Eindhoven University of Technology

MASTER

Exploring the Use of Anthropomorphized Gaze and Movement to Signal Intention in an Industrial Robot Arm

Stolle, Kaj H.P.

Award date:
2023

[Link to publication](#)

Disclaimer
This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain

Exploring the Use of Anthropomorphized Gaze and Movement to Signal Intention in an Industrial Robot Arm

Kaj Hans Peter Stolle

1256270

Department of Industrial Engineering and Innovation Sciences

Eindhoven University of Technology

In partial fulfilment of the requirements for the degree of

Master of Science
in Human Technology Interaction

July 13th 2023

Supervisors:

Sanne Schoenmakers

Haodong Zheng

Raymond Cuijpers

Abstract

This study focused on improving collaboration between industrial robot arms and their operators. A key factor in this collaboration is signalling intention from the robot. By making it more clear what object a robot arm is reaching for, collisions between the robot and its operator can be avoided. This makes for a safer and more efficient working environment.

We explored the use of gaze cues (implemented as humanlike eyes looking at their target) and directed movement (implemented as the robot arm pointing its gripper at its target) as ways for a robot arm to signal intention. While previous studies have demonstrated the positive impact of such cues on task performance and user experience in social and humanoid robots, their application in industrial settings remains limited. However, we expected that we can replicate these positive effects with a robot arm.

This was tested through a collaborative task in virtual reality. In this task, the robot arm picked up coloured objects while participants indicated which object they anticipated the robot arm to be reaching for. It was measured how quickly participants could identify the target before the robot arm reached it. User experience was also measured through the Godspeed questionnaire. Open questions and forced choice preference questions were used to further explain the results. Uncanniness was also measured as a potential confounding effect.

Results showed that both gaze cues and directed movement similarly increased task performance. User experience was improved more by the gaze cues than the directed movement. The best results for both were attained when gaze cues and directed movement were combined. The open and preference questions showed that there were differences between participants on which robot arm they preferred and why.

Based on these findings, industries seeking to increase the efficiency of collaboration between robot arms and their operators should adapt their robot arms to use both directed movement and gaze cues. However, using just directed movement can already prove to be of great benefit with a smaller associated cost than implementing eyes.

Contents

Exploring the Use of Anthropomorphized Gaze and Movement to Signal Intention in an Industrial Robot Arm

Industrial robots are increasingly used to take over tedious tasks from their human counterparts. Nevertheless, human operators are still required to supervise the workings of these machines. This new collaboration between humans and robots does not always proceed without problems. Industrial robots are often separated off from their human operators due to the strength they possess. While this is necessary for safety, it can also cause a great loss of efficiency. Despite engineers' best efforts, it is a common occurrence for industrial robots to respond unexpectedly. For instance, by breaking objects, tipping over stacks, or stopping to work altogether. The steps that need to be taken by a human operator to get close to the scene around the robot all come at the cost of time and money. It is therefore no surprise that many industries seek to enhance the collaboration between machines and their operators. In order to achieve this in practice, it is important to understand the factors that make this collaboration safe, efficient and satisfying for the user.

For an efficient collaboration, both parties need to be aware of each other and know what the other is planning to do next. In the current study we focus on the human operator's understanding of the robot's intentions. Understanding the robot's intentions leads to a safer and more efficient collaboration. Firstly, by avoiding collisions, and secondly by making it easier for the operator to determine what task best fits the robot's current intentions.

In this study we draw insights from two main branches of human-robot interaction (HRI): social HRI and industrial HRI. Where industrial HRI tends to focus on the efficiency of collaboration, social HRI instead tends to focus on closeness, likeability, and connection to (usually humanlike) robots. This study is not the first to argue for bridging the gap between these two domains (Hostettler, Mayer, & Hildebrand, 2022; Onnasch & Hildebrandt, 2022). The goal being to improve the quality of collaboration with industrial robots through the application of concepts from social HRI.

Specifically, this study investigates the use of movement and gaze in an industrial robot arm on both the efficiency and user experience of the collaboration. Gaze and movement are some of the primary means of conveying attention between humans. By looking or pointing at an object it immediately becomes clear what one's attention is focused on. This concept has been studied for use in social HRI as well. However, it remains unclear how this translates to industrial robots, as research on social industrial robots is very limited. There are large differences in both form and function between social and industrial robots. These different embodiments require different types of social behaviour. In the current study we will investigate whether, similarly to a social robot, an industrial robot arm allows for movement and gaze to be used as an intuitive means to convey the robot's intentions.

Joint Actions

A concept at the heart of the current study is joint action. Joint action can be described as "any form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment" (Sebanz, Bekkering, & Knoblich, 2006). This coordination happens naturally between humans, and relies on multiple features. Key among them are the abilities to actively predict what actions one's interaction partner(s) is planning, how these actions influence the environment, and what responses should follow on these actions (Sacheli, Arcangeli, & Paulesu, 2018).

Whilst joint action occurs instinctively between humans, research in HRI has also investigated how to achieve the same common understanding between robots and humans (Bicho, Louro, Hipólito, & Erlhagen, 2009; Dağlarlı, Dağlarlı, Günel, & Köse, 2017). This poses challenges in two directions: the robot's understanding and prediction of human's actions, and the human's understanding and prediction of the robot's actions (Clodic, Pacherie, Alami, & Chatila, 2017). In the present research we seek to add to the literature on human's understanding of a robot's actions. Specifically, we investigate the using cues from gaze and movement, based on biological principles, to make the intentions of a robot arm more clear to the user.

Anthropomorphic Design

Robots are found to be much more intuitive to work with when they mirror behaviour that is naturally expected from humans (Fink, 2012). This is because humans subconsciously re-use the same mental models they use for interacting with humans in their interactions with robots. Designing robots to be perceived as having humanlike qualities is known as anthropomorphic design (Fink, 2012).

Anthropomorphism is the natural phenomenon in which people attribute humanlike qualities to non-human entities (Fink, 2012). Whilst this is a very broad concept, for HRI it mainly concerns the attribution of human mental capacities, like emotions, morality and goals, to a robot's appearance and behaviour (Kühne & Peter, 2023). The perception of a robot's shape and movement has been shown to play a significant role in the attribution of these qualities (Kühne & Peter, 2023). This knowledge is often used in the design of social robots, which are designed with humanlike features in mind. This design helps to give these robots more social presence, making it easier for humans to engage with them (Damiano & Dumouchel, 2018). Indeed, many studies have found that social robots are more likeable when they are more humanlike in their appearance, compared to non-humanlike counterparts (Rau, Li, & Li, 2010). Achieving this is not always easy though, as is seen with the 'uncanny valley' effect (Mori, 1970). This effect describes how people experience negative emotional responses towards robots that are perceived to be close, but not quite close enough, to actual humans in appearance. Avoiding the uncanny valley is a common subject in the design of robots (Walters, Syrdal, Dautenhahn, te Boekhorst, & Koay, 2008).

Nevertheless, theories regarding anthropomorphic design are mainly tested in the context of social robotics. Such research often utilizes robots that possess human-like features like a face, eyes and humanoid body. There is now a growing push to investigate these theories in industrial settings as well (Hostettler, Mayer, & Hildebrand, 2022; Onnasch & Hildebrandt, 2022), as most robots today are in use for industrial settings rather than social settings. A problem with anthropomorphic designs in industrial settings is that, while humanoid designs tend to be more likeable, they can also distract humans from their task, leading to worse performance (Onnasch & Hildebrandt, 2022; Perugia, Paetzel-Prüsmann, Alanenpää, & Castellano, 2021).

An additional idea for anthropomorphic robot design comes in the form of zoomorphism (Bergman, De Joode, De Geus, & Sturm, 2019; Sauer, Sauer, & Mertens, 2021). This is the idea that robot colleagues should seek to emulate the behaviours of animal-companions, rather than human behaviour. Since robot intelligence is not close to the level of intelligence of that of a human, and animals thus make for a more fair comparison.

Gaze in Conveying Robot Intention

Between humans, gaze is used to signal where attention is going. By staring at an object, one communicates intentions with that object. By communicating this intention to one's interaction partner, joint attention is established. This joint attention is a requirement for joint actions to take place (Siposova & Carpenter, 2019). The use of gaze to establish joint attention has been widely applied in the field of HRI as well. From one side in having robots interpret the gaze of humans (Palinko, Rea, Sandini, & Sciutti, 2016; Aronson et al., 2018), and from the other side in having humans interpret the gaze of robots (Admoni & Scassellati, 2017).

As stated earlier, the use of gaze in HRI has mostly been studied and applied in social rather than industrial contexts. One study that did fit an industrial context, namely a handover task with an industrial robot arm, tried implementing anthropomorphism as a factor by adding a flat screen with an image representing eyes to the robot arm. The goal of this study was to investigate how this anthropomorphization influenced the participant's trust in the robot. It concluded that there was no effect of these eyes on trust (Onnasch & Hildebrandt, 2022). However, the study also proposed that the lack of an effect could be due to the implementation of the 'eyes'. The screen showed an image of two eyes and eyebrows that moved around randomly, unrelated to the robot's actions. Accordingly, the study suggested that anthropomorphism should be instrumental, rather than being implemented as a feature that does not relate to the robot's functionalities (Onnasch & Hildebrandt, 2022).

Nevertheless, studies that fall outside of an industrial context can still provide useful insights regarding the use of gaze. Especially when measures of performance are used. For instance, Khoramshahi, Shukla, Raffard, Bardy, & Billard (2016) found that gaze cues from a virtual humanoid robot made it easier to anticipate a robot's movement in a mirroring task. Similarly, Boucher et al. (2012) found that gaze cues from a humanoid robot increased participants' reaction times in a target identification task. These contexts show that gaze cues, referring to a robot's face and eyes facing a target, can be used to make a robot's intentions more clear. The question remains whether this same effect can be achieved in an industrial setting. It is unclear what such gaze cues would look like in a robot arm.

Movement in Conveying Robot Intention

Another important feature in making a robot's intentions more clear is movement. This factor also plays a role in how intelligent and likeable robots are perceived (Kühne & Peter, 2023). In the context of industrial robot arms, people tend to attribute more human-like qualities to arms that move slower, use smooth curves over rigid movement, and use conventional rotations over unnecessary rotations (Hostettler et al., 2022). This perceived humanness is paired with an increase in likeability (Hostettler et al., 2022).

With regards to conveying intentions, 'minimum-jerk' trajectories have been found to be more efficient in expressing a robot arm's goal when compared to other strategies such as 'conventional velocity profiles' (Huber, Rickert, Knoll, Brandt, & Glasauer, 2008) or 'legible' and curved movements (Dragan et al., 2015; Zhao, Shome, Yochelson, Bekris, & Kowler, 2016). Jerk is defined as the rate of change of acceleration. A minimum-jerk trajectory thus minimizes the change in acceleration when moving the robot arm from one position to the other. Interestingly this is also the movement pattern associated with most human movements (Sharkawy, 2021).

Aside from just the factors that make up movement, movement as a whole can signal intention as well. An example for humans or humanoid robots is pointing with the arm, fingers or head. Terzioglu, Mutlu, & Sahin (2020) used principles from character animation to apply this to an industrial robot arm. Their results showed that the movement of a robot arm can be modified to give of the illusion of gaze. In this example, sunglasses were attached to the gripper of the robot arm, to have it represent a face. Likewise, the movement of the last three joints of the arm were manipulated to direct the gripper towards a point of interest. Together, this gave the appearance of the 'face' of the robot gazing towards a target. This form of directed movement resulted in an increase in the perceived likeability of the robot arm. However, although the experiment had the participants collaborating on a task with the robot arm, no measures of task performance were done. Thus, it remained unknown how directed movement affects the efficiency of the collaboration.

A preceding master thesis by Ronckers (2022) investigated a similar topic regarding the use of movement in an industrial robot arm. This study investigated both measures of task performance and user experience. The study had a robot arm picking up objects while participants were tasked to indicate, as quickly as they could, which object they expected the robot arm to pick up. Three conditions for the appearance and movement of the robot arm were compared. In the first condition the robot arm moved using a basic implementation of a minimum-jerk trajectory. In the second condition, a separate humanoid head gazed at the target object, while the robot arm used the same minimum-jerk trajectory. In the third condition, sunglasses were added to the robot arm to represent eyes, and movement used a directed motion where the arm first 'looked' at the target before reaching towards it, as was done by Terzioglu et al. (2020). The study found a preference (in terms of user experience) and better task performance (in terms of participants' reaction speed) for the design using sunglasses and directed movement. However, this study did not separate the conditions of movement type and the presence of sunglasses as 'eyes', so it remains unclear how each factor influenced these results.

Research Question

The current study seeks to build on the gaps found in the literature. Specifically, we investigate the combination of gaze cues and directed movement in signalling the intentions of an industrial robot arm. The study seeks to separate the effects of directed movement and presence of eyes found in Ronckers (2022). In addition, we investigate the potential use of gaze cues to communicate intention in a robot arm. These gaze cues will be presented through humanlike eyes attached to the robot arm. To our knowledge we are the first study to combine the factors of directed movement and gaze cues in an industrial setting. It is currently unknown whether gaze cues or directed movement are better in communicating intention in a robot arm. Likewise, we do not know in what manner these manipulations affect each other.

The set-up of our study was adapted from Ronckers (2022) in order to make our results more comparable. This meant that participants collaborated with a robot arm in a virtual reality environment. Virtual reality was chosen because it is able to simulate a more ecologically valid experience compared to 2D videos, and modifying the appearance of a virtual robot arm is easier than modifying a physical robot arm. In the collaborative task, the robot arm picked objects and the participant was tasked to indicate, as fast as possible, which object they expected the robot arm to pick up.

The robot arm was modified to include eyes near its gripper, meant to represent a face. In Ronckers (2022) and Terzioglu et al. (2020) this was done with sunglasses. Gaze cues were manipulated by having the eyes rotate to face the robot arm's target, similarly to Khoramshahi et al. (2016). The movement of the robot arm was manipulated to use either directed or undirected movement. The directed movement had the robot arm first move its 'face' to 'look' at its target, similar to Terzioglu et al. (2020). The undirected movement followed a more efficient and typical robot arm trajectory. Both movement conditions used minimum-jerk trajectories that were also used by Ronckers (2022).

We used two primary measurements to investigate this question. Firstly, an objective measure of the task performance for the human in the collaboration task. Secondly a subjective measure of the user experience of the participant. Task performance was measured as the time difference between participants correctly identifying the target of the robot arm, and the robot arm reaching its target. User experience was measured using the Godspeed Questionnaire (Bartneck, Kulić, Croft, & Zoghbi, 2009), which measures 5 broad dimensions of user experience. This questionnaire is very widespread, making it useful for comparing results across studies.

In addition to the two primary measures, we also measured how uncanny the robot was perceived using an uncanniness questionnaire (Ho & MacDorman, 2017). This was added to measure the potential confounding effect of uncanniness, which is known to negatively influence how robots are perceived (Mori, 1970; Ho & MacDorman, 2017). This is especially relevant since we are manipulating the robot arm to appear more humanlike. Furthermore, forced choice questions on which condition was preferred, as well as open questions on how participants described the collaboration, were included to provide a better explanation for the results of the primary measures.

