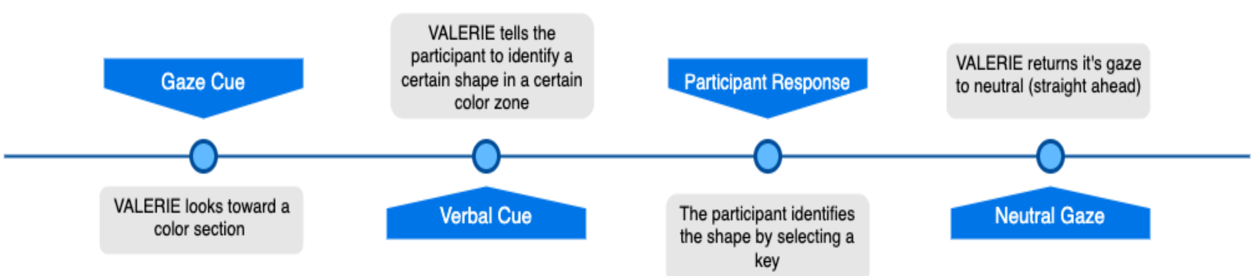


Figure 6: Verbal and Gaze Cue Timings

In between trials, participants (as they were instructed at the beginning of the experiment) returned their hand to the marked start position in front of the keyboard to signal they were ready for the next instruction. After the random trials concluded, the participant was instructed to exit the room. They completed a post-experiment survey to gauge their experience with the robot. They were asked to rate how helpful the eye

gaze was during the task on a scale of 1 to 5 (1 being not helpful, 5 being almost always helpful). They were also asked an open-ended question about what they noticed about the eye gaze.

4 Preparation

4.1 Participants

A SignUp Genius form was sent out to the student body of Union College. Participants voluntarily chose to participate in the experiment and were compensated via Venmo for their participation. Participants covered a wide range of majors and class years. The initial goal for the experiment was 44 participants, doubling the amount used in the Admoni et al. [1] experiment. The reasoning behind doubling the number of participants was to have the same number of participants but within both introduction conditions. Due to time constraints, the ultimate number of participants was 29. This still was ultimately enough to get statistically significant results. Participants were informed that the study would last approximately 30 minutes, though the majority of participants did not need the entire 30 minutes. Participants filled out an informed consent form before they began the task or interacted with the robot.

4.2 Setup

Though the object identification task mimicked the Admoni et al. [1] experiment, there were a few ways in which the setup was different. Because of the time and resource constraints involved in setting up the experiment, the shapes were 2D rather than the 3D ones used in the Admoni et al. [1] experiment. To identify the correct shape in the correct color zone, participants selected keys on a keyboard. The keyboard was covered everywhere other than the keys involved in the experiment. During the experiment, participants were instructed which key to select to identify each shape. There was a distinguished white area in front of the keyboard with the word ”start” written on it. In the instructions, VALERIE tells the participant to return their hand to the marked start position in front of the keyboard to signal they are ready for the next instruction.

Figure 7: Graphical User Interface

```python
def neutral_gaze(self):
    gaze= rospy.Publisher("/face_listener", String, queue_size=10)
    gaze.publish("gaze::0::0::500::500::500")


def speech(self):
    speech = rospy.Publisher("/face_listener", String, queue_size=10)
    speech.publish("say::Hi! My name is Valerie.")
```

Figure 8: ROS Gaze and Speech Code

4.3 Technology

The robot was controlled using ROS (Robot Operating Software). Messages indicating the verbal and eye gaze cues were sent over "topics" (essentially channels) that the robot was subscribed to. The functions used in the experiment were written in Python using the RosPy module. To send a message, a publisher needs to be created that indicates what the message is and what topic it is being sent over. For this experiment, the topic was "/face_listener". An example of speech and gaze cues using RosPy functions is shown in Figure 8. Gaze and speech cues are sent in the same way, just with slightly different syntax to indicate what the robot should do. In the case of gaze, the coordinates indicate where the eye gaze will look. This was manipulated to create the appearance that the robot was looking at the color sections.

The gaze cues were sent by a human researcher through a Graphical User Interface (GUI). The buttons on the GUI called on functions similar to the ones demonstrated in Figure 8. The opening screen of the GUI is shown in Figure 7. The timing of each response was the difference between the sending of the gaze cues

and the participant's key press.

5 Results

It's important to note that participants rarely selected incorrectly based on verbal cues. There were only 15 incorrect selections from 1566 random trials, meaning that participants identified the correct object 99% of the time. Incorrect selections were excluded from the reaction times. Of the incorrect selections, over half of them (8) occurred during the incongruent gaze condition. The other 7 split between the no gaze condition (4) and the congruent gaze condition (3).