We have two main research questions, one for each of our primary measurements:

RQ1: In a collaborative task between a human and a robot arm, how do directed movement and gaze cues affect the task performance of the human?

RQ2: In a collaborative task between a human and a robot arm, how do directed movement and gaze cues affect the user experience of the collaboration?

Earlier studies show that humanlike qualities in robots are strongly related to the likeability and positive user experience of the robot (Fink, 2012; Rau et al., 2010). Nevertheless, the literature also warns for the risk of the uncanny valley (Mori, 1970; Ho & MacDorman, 2017), and that humanlike qualities can distract operators in industrial settings, negatively impacting task performance (Onnasch & Hildebrandt, 2022; Perugia et al., 2021). However, this might only be a problem if the anthropomorphic features do not match the robot's functions (Onnasch & Hildebrandt, 2022), which we will avoid in our design. Our robot arm could also be described as zoomorphic (Bergman et al., 2019; Sauer et al., 2021), with the 'face' and arm not really representing a human, which could lessen the perceived uncanniness and its negative effects. Furthermore, the use of minimum-jerk trajectories has been shown to make the target of a robot arm appear more clear (Huber et al., 2008; Dragan et al., 2015; Zhao et al., 2016). Additionally, directed movement to represent gaze has been found to increase likeability as well (Terzioglu et al., 2020). Finally, gaze cues have successfully been used in social HRI settings to increase the efficiency of collaboration (Khoramshahi et al., 2016; Boucher et al., 2012). It is expected that gaze cues have a similar effect in the task performance of an industrial setting as well.

Based on all this, we hypothesize that both gaze cues and directed movement will yield an increase in task performance, as well as a higher rated user experience. To our knowledge, no studies have directly compared the effects of directed movement and gaze cues. However, based on the hypothesised positive effects of both, we expect that the best task performance and highest rated user experience will be found when directed movement and gaze cues are combined. Likewise, the worst task performance and lowest rated user experience is expected to be found when neither directed movement or gaze cues are used. We do not know which will yield the stronger effect between when only gaze cues or only directed movement are used. Although both are expected to perform better compared to when neither are used.

Method

Experimental Design

The experiment followed a 2x2 within-subjects design with counterbalanced conditions. The first manipulated dimension was the movement trajectory of the robot arm. This utilized either directed or undirected movement. This directed movement was meant to simulate the arm 'gazing' at the object before reaching for it (see Figure 1), whilst the undirected movement followed a more typical robot arm trajectory.

The second manipulated dimension was the movement of the eyes on the robot arm. These either used gaze cues or not. These gaze cues had the eyes move to face towards the object the arm was moving towards (see Figure 2). The non-gaze cues condition had the eyes looking statically ahead over the gripper.

The physical appearance of the robot arm was the same in each condition. Participants completed a collaboration task with the four different conditions in a virtual environment. After each condition they left the virtual environment to fill in a questionnaire on the experience of the interaction.

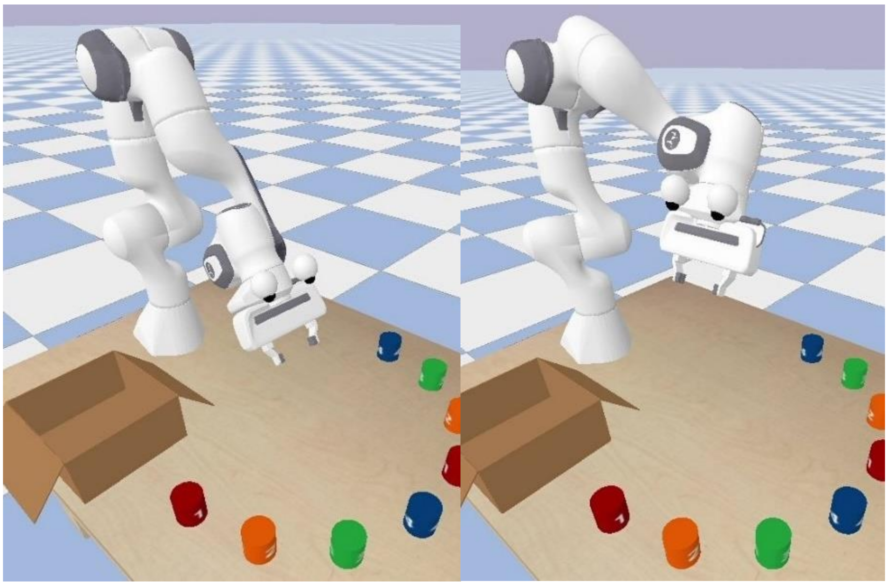

***Figure 1**: Comparison of the robot arm utilizing undirected movement (left) and directed movement (right). Both robot arms are heading towards the red cup on the far left and not using gaze cues.*

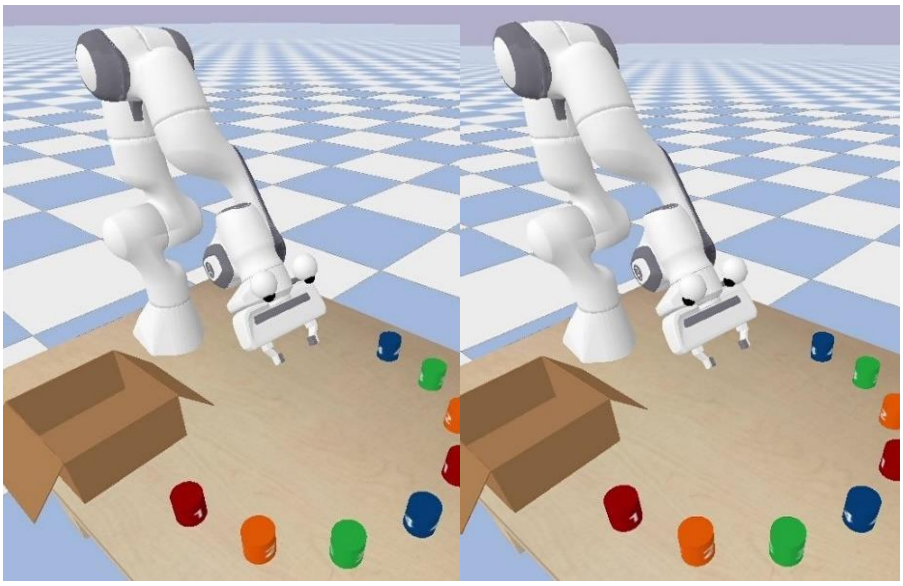

***Figure 2**: Comparison of the robot arm not using gaze cues (left) and using gaze cues (right). Both robot arms are heading towards the red cup on the far left using undirected movement.*

Power analysis

In order to estimate the required amount of participants, an a-priori power analysis was performed using G*Power version 3.1.9.4. It was set to find a medium effect size ($f > 0.25$), with $\alpha = 0.05$ and power = 0.90 for a within-subjects design with four conditions in a repeated measures ANOVA. The result of the power analysis showed that at least 30 participants would be required.

A medium effect size was chosen as smaller effect sizes would not yield enough benefit to make adding new components to a robot arm worthwhile in practice. Additionally, studies comparing the subjective aspects of robots using the Godspeed questionnaire (Bartneck et al., 2009) have found medium to large effect sizes (η^2 = 0.06-0.41; Sauer, Sauer, & Mertens, 2021; Kshirsagar &

Chen, 2015; Ronckers, 2022). Likewise, studies measuring reaction time for identifying the goal of a robot found large effect sizes (η^2 = 0.1-0.2; Cuijpers, Ruijten, & Goor, 2020; Dragan et al., 2015). In Ronckers (2022) this effect size was as large as η^2 = 0.85.

Participants

Participants were partly recruited through the J.F. Schouten database of Eindhoven University of Technology, and partly through convenience sampling in the social network of the researcher. All participants were required to be above 18 years of age, and to have good enough eyesight as to not require glasses.

The study was conducted under the approval of the Ethical Review Board of Human-Technology Interaction at Eindhoven University of Technology. Following the guidelines of the ethical review board, participants received a monetary compensation of €10,00, or 12,00 if they were not a student at Eindhoven University of Technology or Fontys University of Applied Sciences, for their participation.

In total 30 participants (17 female, 13 male) took part in the experiment. Their ages ranged from 18 to 28 (M = 22.30, SD = 2.34). The first 24 participants were distributed among all possible orderings of the four conditions using a counterbalanced design. The remaining 6 participants were randomly assigned to an ordering of conditions.

Materials

Equipment

The experiment was performed in virtual reality using an HTC VIVE Pro Eye head-mounted display. The virtual environment and the simulation of the robot arm was developed using Pybullet, a platform for real-time physics simulations (Coumans & Bai, 2016). The robot arm was a virtual representation of the Franka Emika Panda Robot System.

The experimental set-up had the participant seated at a table in the middle of the lab. A QWERTY keyboard was placed on this table to record button presses of the participant. The 'd', 'f', 'j' and 'k' keys, which had to be pressed by participants during the experiment, were marked by tape with adhesive puddy beneath it to make them easy to locate while wearing the head-mounted display. The experiment leader was in the same room at a separate desk to start the virtual environment and monitor the experiment. The questionnaires were implemented in Limesurvey (Limesurvey Project Team / Carsten Schmitz, 2022) and filled out by the participants on a second computer in the same room.

Virtual environment

The virtual environment was largely adapted from Ronckers (2022). It consisted of a floor, a wooden table, a box placed on the table, eight coloured objects (referred to as 'cups'), a representation of the buttons the participant had to press, an instructional text at certain points during the experiment, and the robot arm itself. An overview of the environment from the point of view of the participant, as well as from the side, can be seen in Figure 3.

The floor consisted of grids of 1.0 by 1.0 m and stretched over the whole of the virtual environment. The table had a wooden texture and was placed on the floor 0.25 m from the participant. It had a width of 1.0 m, depth of 1.5 m, and height of 0.63 m. The box had a cardboard texture and was placed on the table on the opposite left corner from the participant. It had a width and depth of 0.25 m, and height of 0.12 m. The eight cups were placed in a semicircle with a radius of 0.25 m around the middle of the table. Each cup was a cylinder with a height of 6.6 cm and radius of 3.3 cm. There were four colours for the cups, each with a respective number: red (1), orange (2), green (3) and blue (4). Each color appeared twice, once on each half of the table.

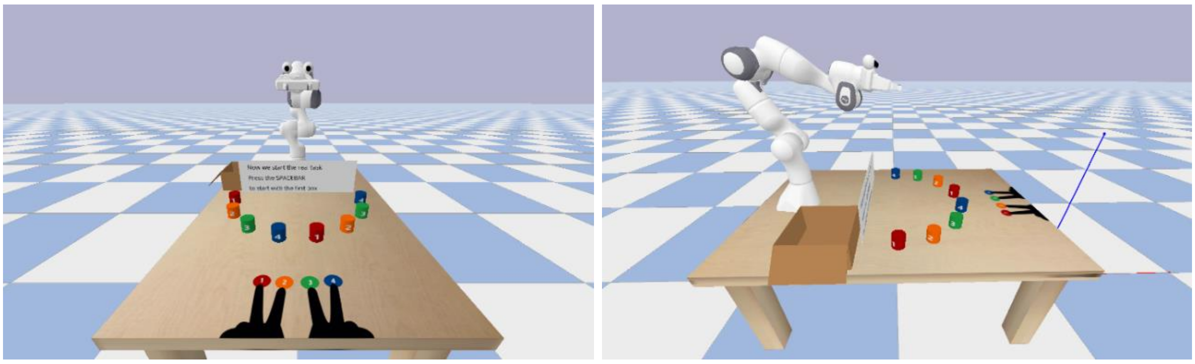

Figure 3*: The virtual environment from the point of view of the participant (left) and as seen from the side (right).*

Representations of the buttons that the participant had to press were placed on their side of the table. This appeared as the shape of two hands with their fingers over the buttons that corresponded to each color of cup. This meant that each button corresponded to two cups in the environment. However, due to the distance between the cups of the same color, it was always clear from context to which of the two cups the button-press was referring to. Pressing the button caused a checkmark to appear on the corresponding cups (see Figure 4). Only one cup colour could be marked like this at a time. Pressing another button caused the checkmark to disappear from the last one. Likewise, pressing the same button again caused the checkmark to disappear from the corresponding cups. Instructional text appeared on a white background in the centre of the table at certain points in the experiment. This text was used to instruct participants in the practice task, to ask participants whether they were ready to continue with the experiment, and to inform participants that they could proceed to the questionnaire at the end of a condition. The text and background were not visible during other points during the experiment. The contents of these texts can be found in Appendix 2.

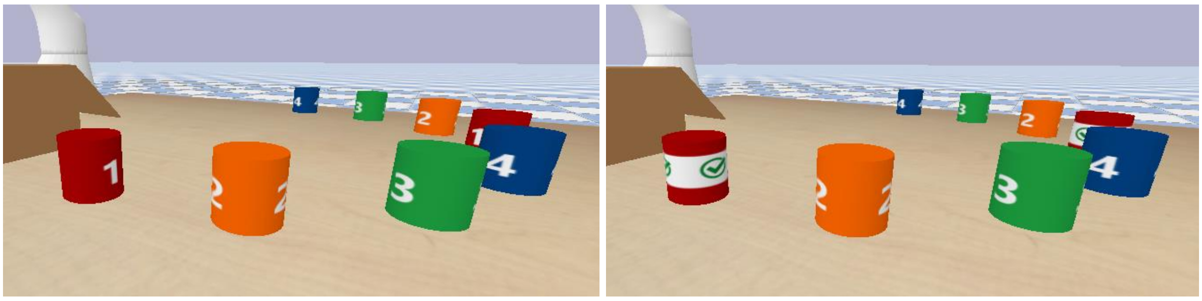

***Figure 4**: The cups before (left) and after (right) the button corresponding to the red cup is pressed.*

The robot arm was placed using its real-world size on the opposite end of the table facing the participant. Each eye consisted of a white sphere with a radius of 2.5 cm, and a smaller black sphere of 1.2 cm centred 2 cm from the white sphere to represent a pupil. The eyes were spaced 10 cm apart and floated 3 cm above the centre of the hand of the robot arm. A close-up of the 'head' of the robot arm can be seen in Figure 5. During the experiment the robot arm would pick up cups, move them to above the box, and then drop them in the box before moving back to its base position.

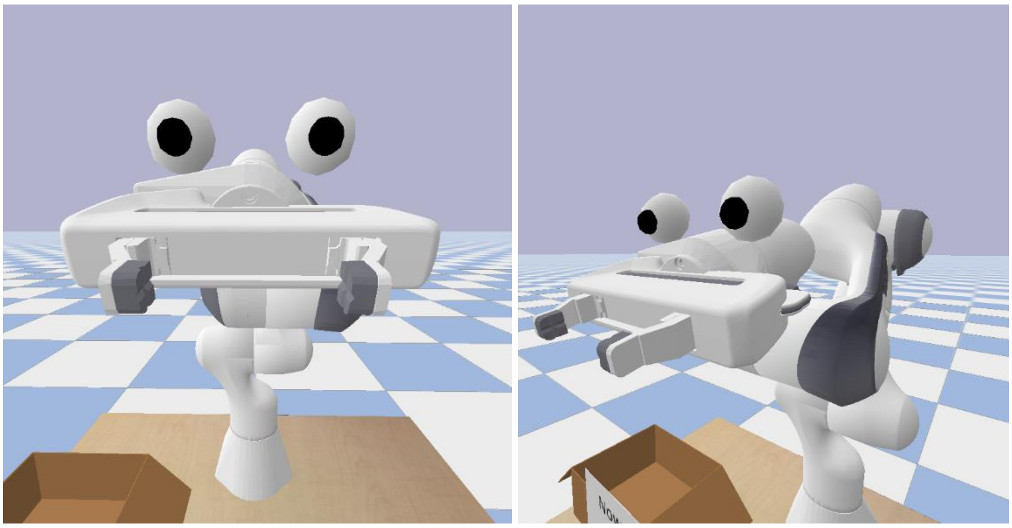

***Figure 5**: Close-up of the head, including the eyes, of the robot arm from two angles.*

Movement Trajectories

Both directed and undirected movement conditions used a minimum-jerk trajectory to reach their target. However, the directed movement differed in that it first moved to a position in which the gripper was pointing towards its target. In both conditions the full trajectory from the base position to grasping a cup took 8.5 seconds. For undirected movement this 8.5 seconds was split up into a section of 1.3 seconds where the robot arm rotated to point its gripper to its target, and a section of 7.2 seconds to move from the aligned position to grasping the target. Moving a grasped cup to the box took approximately 3 seconds and used the same minimum-jerk trajectories in both conditions. The movement of the robot arm from the box back to the base position took approximately 1.4 seconds. This movement did not use minimum jerk, instead following pre-set joint positions in 30 steps. The code for all movements was copied from Ronckers (2022).

The code for the minimum-jerk trajectory was based on the algorithm defined by Flash & Hogan (1985). For more details on how this works, see (Cuijpers et al., 2020). This algorithm is used to calculate the trajectory of the end effector, which in the case of the robot arm was the virtual point in the middle of the gripper. In the current context, this trajectory started at the middle of the gripper in its base position, and ended at the position of the targeted cup. At this point the robot arm would close the gripper, thus grasping the cup. The minimum jerk trajectory is a 5th order polynomial of time, and can be obtained by solving the vector of coefficients (a_i, b_i, c_i, d_i, e_i, f_i) based on the start (t=0), end (t=1), location (x_i), velocity (v_i), and acceleration (a_i) for each coordinate separately (i = x, y, or z) (see Equation 1). With this vector of coefficients we can calculate the desired position, velocity and acceleration at any time between t=0 and t=1.

$$\begin{pmatrix} x_i(t) \\ v_i(t) \\ a_i(t) \end{pmatrix} = \begin{pmatrix} 1 & t & t^2 & t^3 & t^4 & t^5 \\ 0 & 1 & 2t & 3t^2 & 4t^3 & 5t^4 \\ 0 & 0 & 2 & 6t & 12t^2 & 20t^3 \end{pmatrix} \begin{pmatrix} a_i \\ b_i \\ c_i \\ d_i \\ e_i \\ f_i \end{pmatrix} \tag{1}$$

Additional constraints are set for the two conditions to make the gripper reach its target in the desired manner. For undirected movement, the end effector's velocity was set to be zero at the start and end points of each motion. Acceleration was set to zero at the start, and set to 0.014 m/s^2 downward and 0.14 m/s^2 in the direction from the start to the end point of the trajectory on the horizontal plane.

For directed movement, the end effector was first aligned with a vector calculated from the base of the robot arm's gripper towards the goal. During this movement, the base of the gripper remained at the same location, while the tip of the gripper moved to its position on the computed vector. After this aligning movement, the arm proceeded with another minimum-jerk movement similar to that of undirected movement. However, the starting velocity was set to 0.055 m/s^2 downward, as the end effector was already moving downward due to the aligning movement.

The minimum jerk trajectory moving from grasping the cup to dropping it in the box, started with an upward acceleration of 1.11 m/s^2 and zero acceleration on the horizontal plane. At the end of this trajectory, when dropping the object, the acceleration was set to be zero again.

Each trajectory was divided into 100 time steps. Location, velocity and acceleration for the end effector were calculated for each time step of the trajectory. The remaining joints of the robot arm followed the trajectory of the end effector by using the inverse kinematics functionality build into Pybullet (Coumans & Bai, 2016)

Gaze Cues

In no gaze cues condition, the eyes would rotate to move along with the head of the robot arm. In the gaze cues condition, the rotation of the eyes was updated every frame to rotate towards its target. The maximum rotation speed was 4 m/s. This speed was inspected by four researchers to check if it appeared humanoid enough. In order to avoid unnatural angles, the eyes were only able to move a maximum of 55 degrees away from the front of the robot arm. This was set to be close to the 120 degrees vision span of a human eye (Hammoud, 2008).

An additional deviation was added to have the eyes face a virtual target slightly behind the actual cup it was reaching for (see Figure 6). This added distance was set to have the eyes face five degrees away from their actual target. This is based on the behaviour of human eyes when facing objects close to their eyes (Nagamatsu, Sugano, Iwamoto, Kamahara, & Tanaka, 2010) and avoids the eyes appearing like they are squinting.

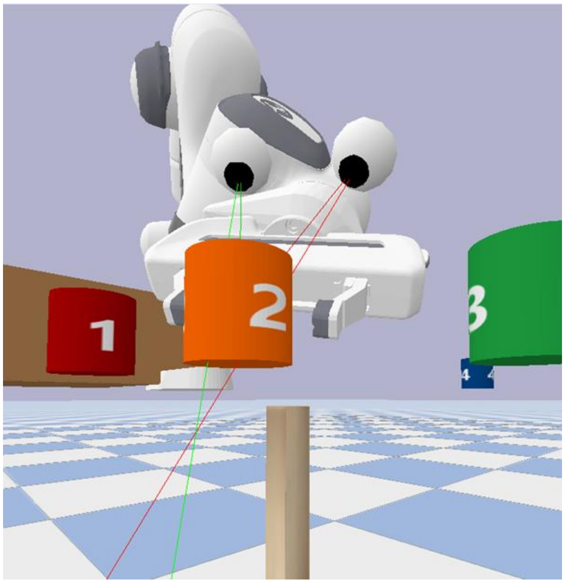

Figure 6*: The eyes of the robot are set to look at a virtual target slightly behind the cup. This avoids the eyes appearing as if they are squinting*

Measurements

Task performance

During the experimental task, participants pressed a button corresponding to which cup they expected the robot arm was reaching for. They were instructed to press this button quickly as possible. Task performance was measured as the time difference between participants pressing the correct button and the robot arm reaching the object.

User Experience

The user experience of the collaboration during each condition was measured using the Godspeed Questionnaire (GSQ) (Bartneck et al., 2009). Questions were rated on a 7-point semantic differential scale. The GSQ consisted of 23 items measuring five categories: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety. Questions were presented in a different randomized order for each participant. The full questionnaire can be found in Appendix 1.

The collected survey responses from the GSQ showed a strong internal consistency across the eight five batteries: anthropomorphism ($\alpha = 0.827$), animacy ($\alpha = 0.865$), likeability ($\alpha = 0.939$), perceived intelligence ($\alpha = 0.819$) and perceived safety ($\alpha = 0.713$).

Uncanniness

How uncanny the robot arm was perceived during each condition was measured using the uncanniness questionnaire (UQ) (Ho & MacDorman, 2017). Questions were rated on a 7-point semantic differential scale. The UQ consisted of seventeen items measuring three categories: perceived humanness, eeriness and attractiveness. The UQ should have consisted of eighteen items, but due to researcher error one item of the eeriness category (ordinary – supernatural) was left out. Questions were presented in a different randomized order for each participant. The full questionnaire can be found in Appendix 1.

The collected survey responses from the UQ showed a strong internal consistency across the three item batteries: perceived humanness ($\alpha = 0.827$), eeriness ($\alpha = 0.887$) and attractiveness ($\alpha = 0.742$).

Open questions

After undergoing all four conditions, participants were also asked to fill in three open questions: *1. How are the robot arms different from each other?; 2. Why would you prefer one robot over the others?;* and *3. Do you have any other comments on the robot or the experiment you would like to share?*. Respectively, these questions served as a manipulation check, to add contexts to the results of the user experience questionnaire, and to check whether anything relevant was missed in the experimental design.

Forced choice questions

At the end of the experiment participants answered five forced-choice questions on their preferred robot design, as was done in Dragan et al. (2015) and Ronckers (2022). These questions asked which implementation of the robot arm participants thought they were the fastest with, which robot arm they found the easiest to work with, which robot arm they liked the most, which robot arm they would prefer to collaborate with in a virtual reality setting, and which robot arm they would prefer to collaborate with in a real-world setting. The full questionnaire can be found in Appendix 1.

Demographics

Finally, participants were asked to fill in their age and the gender they most identified with. These variables mainly served to characterize our sample population.

Procedure

Participants were brought into the VR lab and asked to sign an informed consent form that explained the general procedure in accordance with the Ethical Review Board of Eindhoven University of Technology. After this they were given a more in-depth explanation of what was expected of them. Participants were told that they would enter a virtual environment in which a robot arm would pick up objects, referred to as 'cups', and move them into a box. Two printed-out pictures were used alongside this explanation (see Figure 7). One showed the point of view of the participant, including the buttons they had to press and the robot arm in its base position. The second picture showed the robot arm reaching for a cup. Participants were instructed to press the corresponding button as soon as they thought to know which cup the robot arm was reaching for. Furthermore, it was explained that they were allowed to press another button if they made a mistake, that the four designs differed in their behaviour rather than their appearance, and that there was a slim chance that the arm might get stuck due to errors of the virtual environment. In this case the experiment leader would use a button to reset the position of the arm. This ended up happening in four instances in total, across three participants. If participants had no more questions they were asked to put on the head-mounted display and the experiment leader helped with setting it up comfortably.

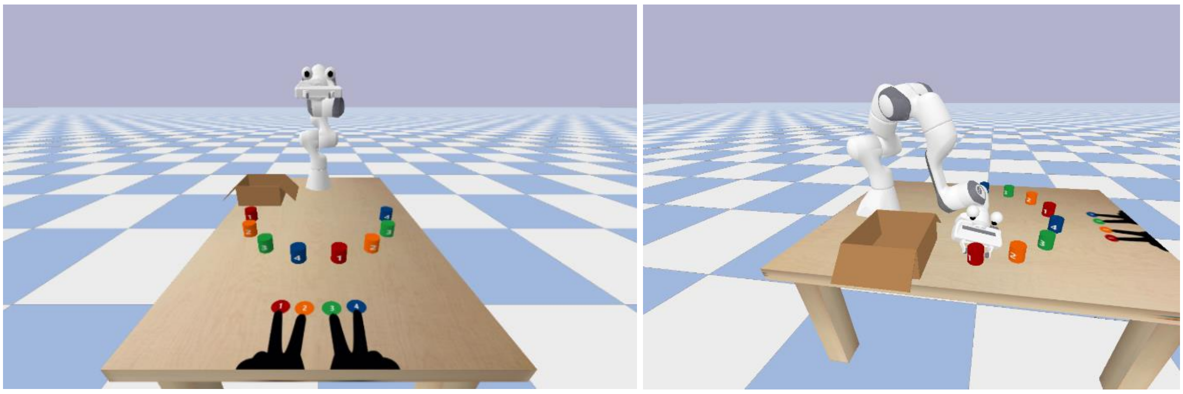

Figure 7*: The two printed out pictures shown to participants during the explanation of the procedure. The left picture shows the robot arm in its base position and the right picture shows the robot arm reaching a cup.*

Participants first underwent a general practice task in the virtual environment. This task served to get participants familiar with the virtual environment and the function of the buttons on the keyboard. There was no robot arm present in the practice task. First, a piece of text in the environment told participants to place their fingers on the buttons of the keyboard as seen in the image on the table. The precise texts used in the experiment can be found in Appendix 2. After 10 seconds, the text updated to encourage participants to press the buttons to check their effects. After an additional 10 seconds, the text told participants they could press spacebar to formally start the practice task. After pressing spacebar, the text changed to tell participants that a red line would appear to move to one of the cups, and that participants should press the button corresponding to that cup as fast as possible (see Figure 8). When the red line reached a cup, this cup disappeared. This practice task was repeated for two cups. There were no measurements made during this practice task. Afterwards, the text in the environment told participants that they could repeat this task by pressing spacebar, or that they could tell the experiment leader that they felt ready to continue with the experiment. The experiment leader started up the next part of the experiment when participants said they were ready to continue.

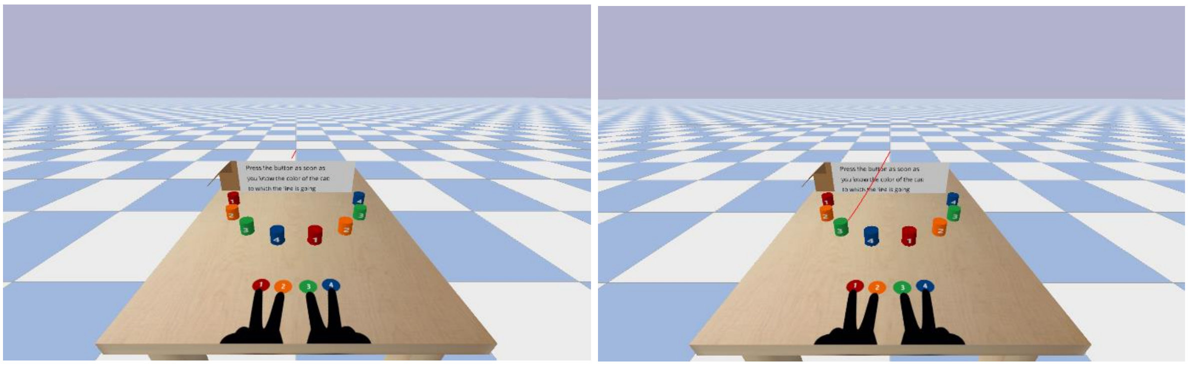

***Figure 8**: The general practice task, in which a red line slowly moves towards one of the cups.*

The next part of the experiment was the collaboration task. This was the main part of the experiment and was repeated for each of the four conditions. It started with the robot arm appearing in the environment. The text in the environment told the participant that there would first be another practice task, which would start when they pressed the spacebar (see Appendix 2 for the full texts). Note that this second practice task was different from the first. It was meant to get participants familiar with the behaviour of the robot arm in the current condition. The second practice task started when participants pressed spacebar. In this practice, the robot arm picked up two cups, selected at random, and moved them to the box. The text in the environment instructed the participant to press the correct button as soon as they knew which cup the robot arm was going to grasp. There were no measurements made during the second practice task. The second practice task could not be repeated. Instead, the text in the environment told participants that the real task would start when they next pressed spacebar.