The first predicted outcome was that participants will be able to identify the correct shapes faster when there was a congruent gaze present. The results of the Admoni et al. [1] experiment found that participants had shorter response times when a congruent gaze was present.

Gaze Type	Mean and SD of Random Trials(ms)
Congruent	2865(266)
Incongruent	3005(272)
No Gaze	2971(260)

1) $p<0.001$ 2) $p=0.001$

Figure 9: Mean Response Times and SD of of Random Trials

This study had similar findings. Figure 9 shows the mean response times to the random trials for each gaze type. A repeated measures ANOVA test was used to evaluate the effect of gaze type on response time. The test found a significant main effect ($F=15.7$, $p< 0.001$). Post hoc tests with a Bonferroni correction were performed to look for individual differences between the gaze types and their effect on response time. It showed that congruent gaze had shorter response times than incongruent gaze ($p<0.001$), with a mean response difference of 1398ms. It also showed that congruent gaze led to shorter response times than the no gaze condition ($p=0.001$), with a mean difference of 1067ms. There was no statistically significant difference in the response times between the incongruent and no gaze conditions.

The way this project differed from the work of Admoni et al. [1] was by introducing gaze cues in the introduction. The goal of the gaze cues in the introduction was to see if participants that had already seen the gaze cues would have a steeper learning curve in response times (meaning they learn to use the gaze cues quicker) than those who didn't see gaze cues in the introduction. Alternatively, they might have had a flatter learning curve, but one that started at a faster time. This would mean that they had already learned

to use the gaze cues from the introduction. But in either case, the goal was to find a difference between the curves depending on the introduction condition. Figure 10 shows the learning curves of the first 9 trials based on introduction type. A repeated measures ANOVA test with introduction type as the between-subjects factor found no statistically significant difference between the response times over 9 trials based on introduction type. It did highlight that participants continually responded faster over the course of the 9 trials regardless of introduction type (which was also observed by Admoni et al. [1]), with a noticeable main effect ($F=7.91$, $p<0.001$). It makes sense that participants would have shorter response times the more they see the gaze cues, but the study was not able to show that introducing gaze cues before the experiment begins leads to participants picking up on the gaze cues even faster.

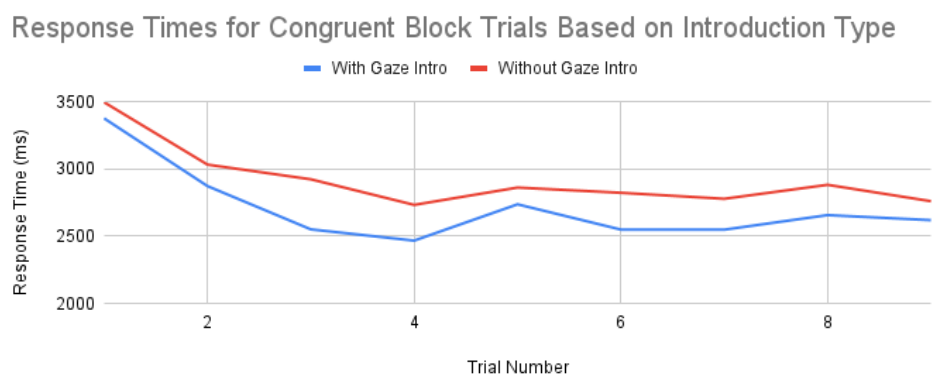


Figure 10: Learning Curve of Congruent Block Trials Based on Introduction Type

The post-experiment survey indicated that participants picked up on the change between congruent, incongruent, and no gaze cues. Participants were asked “How often did you feel that the robot’s eye gaze was helpful in identifying the object?”. Participants were prompted to answer on a scale of 1-5, with 1 being “it was never helpful” and 5 being “it was almost always helpful”. As shown in Figure 11, the majority of the responses were around a 3. This is to be expected, considering the gaze cues were split between congruent, incongruent, and no gaze. Breaking the responses by introduction type did not produce any statistically significant results.

Participants were also asked in the post-experiment survey “What did you notice about the robot’s eye gaze?”. 28 of 29 participants indicated that the gaze was not always helpful. A few participants noted feeling “tricked” by the gaze. Ten participants noted how the gaze was helpful at first and then changed. Considering the first 9 trials of the experiment were congruent, it makes sense that participants would feel that the gaze moved from being helpful to unhelpful. Most participants indicated that the gaze was a mix between helpful and unhelpful. The responses to both survey questions indicate that participants noticed the eye gaze. This backs up the results of the random trials, which indicate that the eye gaze affected the