In the real task, the robot arm filled 2 boxes each with 4 cups selected at random. This made for 8 measurements per condition. For each cup it was measured how long before grasping the cup the participant pressed the corresponding button. Before starting the second box, the text in the environment again told the participant to press the spacebar when they were ready to start. After each box, the cups would instantly appear back at their starting position. In-between moving each cup, the robot arm moved itself back to its base position. There was no text in the environment during the real task.

After the collaboration task, the text in the environment instructed participants to take off the head mounted display and fill in the questionnaire on a separate computer in the room. The questionnaire consisted of the Godspeed and uncanniness questions. After filling in the questionnaire, participants returned to the virtual environment and started the next condition.

After the fourth condition, the questionnaire also included, in order, the open questions, the forced-choice questions and the demographic questions. Afterwards participants were paid and debriefed on the goals of the experiment and what measurements were made. The full experiment lasted approximately 45 minutes.

Statistical Analysis

Task Performance

A linear mixed-effects model was created to compare the participants' task performance across the four conditions. Task performance was the dependent variable. This was measured as the time difference between the robot arm grasping its target and the participant pressing the corresponding button. The two independent variables were whether the robot arm used directed movement and whether gaze cues were used. Inter-participant variability was included as a random effect. The models were fitted using restricted maximum likelihood (REML) with unstructured covariance between random effects with the Kenward-Roger method to obtain degrees of freedom per model.

User Experience & Uncanniness

For each of the eight categories of the Godspeed and uncanniness questionnaires as the dependent variable, a two-way repeated measures ANOVA was performed to compare the effects of movement type and presence of gaze cues, while controlling for inter-participant differences. The five Godspeed categories served as one of our primary measures, whilst the three uncanniness categories were included to check whether uncanniness was a confounding variable.

Forced Choice Preference & Open Questions

A thematic analysis was performed on the open questions. For the first question (*"How are the robot arms different from each other?"*) we first divided the contents of the responses as either describing the movement of the robot arm, the gaze cues, or neither of these two. For each of these three categories we then investigated what themes emerged regarding the differences between conditions.

For the second question (*"Why would you prefer one robot over the others?"*) we first divided the contents of the responses as either attributing preference to the movement of the robot arm, the gaze cues, or neither of these two. For each of these three categories we then investigated what themes emerged in regard to the participants preferences.

The third question *("Do you have any other comments on the robot or the experiment you would like to share?")* was the most exploratory. For this question we were curious if participants noted anything that was not considered by the experimenter. We first divided the contents of the responses as either being positive, negative, or neutral to the study design or robot arm. For each of these categories we then investigated what themes emerged regarding these additional notes.

For the forced-choice preference questions we compared, for each question, the frequencies of how often each condition was preferred.

Exploratory Analyses

Two additional linear mixed-effects models were created to check additional confounding effects on task performances. The first model checked for differences between the different cups. The second model checked for differences between the orderings in which the four conditions were presented to the participants. The models were fitted using restricted maximum likelihood (REML) with unstructured covariance between random effects with the Kenward-Roger method to obtain degrees of freedom per model. Both models had task performance as the dependent variable, directed movement and gaze cues as two independent variables, and inter-participant variability was included as a random effect. For the first model the cups were included as independent dummy variables as well. For the second model the order in which each condition was presented to the participant was included as an independent variable as well.

The effects of ordering were also checked in the user experience and uncanniness variables. To do this, a two-way repeated measures ANOVA was performed for each of the eight Godspeed and uncanniness categories as the dependent variable. The independent variables were whether directed movement was used, whether gaze cues were used, and the order in which each condition was presented to the participant. The repeated measure controlled inter-participant differences.

Results

Pre-processing and Assumption Checking

Task Performance

The task performance dataset was first checked for outliers. Four observations were removed because they were empty. This meant participants forgot to press a button in those instances. Additionally, four observations were removed because the robot arm got stuck and had to be reset by the experiment leader, making the measurement inaccurate. Furthermore, there were eight observations where the participant pressed the correct button only after the cup was already grasped by the robot arm. These instances likely represent mistakes from the participants. Still, because these mistakes only made up eight observations out of the remaining 952, their impact on the dataset was considered legible. They were thus retained in the dataset. There were no other abnormalities in the dataset.

Normality was checked and rejected for all conditions using both Shapiro-Wilk W tests for normality and joint skewness and kurtosis tests for normality. The test results for respectively the conditions no gaze cues & undirected movement, gaze cues & undirected movement, no gaze cues

& undirected movement, and gaze cues & directed movement for the Shapiro-Wilk W test were: (W, = .908, $p < 0.000$), (W, = .919, $p < 0.000$), (W, = .968, $p < 0.000$) and (W, = .972, $p < 0.000$). For the joint skewness and kurtosis test: ($X^2(2, N = 240) = 53.77$, $p < 0.000$), ($X^2(2, N = 238) = 183.68$, $p <$ 0.000), ($X^2(2, N = 237) = 37.02$, $p < 0.000$) and ($X^2(2, N = 237) = 10.44$, $p < 0.005$).

Inspecting the histograms of the anticipation time across the four conditions (see Figure 9) shows that, for the first condition (no gaze cues & undirected movement), there are some data points that skew the data slightly to the left. The histogram for the second condition (gaze cues & undirected movement) shows a bimodal distribution with peaks around t = -7 and t = -2. This effect is seen to a lesser extent in the histograms for the remaining two conditions as well. There were no transformations that resulted in normality for all conditions. However, as linear mixed models have been shown to be robust against violations of normality (Schielzeth et al., 2020), the analysis proceeded with the untransformed dataset.

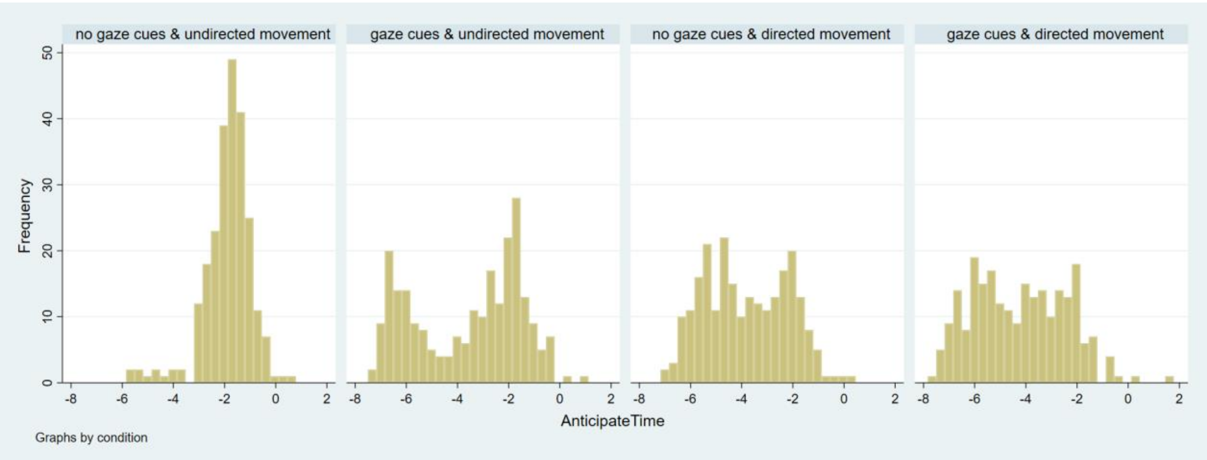

Figure 9*: Histograms for anticipation time (time in seconds between participants pressing the correct button and the robot arm grasping the cup) in each experimental condition.*

User Experience and Uncanniness

For each of the eight variables retrieved from the GSQ and UQ (anthropomorphism, animacy, likeability, perceived intelligence, perceived safety, perceived humanness, eeriness and attractiveness), normality was checked using both Shapiro-Wilk W tests for normality and joint skewness and kurtosis tests for normality. Additionally, heteroscedasticity was checked using Levene’s test for equality of variance. Across these tests, a single rejection was found. Namely for the joint skewness and kurtosis test for perceived safety in the no gaze cues & directed movement condition ($X^2(2, N = 30) = 6.25$, $p = .044$). The histogram of this variable (see Figure 10) shows that it follows a somewhat uniform distribution between the values of 3 and 6.5. The analysis proceeded with this variable, as previous research has shown that a repeated measures ANOVA is generally robust to violations of normality (Blanca, Arnau, García-Castro, Alarcón, & Bono, 2023).

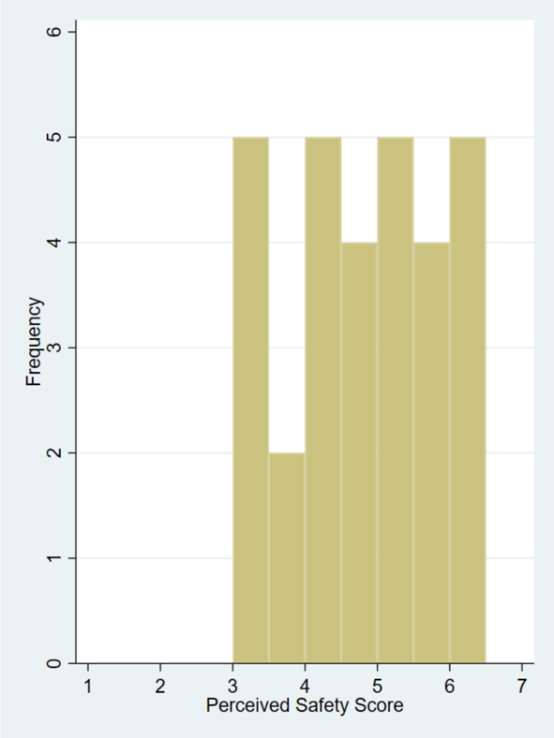

***Figure 10**: Histogram of the Perceived Safety variable in the no gaze cues & directed movement condition. Normality was rejected for this variable. It shows a somewhat uniform distribution between the values of 3 and 6.5.*

Effects of Directed Movement and Gaze Cues on Task Performance

The results of the mixed model revealed a statistically significant difference between the four conditions on anticipation time ($F(3, 919.07) = 113.14$, $p < 0.000$). The variables of directed movement and gaze cues were both statistically significant ($t = -14.13$, $p < 0.000$) and ($t = -12.41$, $p < 0.000$), as was their interaction ($t = 6.70$, $p < 0.000$). Comparing the mean anticipation times for the four conditions (see Figure 11) revealed that participants were fastest when gaze cues and directed movement were combined ($M = -4.242$, $SD = 1.819$). The combination of directed movement without gaze cues provided the second-fastest anticipation times ($M = -3.839$, $SD = 1.627$), followed by undirected movement with gaze cues ($M = -3.594$, $SD = 2.106$). The lowest anticipation times were found when neither directed movement nor gaze cues were used ($M = -1.877$, $SD = .929$). Note that anticipation time is coded as negative. It represents the time in seconds the participant pressed the corresponding button *before* the robot arm grasped the cup.

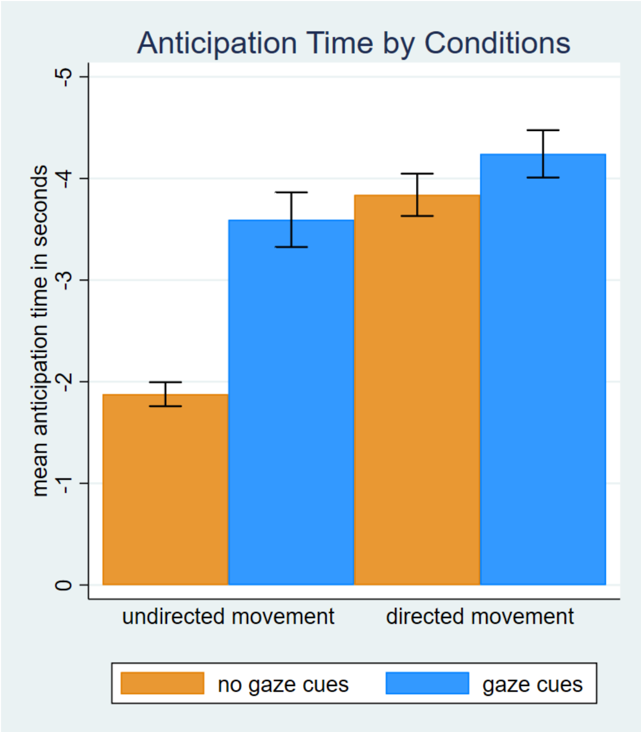

Figure 11*: Mean anticipation time per condition with error bars indicating the 95% confidence interval. Anticipation time is defined as the time difference between the robot arm grasping the cup, and the participant pressing the corresponding button.*

Effects of Directed Movement and Gaze Cues on User Experience and Uncanniness

A two-way repeated measures ANOVA was performed on each of the eight variables that made up the GSQ and UQ. Post-hoc pairwise mean comparisons (Tukey corrected with equal variances) were made for the significant effects. Mean values for all variables across all conditions can be found in Figure 12.

For anthropomorphism there was a significant effect for gaze cues ($F(1,120) = 13.40$, $p < .001$) with a pairwise mean contrast of .657 (*SD* .212). For animacy there was a significant effect for gaze cues ($F(1,120) = 36.91$, $p < .000$) with a pairwise mean contrast of .939 (*SD* .191). There also was a significant interaction effect between directed movement and gaze cues ($F(1,120) = 12.93$, $p < 0.001$). This interaction effect is reflected in the mean scores, where the presence of gaze cues scores similarly between directed and undirected movement, but there is a large pairwise mean contrast (1.34 (SD .264)) between directed and undirected movement when no gaze cues are used (see Figure 12). For likeability there were significant effects for directed movement ($F(1,120) = 5.68$, $p < 0.024$) with a pairwise mean contrast of .543 (*SD* .218), gaze cues ($F(1, 120) = 27.38$, $p < 0.000$) with a pairwise mean contrast of .810 (*SD* .210), as well as an interaction effect ($F(1, 120) = 5.35$, $p < 0.028$). This interaction effect is reflected in the mean scores, where the presence of gaze cues scores similarly between directed and undirected movement, but there is a large pairwise mean contrast (1.14 (*SD* .289)) between directed and undirected movement when no gaze cues are used (see Figure 12). There were no significant effects for either perceived intelligence or perceived safety.

For perceived humanness there was a significant effect for gaze cues (*F*(1, 120) = 13.40, *p* < 0.001) with a pairwise mean contrast of .657 (*SD* .212). For eeriness there were significant effects for directed movement (*F*(1, 120) = 8.81, *p* < 0.006) with a pairwise mean contrast of 0.485 (*SD* .192), gaze cues (*F*(1, 120) = 58.63, *p* < 0.000) with a pairwise mean contrast of .894 (*SD* .179), as well as an interaction effect (*F*(1, 120) = 14.58, *p* < 0.001). This interaction effect is reflected in the mean scores, where the presence of gaze cues scores similarly between directed and undirected movement, but there is a large pairwise mean contrast (1.317 (*SD* .241)) between directed and undirected movement when no gaze cues are used (see Figure 12). For attractiveness there was a significant effect for gaze cues (*F*(1, 120) = 9.62, *p* < 0.004) with a pairwise mean contrast of .x (*SD* .186).

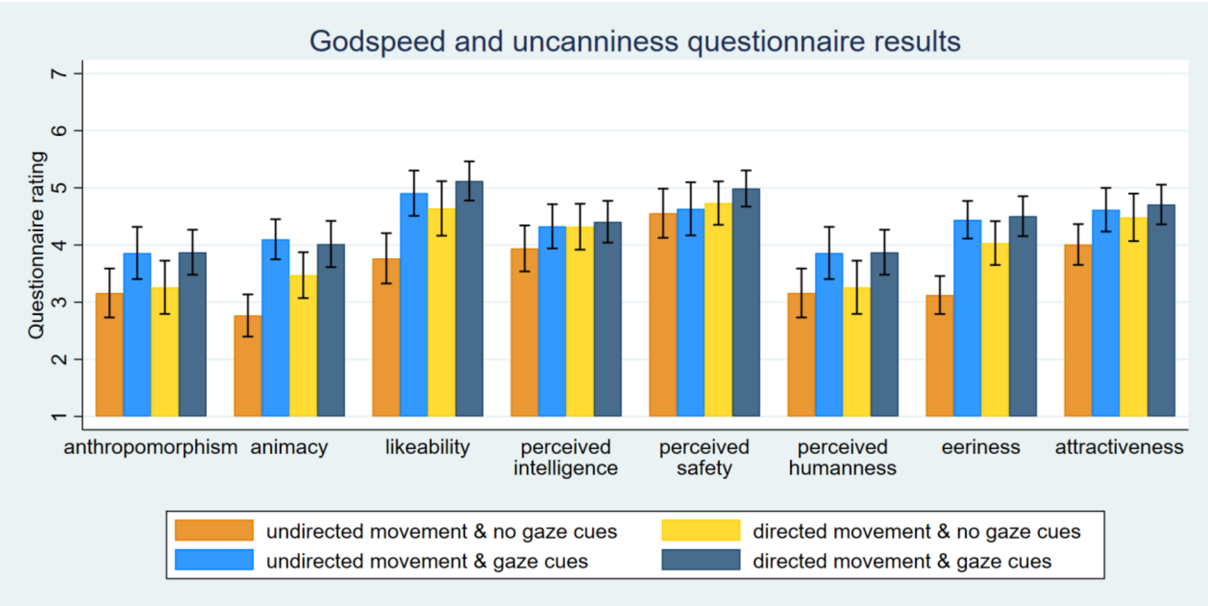

Figure 12*: Mean scores with error bars indicating the 95% confidence interval for each of the Godspeed and uncanniness questionnaire categories.*

Forced Choice Preference Questions

The frequencies of the answers for the forced choice questions are displayed in Figure 13. Participants judged themselves to be fastest with the fourth (directed movement & gaze cues)) condition, followed by the second design (undirected movement & gaze cues). Similarly, they found the fourth condition to be easiest to work with, although the second and third (directed movement & no gaze cues) design score quite close to it. The second condition seems to be the most liked, and most preferred to work with in both a virtual and real-world setting. It is followed by the fourth condition in all these three cases. Generally, there seems to be a preference for design two and four, both of which utilized gaze cues. An exception is the second question, which is the only question in which a condition not using gaze cues made it to the second place.

On average across all questions, the condition with undirected movement & no gaze cues was preferred by 10% of participants, undirected movement & gaze cues by 34.67%, directed movement & gaze cues by 22%, and directed movement & gaze cues by 33.33%.

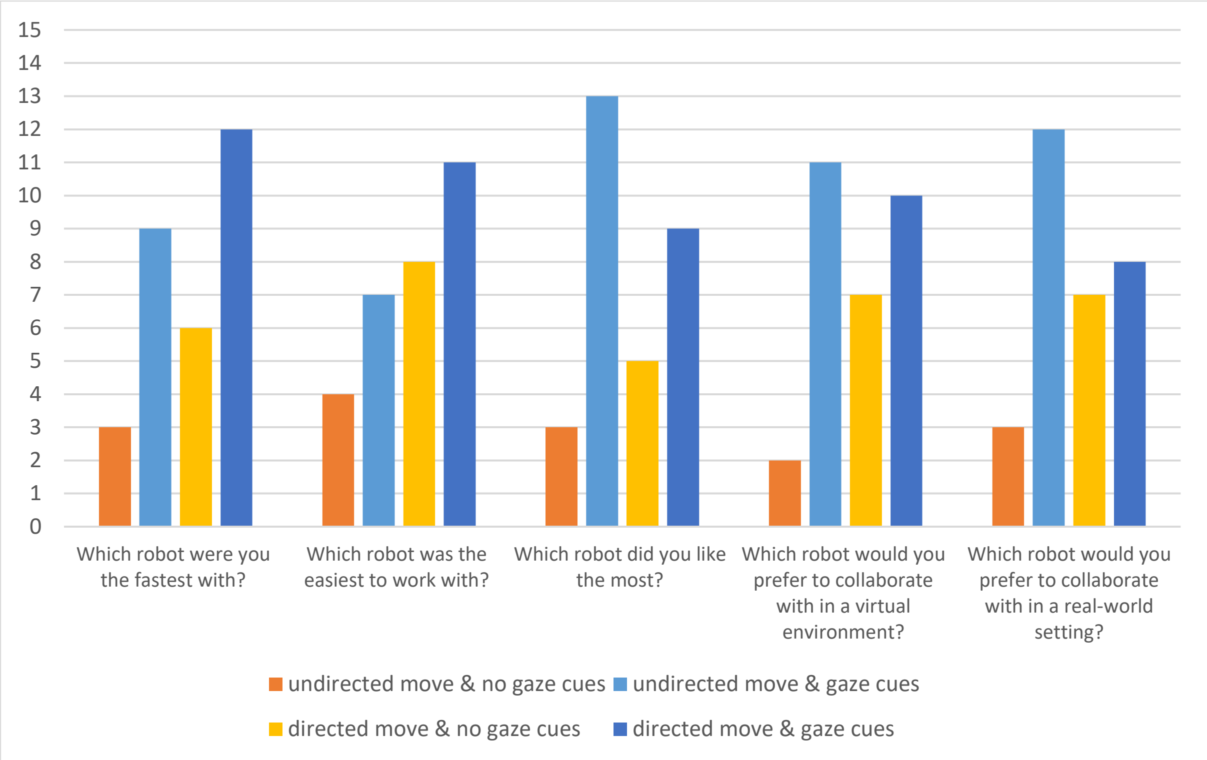

***Figure 13**: Preference frequencies for each condition over each question*

Open Questions

Question 1: How are the robot arms different from each other?

29 out of the 30 participants explicitly mentioned factors related to the movement of the robot arm in their answers. The one participant that did not mention movement instead wrote about how the designs differed in their "attitude" towards the participant. Seven participants mentioned how the different robot arms "moved different parts of their bodies". Additionally, five participants described the robot arm to be moving its "head" first, or "establishing eye contact" before moving. An additional six participants described the movement difference as moving directly towards the object or not. Eleven participants described the movement as differing in terms of "smoothness", "rigidness", "fluidity" or "flexibility". This movement was sometimes referred to as being more or less "humanlike" or "machine-like".

21 out of 30 participants explicitly mentioned differences in the eyes between conditions. However, with some words (e.g., "gaze") it can be hard to interpret whether participants referred to the movement or eyes of the robot arm. Eleven participants described that the eyes "looked" at

their target while moving. Two participants wrote that “the intentions were communicated more clearly” when the eyes targeted these objects. An additional two participants wrote that the “eyes made the whole seem less mechanical”.

Seven participants explicitly referred to disliking the undirected movement, in 4 cases paired in the condition without gaze cues. Participants described that this robot arm was very hard to predict, as it only became clearer later on in the movement which cup it was reaching for. Two participants describe it as the robot arm “changing plans at the end”. Four participants refer to the robot arm as “a trickster” or “deceptive on purpose”.

Question 2: Why would you prefer one robot over the others?

Five participants mentioned an explicit preference for directed movement over undirected movement. These participants describe this as stemming from a preference for the “smoothness” or “less machine-likeness” of the movement.

In contrast, seven participants noted an explicit preference for undirected movement. Again, this movement is described as more “smooth”, “natural” or “fluid”. Further reasoning differs between these participants. Three mention a dislike for the directed movement because it “moved weirdly”, or was “distracting”. One participant mentions that their preference would be different “was it not for the outer cups”, to which the directed movement “moved weirdly”. Of the seven participants that preferred undirected movement, five explicitly mentioned they preferred it in combination with the gaze cues.

Sixteen participants noted gaze cues as being important for their choice of preference. Ten of these explicitly mention that these eyes “made the intentions of the robot clear earlier”.

Only three participants were explicitly against the use of eyes. One describes them as “very cute”, but that the condition without eyes was better for prediction. Another writes that “eyes added to the uncanniness when it moved too much or not at all”. The third one states they would have preferred the eyes to not have been present at all.

Eight participants did not explicitly refer to the use of directed movement or gaze cues in their response. Five of them mentioned that “predictability” was the key factor that caused their preference, but did not explain what feature caused this predictability. Three participants noted that “human-likeness” was the key factor that caused their preference, but again did not explain what caused this human-likeness.

We also see that some participants attribute humanlike qualities to the robot. One participant described how the undirected no gaze robot arm “was mean” and “did not know what it was doing”, and that the other designs were “nicer”. Another participant describes the robot arms using gaze cues as “more goofy, like a puppy playing a game”.

Question 3: Do you have any other comments on the robot or the experiment you would like to share?

Five participants described the eyes to be “shaking” or “vibrating” during some moves. One described this as “distracting”, one as “less alive and intelligent”, and one as “derpy, which makes it fun”. An additional three participants noted a dislike for the eyes “staring” straight ahead in the rest position. One described it as “staring at me without thought behind it”, and another described it as “scary”.

Four participants noted that the directed gaze seemed to have “trouble with objects on the right”, to which the robot arm seemed to need a longer time to reach.

Six participants made note of problems with the experimental setup. Two write that they disliked sitting down and would have liked to move around in the virtual environment. One participant disliked the keyboard position, as it cramped up their wrists. Another participant complained that the environment was not very engaging and suggests that this might have made the robot appear more lifeless. Another participant suggested adding colour to the eyes of the robot arm to make it stand out more. And finally, a participant suggested adding noise to the movement of the robot arm, to make it more realistic and engaging.

Exploratory analyses

Effects of Cup Positions on Task Performance

A second mixed model on task performance also included the differences between cups. The model revealed a significant effect on anticipation time ($F(10, 912.06) = 72.60$, $p < 0.000$). The results for directed and gaze cues remained statistically significant ($t = -16.03$, $p < 0.000$) and ($t = -14.22$, $p < 0.000$), as did their interaction ($t = 7.71$, $p < 0.000$).

The cups were coded from 0 to 7, ordered from left to right from the point of view of the participant. The test showed that cups 1 ($t = -3.63$, $p < 0.000$), 2 ($t = -3.96$, $p < 0.000$), 3 ($t = -9.21$, $p < 0.000$), 4 ($t = -8.29$, $p < 0.000$) and 7 ($t = 3.97$, $p < 0.000$) were significantly different from cup 0. Inspecting the mean anticipation times (see Figure 14) revealed a trend where the cups in the middle are easier to anticipate (quicker anticipation times) than those on the edges (slower anticipation times).

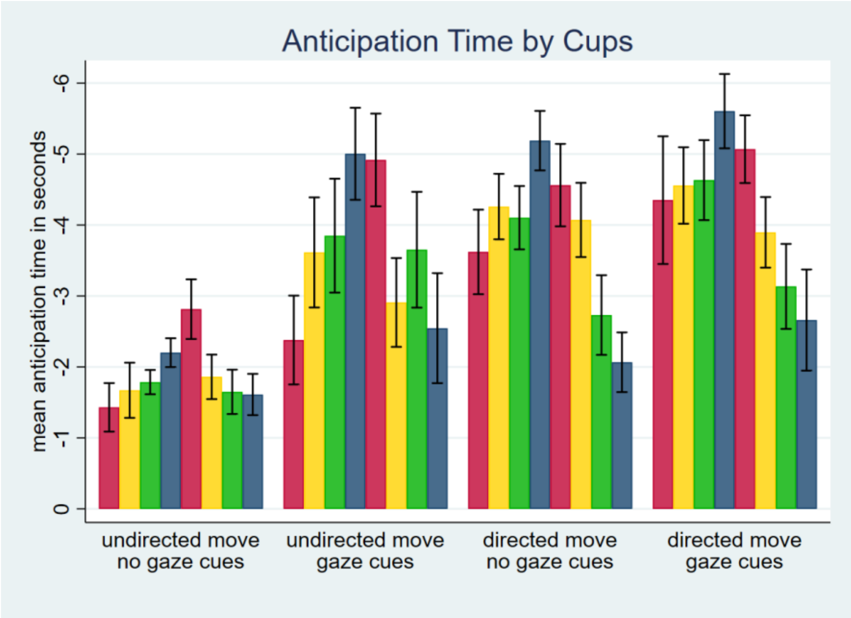

Figure 14: *Mean anticipation time with error bars indicating the 95% confidence interval per condition and cup. Anticipation time is defined as the time difference between the robot arm grasping the cup, and the participant pressing the corresponding button.*

Effects of Condition Ordering on Task Performance

A third mixed model for task performance also included the order in which each condition was introduced to the participant. The model revealed a significant effect on anticipation time ($F(4, 918.09) = 87.19$, $p < 0.000$). The results for directed and gaze cues remained statistically significant ($t = -14.28$, $p < 0.000$) and ($t = -12.44$, $p < 0.000$), as did their interaction ($t = 6.87$, $p < 0.000$). The ordering variable itself had a significant effect on anticipation time ($t = -2.66$, $p < 0.008$), although the coefficient (-0.117; $SD = .044$) was very small compared to the coefficients of directed movement and gaze cues.

Inspecting the mean anticipation times (see Figure 15) shows that the orderings affect each conditions differently. Additional one-way ANOVAs for each of the conditions of ordering on anticipation time reveal that there is no significant difference between orderings in the undirected movement & no gaze cues condition, and directed movement & gaze cues condition. However, there was a significant difference in the undirected movement & gaze cues condition ($F(3,238) = 7.74$, $p < 0.000$) and directed movement & no gaze cues condition ($F(3,237) = 5.69$, $p < 0.001$). For directed movement & no gaze cues, there seems to be no clear pattern between the orderings (see Figure 15). However (undirected movement & gaze cues) shows an upward trend (see Figure 15). A regression analysis showed the coefficient of this slope to be -.562 seconds.