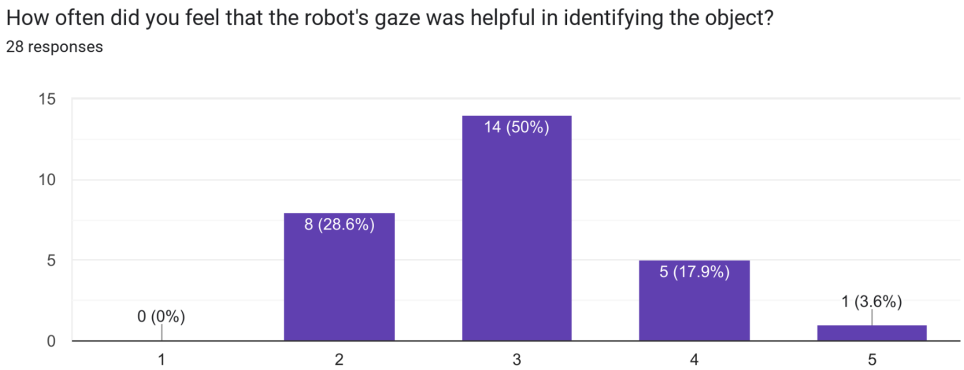


Figure 11: Responses to Post-Experiment Survey

speed with which they responded to the verbal instruction.

6 Discussion

It's necessary to note an obvious limitation in the study. The inability to use VALERIE was not anticipated while the study was developed. Figure 12 shows what VALERIE looks like compared to what participants saw. Although the best possible effort was made to mimic what participants would have seen with VA-

Figure 12: The Setup Compared to VALERIE

LERIE, there is no way to know how this affected the results.

Using gaze cues in the introduction did not prove to have a significant effect on the speed in which participants learned to use the gaze cues. Whether or not the participants saw gaze cues in the introduction did not affect the speed at which they learned to use them in the object identification task. There are a number of outlying factors that may have contributed to this. For one, the participant might be consumed

with listening to the instructions and the questions that they barely notice the eye gaze. There is also the subjectivity of the "human-like" gaze. VALERIE's face could not interpret the human's face in front of them, meaning there was no way to establish eye contact. Moving the pupils off to the side while asking the questions was an attempted way to mimic human gaze cues. Participants may not have interpreted this as human-like. In future research, it would be interesting to see if using a 3D robot face with the human-like introduction would lead to a steeper learning curve compared to no gaze in the introduction. Human-likeness would be easier to replicate with a 3D robot.

Human-Robot interaction is continually becoming more and more integrated into everyday life. One of the settings that can benefit the most is the classroom. This experiment showed how a 2D robot face can influence participants in a similar way to a 3D robot face. Projecting a 2D robot face on a monitor is significantly more accessible than using a 3D robot. This could have beneficial implications for the effectiveness of robot tutoring coming from a 3D robot. In the future, it would be interesting to use a 2D robot for other collaborative tasks, such as language acquisition or subject-specific learning tasks.

7 Conclusion

This study aimed to see how meaningful eye gaze from a robot with a 2D face affects human-robot interaction when completing an object identification task. An experiment was performed using the face of the Union College Social Robot VALERIE to see whether directional eye gaze from a robot influences the time is takes a human participant to identify the correct shape in the correct color zone based on a verbal instruction. The first hypothesis was that participants would identify the correct object faster when directional eye gaze was helpful (meaning VALERIE looked toward the correct color zone). The second hypothesis was that if the robot uses human-like gaze before the object identification task begins, participants will learn to use the helpful directional eye gaze faster than if they did not see any gaze cues in the introduction. The experiment found that participants did identify the correct shape faster when the directional gaze from VALERIE was helpful. The introduction condition did not make a significant difference in how quickly participants learned to use the helpful gaze. The post-experiment survey found that participants picked up on the change between congruent, incongruent, and no gaze cues.

References

[1] Henny Admoni, Christopher Datsikas, and Brian Scassellati. "Speech and gaze conflicts in collaborative human-robot interactions". In: *Proceedings of the annual meeting of the cognitive science society*. Vol. 36. 36. 2014.

[2] Jean-David Boucher et al. "I Reach Faster When I See You Look: Gaze Effects in Human–Human and Human–Robot Face-to-Face Cooperation". In: *Frontiers in neurorobotics* 6 (May 2012), p. 3. DOI: 10.3389/fnbot.2012.00003.

[3] Kyveli Kompatsiari et al. "The Importance of Mutual Gaze in Human-Robot Interaction". In: *ICSR*. 2017.

[4] Dieta Kuchenbrandt, Nina Riether, and Friederike Eyssel. "Does anthropomorphism reduce stress in HRI?" In: *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. 2014, pp. 218–219.

[5] Andrea Pierno et al. "When Gaze Turns into Grasp". In: *Journal of cognitive neuroscience* 18 (Jan. 2007), pp. 2130–7. DOI: 10.1162/jocn.2006.18.12.2130.