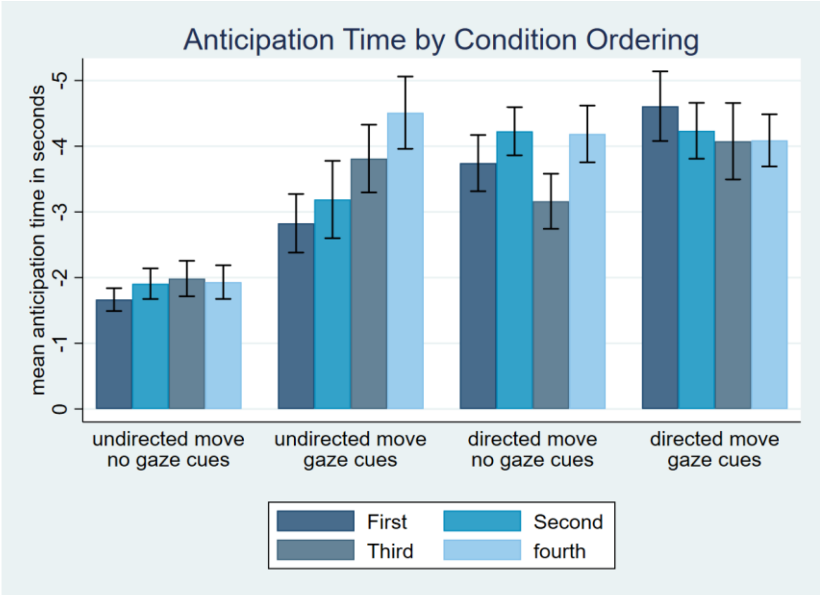

***Figure 15**: Mean anticipation time with error bars indicating the 95% confidence interval per condition and the order this condition was introduced. Anticipation time is defined as the time difference between the robot arm grasping the cup, and the participant pressing the corresponding button.*

Effects of Ordering on User Experience & Uncanniness

In order to check for the possibly confounding effect of the order in which conditions were presented on the GSQ and UQ items, an additional factorial ANOVA was performed for each item with the independent variable of condition ordering, alongside directed movement and gaze cues. A significant effect for ordering was found for two of these, namely likeability ($F(3,120) = 4.11$, $p < 0.008$) and attractiveness ($F(3,120) = 4.38$, $p < 0.006$). The mean distributions can be seen in Figure 16. These both show a similar pattern, with the scores lowering for the first three conditions, then rising again.

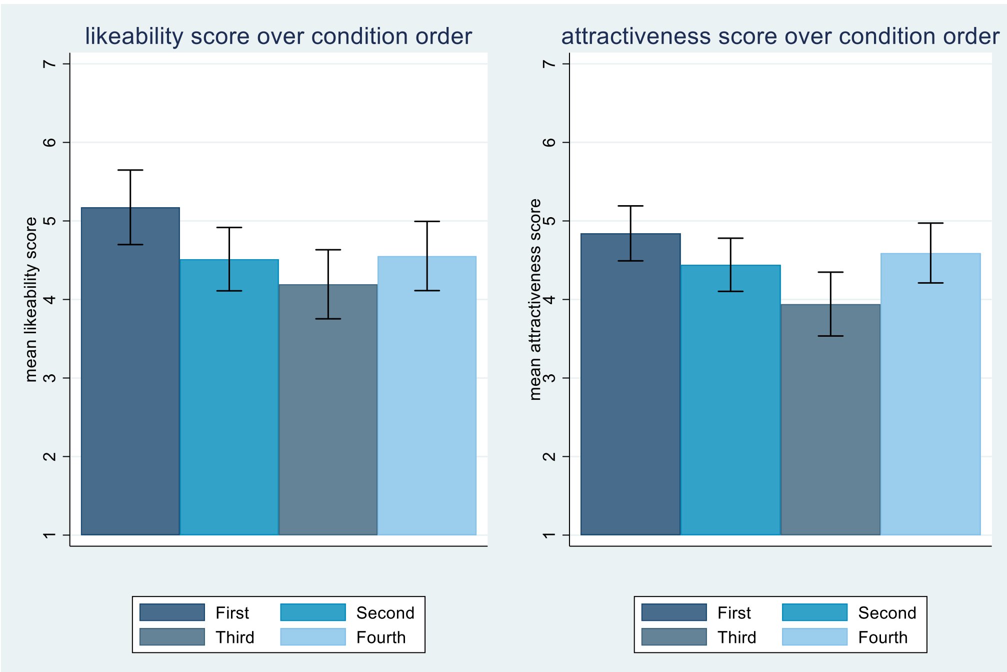

***Figure 16**: Mean likeability (left) and attractiveness (right) scores for the order in which conditions were presented. Error bars indicate the 95% confidence interval.*

Differences in movement of the robot arm between conditions

As was described in the Method, the movement trajectory of the robot arm in both the directed and undirected condition was set to take 8.5 seconds. However, inspecting the time it took each arm to reach each cup in reality (see Figure 17) revealed that the movement time for directed movement condition was closer to 8.3 seconds, and 8.4 seconds for undirected movement. Furthermore, there seemed to have been a deviation for the two cups on the most right, respectively taking reliably closer to 8.55 seconds and 8.75 seconds. It is unknown what caused this discrepancy between the current experiment and Ronckers (2022).

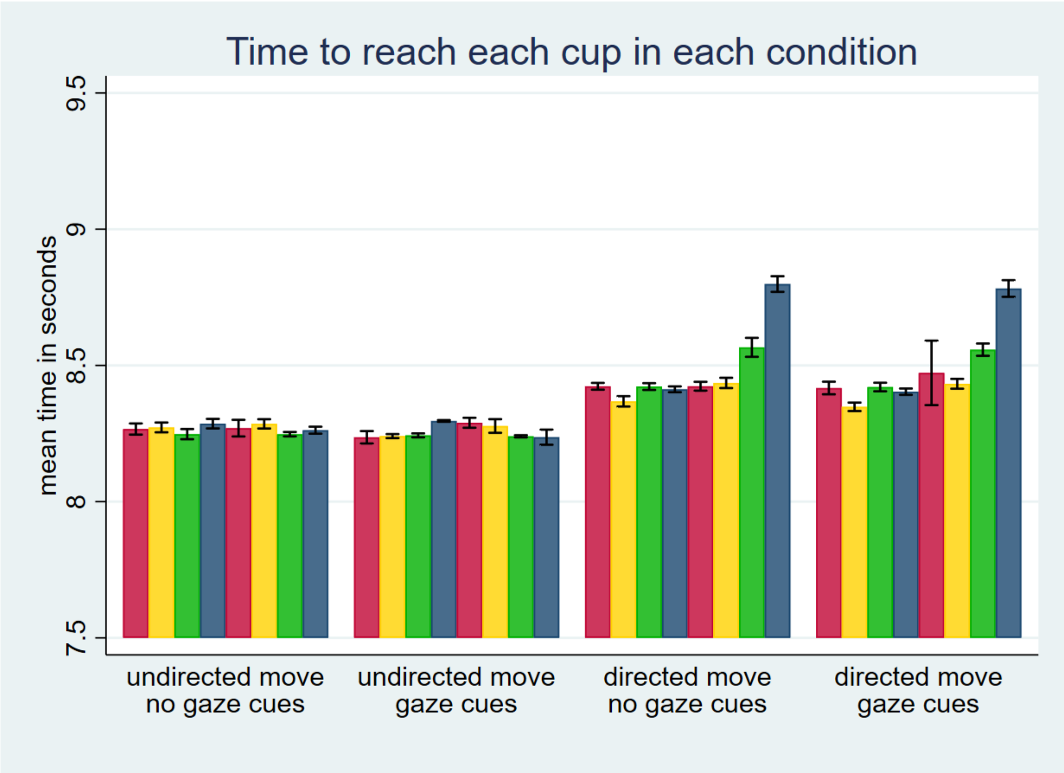

***Figure 17**: Bar graph showing the mean time it took each arm to reach each individual cup. Error bars indicate the 95% confidence interval. The gaze movement took slightly longer, and that there was a deviation for the two rightmost cups in the gaze movement condition.*

Discussion

Main Findings

Task Performance

We found that both gaze cues and directed movement made participants significantly faster in identifying the target of the robot arm. The effect of directed movement was larger than the effect of gaze cues. However, the largest effect was found when both manipulations were combined.

Our results are in line with the expected outcomes. Although we were unsure which condition would result in better task performance between gaze cues or directed movement. Whilst it turned out that directed movement increased the anticipation time the most, the outcomes between the two conditions only differ by 0.245 seconds. Positive effects for gaze cues on task performance have previously been found for movement anticipation tasks (Khoramshahi et al., 2016) and target identification tasks (Boucher et al., 2012). However, the current study is the first to investigate these effects in an industrial robot arm. Both these earlier studies used humanoid robots, and very different experimental tasks. Unfortunately, the differences make the effect sizes of these studies very hard to compare to ours.

We can use Ronckers (2022) as a standard to compare our results against, as most of our experimental set-up was adapted from that study. In Ronckers (2022) the robot arm using directed movement and sunglasses performed better in terms of response time compared to the robot arm using solely directed movement. In our study, the condition that best matches the manipulation of Ronckers (2022) is the directed movement and no gaze cues condition, as Ronckers (2022) did not use gaze cues. It is then strange that our results showed a much smaller effect size in this condition. Instead, the effect size that is closest to Ronckers (2022) was found for the directed movement and gaze cues. A difference between the studies is that our experiment had humanlike eyes attached to the robot arm in all conditions, whilst Ronckers (2022) only attached sunglasses in the directed movement condition. It is possible that the presence of the eyes by themselves, even without gaze cues, made the intentions of the robot arm more clear. However, this cannot be said with certainty in the current study.

The large increase in anticipation time when either gaze cues or directed movement is used, highlights that people benefit from social cues that communicate intention in a robot arm. The presence of either cue is already of benefit to the user, and the effect sizes of the two are similar. This shows that this communication of intention can be done in multiple forms, although the best effects are seen when communication channels are combined. This shows a form of redundancy gain. Other studies have also argued for multi-modal communication in social robots (Tatarian et al., 2022).

When it comes to practical significance, either gaze cues or directed movement would suffice in shortening collaboration times. Although the best outcome would utilize both, considering the implementation costs of adding eyes with accurate functionality, changing the trajectories to use directed movement will probably be a more cost-effective measure to increase task performance.

User Experience

Based on earlier findings in social HRI, we hypothesised that both directed movement and gaze cues would lead to higher user experience ratings in an industrial robot as well. This turned out to be only partially correct. Gaze cues increased ratings for anthropomorphism, animacy and likeability. Directed movement only increased the likeability of the robot arm, whilst combining directed movement with gaze cues gave a larger increase in likeability. Nevertheless, we found no effects of directed movement nor gaze cues on perceived intelligence and perceived safety.

A previous implementation of eyes in an industrial robot arm found that this negatively affected on the user experience, specifically trust (Onnasch & Hildebrandt, 2022). The study suggested that anthropomorphic features in industrial robots should serve a functional purpose, which was not the case in their study. It seems that this was achieved for our study though. The functional goal of the eyes was to provide gaze cues, and these gaze cues were shown to improve the user experience. This shows that functional implementations of anthropomorphic features are indeed possible, and can yield benefits, for industrial robots.

The questions making up anthropomorphism and animacy are quite similar to each other, as both ask about the humanness and liveliness of the robot arm. Anthropomorphism and animacy both score significantly higher when gaze cues were used. This could imply that people recognize a sense of humanness from the eye movement. This same effect was not achieved by the use of directed movement. This is odd, as directed movement did affect these categories in Terzioglu et al. (2020). It could be that the different nature of the tasks between studies caused this effect. Alternatively, it could be that the presence of the eyes, whether they used gaze cues or not, influenced the participants more than the movement of the arm did.

Both gaze cues and directed movement increased the likeability of the robot, although the effect for gaze cues was stronger. Still, the highest likeability score was found by combining the manipulations. In the literature it is commonly found that social robots are perceived as more likeable when they are more humanlike (Fink, 2012; Hostettler et al., 2022), which could explain the increase in likeability found here. However, it is weird that this is not reflected in the anthropomorphism and animacy score, where only a significant effect for the gaze cues was found. Then again, the effect of the gaze cues on likeability was larger than that of directed movement. The increase in likeability from directed movement could also have been due to a form of zoomorphism (Bergman et al., 2019; Sauer et al., 2021) rather than anthropomorphism. However, in that case an increase in animacy would have been expected for directed movement as well.

There were no differences between conditions for perceived intelligence and perceived safety. Their ratings were not particularly high, nor particularly low. The safety variable might be less trustworthy in an experimental setting though, especially considering the collaboration took place in virtual reality, where participants might have experienced less risks. It is interesting that no increase in perceived intelligence was found though, as factors of anthropomorphism, likeability and intelligence are often related to each other, as was also found in Terzioglu et al. (2020). A meta-analysis of the Godspeed questionnaire (Weiss & Bartneck, 2015) suggests that perceived intelligence is affected more by the interaction scenario than the actual behaviour of the robot. It could be that the role of the robot arm in the packing task was deemed to not require intelligence.

Ronckers (2022) serves as a good comparison for our results In that study, the score for all five Godspeed categories was significantly higher for the robot arm using directed gaze and sunglasses. Our findings for anthropomorphism, animacy and likeability were similar to Ronckers (2022). Both in terms of their values and their contrast. This is odd though, as Ronckers (2022) did not use gaze cues, which was the factor that mainly influenced these variables in our study. The results from our study suggest that directed movement by itself should not have the effects seen in Ronckers (2022). The only difference between the studies is the lack of sunglasses in the undirected movement condition of Ronckers (2022). This would suggest that the mere presence of the sunglasses, or eyes, had a large confounding effect on the user experience.

It is also odd that Ronckers (2022) found an increase in perceived intelligence, which was not found for us. Especially considering how the interaction scenario, which was the same between the studies, has been found to affect this variable the most (Weiss & Bartneck, 2015). Instead, the mean perceived intelligence score found in our study is closest to the perceived intelligence of the undirected movement condition in Ronckers (2022). This implies that the presence sunglasses made the robot appear more intelligent compared to the presence of eyes, whether these eyes provided gaze cues or not. The perceived safety score was also a lot higher for both conditions in Ronckers (2022) compared to the values found in our study. Again, this could imply that the presence of the eyes made the robot appear less safe.

Uncanniness as a Confound

Next to the user experience variables, we also measured how uncanny the robot arm was perceived. The questions for perceived humanness and attractiveness are very similar to anthropomorphism and animacy. This explains why the significant effects and effect sizes of these categories are the same as those found for anthropomorphism.

Eeriness was the main category of interest. This measured feelings of unease and unnaturalness towards the robot. There was a significant increase in eeriness for both directed movement and gaze cues, and an even stronger effect when both were combined. This tells us that participants felt more uneasy towards the robot arm if it used gaze cues, directed movement, or combined the two. It makes sense that the robot was perceived as more uncanny when it expressed less machinelike features (i.e., gaze cues and directed movement). The literature warns that this uncanniness can negatively impact the user experience (Mori, 1970; Ho & MacDorman, 2017). However, our results show that scores for likeability and eeriness go paired with each other. Both have similar effect sizes and the same significant effects. Although the likeability score is consistently higher than the eeriness score. This could mean that both the gaze and directed movement caused an increase in the affective response to the robot. This response was not negative though, as seen by the increase in likeability. This suggests that, whilst directed movement and gaze cues caused uncanniness, this uncanniness was not strong enough to negatively impact the user experience.

Forced Choice Preference Questions

The forced choice preference questions generally showed a preference for a robot arm using gaze cues. However, between directed and undirected movement, there was a discrepancy between which robot arm participants found the most efficient, and which robot arm they would prefer to work with in practice. Specifically, directed movement combined with gaze cues, whilst preferred for efficiency and making the task easier, was not preferred when asked about likeability and which was preferred to work with in practice. This implies that people find likeability more important than efficiency when it comes to choosing a robot interaction partner.

One outlier in the results is found in the question on which robot participants found easiest to work with. This was the only question in which the robot arm using directed movement and no gaze cues receives second place. For all other questions, the top two answers were always robots using gaze cues. The number one choice in this question was the robot using directed movement combined with gaze cues. This implies that directed movement is seen as easiest to work with, although again we see that the ease of work is not necessarily reflected in likeability and choice in practice.

It should be noted that the answers say nothing about how much each preferred choice was favoured over the others. The fact that participants are divided in which condition they preferred, might also be a problem for practical implementations. If we take the preferred option in the question *"which robot would you prefer to work with in a real-world setting"* as indicative of what would be implemented in a practical setting, a robot arm using undirected movement and gaze cues would be chosen. This robot arm was preferred by 36.67% of participants. However, the remaining 63.33% would then have to work with a robot that does not have their preference.

Open Questions

The first open question mainly served as a manipulation check. It was used to find out whether participants noticed the differences between conditions and how they were interpreted. Almost all participants noted differences in the movement of the arm. Generally the answers showed that participants noticed the difference in movement as intended. However, only 21 out of 30 participants talked about gaze cues in their answer. This could mean that these participants did not notice the gaze cues manipulation, or that they were more intrigued by the movement differences and therefore wrote about these instead. Considering that gaze cues had more effect on the measured user experience than directed movement, it is interesting that gaze cues got written about less. This could imply that the increased liking from gaze cues happens more subconsciously.

Then again, in the preference question we did see a clear explicit preference for the two designs using gaze cues. Perhaps the effects of gaze cues would be even bigger if people were explicitly made aware of them.

The second question asked participants to explain why they preferred certain robot arms over the others. Although movement was described more in the first question, in the second question more participants attributed their preference as stemming from the gaze cues. These answers mainly describe how gaze cues helped in communicating intention towards the cups. However, there was also a small group of participants that was explicitly against the use of gaze cues. One described them as cute, but that this was distracting, while another described them as uncanny. Likewise, there were some strongly divided opinions on whether directed movement was preferred or not. Those that liked directed movement describe its smoothness. Those that disliked it describe it as being distracting and moving weirdly. The risk of anthropomorphised features in industrial robots distracting users from their task was also described in Onnasch & Hildebrandt (2022) and Perugia et al. (2021). However, interestingly it seems that this distractibility was only a problem for certain participants, and was not reflected in the task performance measure.

Some participants wrote that specifically the predictability of one robot arm made them prefer it over the others. This is interesting for the preference questions, where we see a discrepancy between which robot was easiest to work with (thus most predictable), and which was actually preferred to work with. Most participants seem to prefer working with the robot arm that they like the most, rather than the robot arm with whom it is easiest to work with. It would be interesting to see whether this changes if participants were invested in the success of the collaboration, as would be the case in a real world industrial setting.

Some of the answers also voiced a strong dislike for the robot arm using undirected movement and no gaze cues. This is in line with this condition consistently scoring lowest in the user experience, preference and task performance measures. Interestingly there were no responses in the open questions that expressed explicit preference for this condition, yet the preference question still had a small minority that preferred this condition.

These answers all suggest that there are strong differences between participants in whether and why they prefer gaze cues and directed movement. This is unfortunate, as it means there is no single ‘best’ choice for user experience. Instead, our results indicates that different types of users prefer different designs. The current study did not investigate what underlying personal features cause these preferences. It might also be possible to shift these preferences, for instance through repeated interactions, or by explaining beforehand how these manipulations are meant to communicate intention. This could make it so that there arises a single preferred ‘best’ design.

Exploratory Findings

Bimodal distribution of Task Performance

The histograms of the task performance variable showed a bimodal distribution for the robot arm using undirected movement and gaze cues. It had peaks around t = -7 and t = -2. To a lesser extent this bimodal distribution was also seen in the robot arm using directed movement and no gaze cues, and the robot arm using directed movement and gaze cues. Interestingly, Ronckers (2022) had a similar problem with a bimodal distribution in task performance. However, this was only in the condition with a separate robot head giving gaze cues. There was no analogous condition in the current study. One explanation for the distribution could be that the peak at t = -2 represents participants that were unable to pick up on the cues of the gaze and directed movement. Therefore they only noticed what cup was going to be grasped by the robot arm quite late. This idea is supported by the single peak at t = -2 for the robot arm using undirected movement and no gaze cues, in which there were no cues from the eyes or the arm movement.

To test the idea that the two peaks represented participants that did, or did not, pick up on gaze cues, the data was split based on which participants mentioned the differences in gaze in the open questions. This was not possible to do with movement differences, as almost all participants mentioned movement differences in the open questions. This split had no result though, as it seemed the participants were about equally divided among the two distributions.

An explanation for the two peaks could instead be that participants only noticed gaze cues later on in the experiment, after having experienced the other conditions. This would be in line with the fact that, for the robot arm using undirected movement and gaze cues, anticipation time significantly increased if the condition was presented later in the experiment. Although this sounds like a likely explanation, a problem with it is that a similar effect is not seen for the other two robot arms. Thus, while it remains unclear what caused this distribution, it does point to an unknown confounding factor that causes differences between individuals.

Effects of Cup Positions on Task Performance Measures

Inspecting the time it took the robot arm to reach each cup revealed that robot arms using directed movement had trouble with reaching the green and blue cup on the right. These cups were consistently reached later than the other cups. It is unknown what caused this discrepancy between the current experiment and Ronckers (2022). In order to alleviate potential problems with this discrepancy, the dependent variable for task performance was changed from the time between the start of the movement of the robot arm and the participant pressing the button (as was done in Ronckers (2022)), to the time difference between the participant pressing the button and the arm grasping the cup.

Nevertheless, we still found large differences in anticipation time between the different cups across all conditions. Specifically, the cups in the middle were consistently easier to anticipate than those on the edges. In Ronckers (2022), the same pattern was found for the robot arm using undirected movement. However, there were no differences between cups for the robot arm using directed movement in Ronckers (2022). This is a big discrepancy, and again points to a confounding effect of the presence of the humanlike eyes, as this was the main difference between studies. Alternatively, the difference in anticipation time could have been caused by the inconsistent time it took the robot arm using directed movement to reach each cup. However this does not explain why the effect occurred on both left and right sides of the table.

Limitations

There were some technical problems with the simulation of the robot arm and the virtual environment. For reasons that could not be figured out within the timeframe of the project, the robot arm did not always perform consistently. The biggest problem was that sometimes, while using directed movement, the robot arm would not grasp the cup, even though it reached its destination. When this happened, the robot arm responded as if it never reached its target, and kept trying to move forward towards the cup in a never-ending loop. Because this problem was already identified during piloting, a reset function was added to be controlled by the experiment leader. This made the arm move back to its base position after which it tried moving to the cup again. For some reason, the arm would then successfully be able to grasp the cup. This reset function was needed four times in total throughout all trials of the entire experiment and happened for three participants.

Furthermore, the robot arm sometimes dropped the yellow cup prematurely when moving it to the box. This occurred a total of 23 times (roughly one third of all yellow cups) and affected both the yellow cup on the left and right side of the table. It is unknown what caused this problem and why it specifically affected the yellow cup, as the specifications for all cups were the same. Through a timer that started when a cup was picked up, the cups were programmed to disappear shortly after being dropped by the robot arm. This happened whether the cup was dropped in the box or on the table. This way, a cup that was dropped on the table would not be distracting the participant. Nevertheless, problems with the robot arm might have influenced the collaboration for the participants that experienced it. One participant wrote in the open questions that seeing the robot arm drop a cup made it appear more incompetent and unintelligent.

Aside from these outright mistakes, there also were inconsistencies in how much time it took for the arm using directed movement to reach each cup. The robot arm using directed moved consistently needed a bit more time to reach the two cups on the far right. One participant described noticing this in the open questions as well. It remains unknown what caused this inconsistency, as it was not seen in Ronckers (2022). Further analysis showed that there were large differences in task performance between cups as well, with those in the middle having consistently faster anticipation times. This is not necessarily a problem, as statistical models can correct for this. However, it should be considered when it comes to practical implementations. Ronckers (2022) also found significant differences between cup positions and reaction times, although the effects were larger in our study for unknown reasons.

The movement of the eyes could also have been implemented better. Four participants described the eyes as "shaking". In its current implementation, the eyes' position and orientation would update every frame to match the position of the arm and (if required) the gaze cues. This could have given the impression of shaking.

Finally, there were also some problems in the questionnaire. First, one of the questions for eeriness, (ordinary – supernatural), was left out due to researcher error. Additionally, because the open questions were posed only after the user experience questionnaire, the vocabulary used by participants in the open questions might have been influenced by the words in the questionnaire. This was seen in some participants that essentially just wrote what they filled in in the questionnaire. Finally, in the answers to the open questions, it was not always clear to what manipulations the participants referred. For instance, a word as "gaze" could refer to the gaze cues of the eyes, or the gripper of the robot arm pointing towards the target in directed movement. The open questions could be rephrased to ask about these manipulations specifically.

Implications & Future Research

The manipulations of directed movement and gaze cues influenced the task performance and user experience in different ways. For task performance, the effects of directed movement were larger than the effects of gaze cues. Although the effect sizes were relatively similar. For the user experience, gaze cues showed more significant and stronger effects compared to directed movement. The preference questions also showed that participants liked the robot arms using gaze cues more overall. Nevertheless, for both task performance and user experience, the best results were found when directed movement and gaze cues were combined.

Therefore, when it comes to applying our findings in practice, the best option would be to implement both directed movement and gaze cues. However, the costs of implementing gaze cues through added eyes are probably higher compared to the costs of changing the movement trajectories of the robot arms. Combined with the fact that directed movement has a larger effect on task performance than gaze cues, companies might choose to only implement directed movement. This will come at the cost of a potentially better user experience for the operators. In this trade-off, employers should consider that an increase in user experience might make operators more engaged in their jobs (Meneweger, Wurhofer, Fuchsberger, & Tscheligi, 2018), and that this engagement might lead to better task performance overall (Markos, 2010). Nevertheless, we do not know how our findings will translate to an actual factory setting, as the participants in our sample did not represent typical factory operators.

An unexpected finding was the strong contrast in the open questions between how participants responded to both directed movement and the gaze cues. There were some that clearly liked the gaze cues, and some that clearly disliked them. The same goes for movement. This division was reflected in the preference questions as well, where there was no obvious preference for a single robot arm either. Some participants wrote that predictability was most important for their preference, although the preference questions show that likeability is more important for most participants. This is a problem for practical implementations, as some users might work better with a different robot arm than other users. It would be interesting to see if we can change the opinions of users that disliked certain robot arms to instead view them more favourably. For instance, through repeated interaction or by explaining how the manipulations are meant to communicate intention.

It was also odd that some participants did not mention gaze cues in the open questions at all, especially since gaze cues strongly influenced the user experience. The bimodal distribution found in task performance could also have been due to differences in how people perceive and interpret the manipulations. Perhaps some users do not notice the gaze cues at all. These users might still benefit from the gaze cues, if they were made aware of them. It would be interesting to see how this would affect the effect size of gaze cues compared to directed movement. Future research could also to investigate how personal factors influence people's interpretation of the movement and gaze cues of the robot arm. For practical significance, it would be most interesting to see what these personal factors are like in actual factory workers.

The context around the collaborative task could also be changed to better match a real-world setting. Participants would likely prefer a more efficient robot over a more likeable robot if there was something to gain from an efficient collaboration, for instance a monetary reward. Likewise it would be interesting to see how changing the viewing angle for the collaboration task, or the positions of the cups could change the task performance and user experience. One participant also described a desire to move around the table, in order to view the robot arm from different angles. Furthermore, the effects could be different if participants were explicitly told to focus on the directed movement or the gaze cues, and their workings were explained beforehand. Earlier research has shown that explanations for a robot's behaviour can cause robots to be perceived more favourably due to top-down processing (Stenzel et al., 2012). We also do not know how the task performance and user experience of the collaboration would change over repeated interactions. The current results only represent a first impression of the arm. The significant effect of condition ordering found on task performance already suggests that people learn to make better use of the gaze cues over repeated interactions.

We also do not know how the mere presence of the eyes on the robot affected the outcome of the experiment. There were quite some discrepancies between our results and those found by Ronckers (2022). We were mainly surprised to find no significant effect for directed movement in most user experience categories, while this was the case in Ronckers (2022). Instead, the results that best match the results of Ronckers (2022) were found by the robot arm using both directed movement and gaze cues. This begs the question of how Ronckers (2022) achieved these effects without the use of gaze cues. This could stem from how the eyes were presented, as this is the only major difference between the studies. Our study had humanlike eyes present in every condition, whilst Ronckers (2022) either had either nothing or sunglasses to represent eyes. It could also be that the sunglasses, whilst not giving explicit gaze cues, still gave the same effect of gaze cues by their presence alone. This could mean that people 'fill in' the gaze cues themselves.

Future research could further investigate how different embodiments of vision (or lack thereof) affect task performance and user experience. It could be that the presence of the eyes, when no gaze cues were used, distracted the participants from their task. This would be an example of an anthropomorphic feature serving no functional purpose and thus being distracting, as was found by Onnasch & Hildebrandt (2022). Perugia et al. (2021) links this distractibility to uncanniness. Although uncanniness was not thought to have negatively influenced our outcome, we did see an increase in uncanniness when gaze cues were used. It would be interesting to see how different embodiments of vision, whether they use gaze cues or not, influence this uncanniness and distractibility. For instance using the sunglasses from Ronckers (2022), or making different variations to the dimensions and position of the eyes, or adding features to represent saccades and blinking behaviour.

Conclusion

The present study aimed to investigate the use of gaze cues and directed movement in making the intentions of an industrial robot arm more clear. It was found that both directed movement and gaze cues had a similar positive effect on task performance, with the effects of directed movement being slightly stronger. The best task performance was found when both directed movement and gaze cues were combined. Although, whether these gaze cues should be in combination with directed or undirected movement differed among participants.

Based on the results of the current study, optimal collaboration with a robot arm in an industrial setting would have the robot arm utilize both directed movement and gaze cues. This would achieve both the most efficient task performance and yield the best user experience for the operator. However, the implementation costs of attaching eyes with adequate functionality to existing robot arms is probably much greater than the costs of adapting the movement trajectories of an existing robot arm. The benefit to task performance provided by gaze cues when directed movement is already implemented is also relatively small. Not implementing gaze cues will come at the cost of a potentially better user experience though, so the industry might have to make a trade-off there. Still, implementing just directed movement could be a valid cost-effective measure to make the intentions of an industrial robot arm more clear. By making these intentions more clear, collisions between the robot arm and its operators can be more easily avoided. Thus, improving both efficiency and safety in industrial human-robot collaboration.

References

Admoni, H., & Scassellati, B. (2017). Social Eye Gaze in Human-Robot Interaction: A Review. *Journal of Human-Robot Interaction*, *6*(1), 25. https://doi.org/10.5898/jhri.6.1.admoni

Aronson, R. M., Santini, T., Kübler, T. C., Kasneci, E., Srinivasa, S., & Admoni, H. (2018). Eye-Hand Behavior in Human-Robot Shared Manipulation. *ACM/IEEE International Conference on Human-Robot Interaction*, *18*, 4–13. https://doi.org/10.1145/3171221.3171287

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Bergman, M., De Joode, E., De Geus, M., & Sturm, J. (2019). Human-cobot teams: Exploring design principles and behaviour models to facilitate the understanding of non-verbal communication from cobots. *CHIRA 2019 - Proceedings of the 3rd International Conference on Computer-Human Interaction Research and Applications*, 191–198. https://doi.org/10.5220/0008363201910198

Bicho, E., Louro, L., Hipólito, N., & Erlhagen, W. (2009). A dynamic field approach to goal inference and error monitoring for human-robot interaction. *Adaptive and Emergent Behaviour and Complex Systems - Proceedings of the 23rd Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour, AISB 2009*, (January), 31–37.

Blanca, M. J., Arnau, J., García-Castro, F. J., Alarcón, R., & Bono, R. (2023). Non-normal Data in Repeated Measures ANOVA: Impact on Type I Error and Power. *Psicothema*, *35*(1), 21–29. https://doi.org/10.7334/psicothema2022.292

Boucher, J. D., Pattacini, U., Lelong, A., Bailly, G., Elisei, F., Fagel, S., … Ventre-Dominey, J. (2012). I reach faster when i see you look: Gaze effects in human-human and human-robot face-to-face cooperation. *Frontiers in Neurorobotics*, *6*(MAY), 1–11. https://doi.org/10.3389/fnbot.2012.00003

Clodic, A., Pacherie, E., Alami, R., & Chatila, R. (2017). Key Elements for Human-Robot Joint Action. *Studies in the Philosophy of Sociality*, 159–177. https://doi.org/10.1007/978-3-319-53133-5_8

Coumans, E., & Bai, Y. (2016). *PyBullet, a Python module for physics simulation for games, robotics and machine learning*. Retrieved from pybullet.org

Cuijpers, R. H., Ruijten, P. A. M., & Goor, V. J. P. va. den. (2020). Legibility of Robot Approach Trajectories with Minimum Jerk Path Planning. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-030-62056-1_33

Dağlarlı, E., Dağlarlı, S. F., Günel, G. Ö., & Köse, H. (2017). Improving human-robot interaction based on joint attention. *Applied Intelligence*, *47*(1), 62–82. https://doi.org/10.1007/s10489-016-0876-x

Damiano, L., & Dumouchel, P. (2018). Anthropomorphism in human-robot co-evolution. *Frontiers in Psychology*, *9*(MAR), 1–9. https://doi.org/10.3389/fpsyg.2018.00468

Dragan, A. D., Bauman, S., Forlizzi, J., & Srinivasa, S. S. (2015). Effects of Robot Motion on Human-Robot Collaboration. *ACM/IEEE International Conference on Human-Robot Interaction*, *2015-March*, 51–58. https://doi.org/10.1145/2696454.2696473

Fink, J. (2012). Anthropomorphism and human likeness in the design of robots and human-robot interaction. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7621 LNAI*(October), 199–208. https://doi.org/10.1007/978-3-642-34103-8_20

Flash, T., & Hogan, N. (1985). The coordination of arm movements: An experimentally confirmed mathematical model. *Journal of Neuroscience*, *5*(7), 1688–1703. https://doi.org/10.1523/jneurosci.05-07-01688.1985

Hammoud, R. I. (2008). Passive Eye Monitoring: Algorithms, Applications and Experiments. *Springer Science & Business Media*, 481. Retrieved from

http://www.springer.com/us/book/9783540754114
Ho, C. C., & MacDorman, K. F. (2017). Measuring the Uncanny Valley Effect: Refinements to Indices for Perceived Humanness, Attractiveness, and Eeriness. *International Journal of Social Robotics*, *9*(1), 129–139. https://doi.org/10.1007/s12369-016-0380-9
Hostettler, D., Mayer, S., & Hildebrand, C. (2022). Human-Like Movements of Industrial Robots Positively Impact Observer Perception. *International Journal of Social Robotics*, 1–19. https://doi.org/10.1007/s12369-022-00954-2
Huber, M., Rickert, M., Knoll, A., Brandt, T., & Glasauer, S. (2008). Human-robot interaction in handing-over tasks. *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN*, 107–112. https://doi.org/10.1109/ROMAN.2008.4600651
Khoramshahi, M., Shukla, A., Raffard, S., Bardy, B. G., & Billard, A. (2016). Role of gaze cues in interpersonal motor coordination: Towards higher affiliation in human-robot interaction. *PLoS ONE*, *11*(6). https://doi.org/10.1371/JOURNAL.PONE.0156874
Kshirsagar, A., & Chen, J. (2015). *Toward bridging the gap between Social Robots and Industrial Robots*.
Kühne, R., & Peter, J. (2023). Anthropomorphism in human–robot interactions: a multidimensional conceptualization. *Communication Theory*, *33*(1), 42–52. https://doi.org/10.1093/ct/qtac020
Markos, S. (2010). *5. Employee_Engagement_The_Key_to_Improving.pdf*. *5*(12), 89–96.
Meneweger, T., Wurhofer, D., Fuchsberger, V., & Tscheligi, M. (2018). Factory workers' ordinary user experiences: An overlooked perspective. *Human Technology*, *14*(2), 209–232. https://doi.org/10.17011/ht/urn.201808103817
Mori, M. (1970). The Uncanny Valley: The Original Essay by Masahiro Mori. *IEEE Robotics & Automation Magazine*, *12*(Figure 1), 1–6. Retrieved from https://spectrum.ieee.org/automaton/robotics/humanoids/the-uncanny-valley%0Ahttps://spectrum.ieee.org/automaton/robotics/humanoids/the-uncanny-valley%0Ahttps://spectrum.ieee.org/automaton/robotics/humanoids/the-uncanny-valley
Nagamatsu, T., Sugano, R., Iwamoto, Y., Kamahara, J., & Tanaka, N. (2010). User-calibration-free gaze tracking with estimation of the horizontal angles between the visual and the optical axes of both eyes. *Eye Tracking Research and Applications Symposium (ETRA)*, *1*(212), 251–254. https://doi.org/10.1145/1743666.1743725
Onnasch, L., & Hildebrandt, C. L. (2022). Impact of Anthropomorphic Robot Design on Trust and Attention in Industrial Human-Robot Interaction. *ACM Transactions on Human-Robot Interaction*, *11*(1). https://doi.org/10.1145/3472224
Palinko, O., Rea, F., Sandini, G., & Sciutti, A. (2016). Eye tracking for human robot interaction. *Eye Tracking Research and Applications Symposium (ETRA)*, *14*(March), 327–328. https://doi.org/10.1145/2857491.2888590
Perugia, G., Paetzel-Prüsmann, M., Alanenpää, M., & Castellano, G. (2021). I Can See It in Your Eyes: Gaze as an Implicit Cue of Uncanniness and Task Performance in Repeated Interactions With Robots. *Frontiers in Robotics and AI*, *8*, 78. https://doi.org/10.3389/frobt.2021.645956
Rau, P. L. P., Li, Y., & Li, D. (2010). A cross-cultural study: Effect of robot appearance and task. *International Journal of Social Robotics*, *2*(2), 175–186. https://doi.org/10.1007/s12369-010-0056-9
Ronckers, M. L. M. (2022). *Animal-like Referential Gaze in a Robot Arm improves Human-Robot Collaboration: Going Beyond Humanoid Design*. Eindhoven University of Technology.
Sacheli, L. M., Arcangeli, E., & Paulesu, E. (2018). Evidence for a dyadic motor plan in joint action. *Scientific Reports*, *8*(1), 5027. https://doi.org/10.1038/s41598-018-23275-9
Sauer, V., Sauer, A., & Mertens, A. (2021). Zoomorphic Gestures for Communicating Cobot States. *IEEE Robotics and Automation Letters*, *6*(2), 2179–2185. https://doi.org/10.1109/LRA.2021.3060416
Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., … Araya-

Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, *11*(9), 1141–1152. https://doi.org/10.1111/2041-210X.13434

Schmitz, C., & Limesurvey-Project-Team. (2022). *LimeSurvey: An Open Source survey tool*. Retrieved from www.limesurvey.org

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Sharkawy, A.-N. (2021). Minimum Jerk Trajectory Generation for Straight and Curved Movements: Mathematical Analysis. In Y. Sergey Y. (Ed.), *Advances in Robotics: Reviews, Book Series, Vol. 2* (pp. 187–201). Retrieved from http://arxiv.org/abs/2102.07459

Siposova, B., & Carpenter, M. (2019). A new look at joint attention and common knowledge. *Cognition*, *189*, 260–274. https://doi.org/10.1016/j.cognition.2019.03.019

Stenzel, A., Chinellato, E., Bou, M. A. T., Del Pobil, Á. P., Lappe, M., & Liepelt, R. (2012). When humanoid robots become human-like interaction partners: Corepresentation of robotic actions. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(5), 1073–1077. https://doi.org/10.1037/a0029493

Tatarian, K., Stower, R., Rudaz, D., Chamoux, M., Kappas, A., & Chetouani, M. (2022). How does Modality Matter? Investigating the Synthesis and Effects of Multi-modal Robot Behavior on Social Intelligence. *International Journal of Social Robotics*, *14*(4), 893–911. https://doi.org/10.1007/s12369-021-00839-w

Terzioglu, Y., Mutlu, B., & Sahin, E. (2020). Designing social cues for collaborative robots: The role of gaze and breathing in human-robot collaboration. *ACM/IEEE International Conference on Human-Robot Interaction*, 343–357. https://doi.org/10.1145/3319502.3374829

Walters, M. L., Syrdal, D. S., Dautenhahn, K., te Boekhorst, R., & Koay, K. L. (2008). Avoiding the uncanny valley: Robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Autonomous Robots*, *24*(2), 159–178. https://doi.org/10.1007/s10514-007-9058-3

Weiss, A., & Bartneck, C. (2015). Meta analysis of the usage of the Godspeed Questionnaire Series. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, *2015-Novem*(March 2017), 381–388. https://doi.org/10.1109/ROMAN.2015.7333568

Zhao, M., Shome, R., Yochelson, I., Bekris, K., & Kowler, E. (2016). An experimental study for identifying features of legible manipulator paths. *Springer Tracts in Advanced Robotics*, *109*, 639–653. https://doi.org/10.1007/978-3-319-23778-7_42

Appendix 1 – Questionnaire items as presented in the experiment

Godspeed Questionnaire

The following questions are about how you perceived the robot-arm and your interaction with it.

Please rate your impression of the robot on these scales:

1. Fake – Natural
2. Machinelike – Humanlike
3. Unconscious – Conscious
4. Artificial – Lifelike
5. Moving Rigidly – Moving Elegantly

6. Dead – Alive
7. Stagnant – Lively
8. Mechanical – Organic
9. Inert – Interactive
10. Apathetic – Responsive

11. Dislike – Like
12. Unfriendly – Friendly
13. Unkind – Kind
14. Unpleasant – Pleasant
15. Awful – Nice

16. Incompetent – Competent
17. Ignorant – Knowledgeable
18. Irresponsible – Responsible
19. Unintelligent – Intelligent
20. Foolish – Sensible

21. Anxious – Relaxed
22. Agitated – Calm
23. Quiescent - Surprised

Uncanniness Questionnaire

The following questions are about how you perceived the robot-arm and your interaction with it.

What did you think about the movements of the robot?

1. Synthetic – Real
2. Inanimate – Living
3. Human-made – Humanlike
4. Mechanical Movement – Biological Movement
5. Without Definite Lifespan – Mortal

What are your feelings about the robot?

1. Reassuring – Eerie
2. Dull – Freaky
3. Uninspiring – Spine-tingling
4. Boring – Shocking
5. Predictable – Thrilling
6. Bland – Uncanny
7. Unemotional – Hair-raising
8. Plain – Weird

What did you think of the robot's appearance?

1. Ugly – Beautiful
2. Repulsive – Agreeable
3. Crude – Stylish
4. Messy – Sleek

Open Questions

Finally, we have some open questions regarding the experiment as a whole. Please let us know what you thought of the four robot-arms you experienced.

1. How are the robot-arms different from each other? Please elaborate on your experience.
2. Why would you prefer one robot over the others? Please elaborate on your experience.
3. Do you have any other comments on the robot or the experiment you would like to share?

Forced Choice Preference Questions

1. Which robot were you the fastest with?
2. Which robot was the easiest to work with?
3. Which robot did you like the most?

4. Which robot would you prefer to collaborate with in a virtual environment (VR)?
5. Which robot would you prefer to collaborate with in a real-world setting?

Appendix 2 – Texts used in the Virtual Environment

General Practice

Starting text:

"Make sure your fingers are on the right buttons. See instructions on the table"

After 10 seconds:

"Try to add a label to the red cans. Press the corresponding button"

After an additional 10 seconds:

"Try to add labels to the other cans. Press the other buttons"

After an additional 10 seconds:

"To start a practice task: Press the SPACEBAR"

After pressing the spacebar, while the general practice task is happening:

"Press the button as soon as you know the colour of the can to which the line is going"

After the general practice task is complete:

"To repeat this practice task: Press the SPACEBAR Otherwise indicate to the experiment leader that you want to continue"

Collaboration Task

Starting text:

"First we do 1 practice round with 2 cans. Make sure your fingers are on the right buttons Press SPACEBAR to start"

During second practice:

"Press the button as soon as you know the colour of the can which the robot is going to collect"

Before starting first box

"Now we start the real task Press the SPACEBAR to start with the first box"

Before starting second box:

"To continue with the next box press the SPACEBAR"

After completing second box:

This was the end of this task. You can take off the headset and fill in the questionnaire