UMEÅ UNIVERSITET

Predictive gaze in action selection within virtual reality

Elpida Bampouni

Acknowledgements

I would like to express my gratitude to my friends and family for supporting me each in their own way during the course of my studies. I am also thankful to the many teachers I have had for facilitating my learning experience and overall development.

Definitions

Manual/Bimanual manipulation task: A task that requires one hand/both hands activity to be completed e.g. grasping, pointing or rotating an object to fit certain requirements and so on.

Abstract

The human gaze is pivotal in motion planning and control. Gaze is typically directed at visual target sites prior to physical interactions with them. This proactive gaze (PEG) behavior has been observed in a multitude of physical situations. However, PEG has not been examined in virtual reality (VR). Identification of PEG in VR could be helpful for digital human modeling applications and human-robot interactions. In this study we asked 10 participants to perform a pick-and-place (PAP) task in VR while we were tracking gaze behavior. Our results indicate that PEG also occurs in VR. Furthermore, the action to reach directly towards the PAP object or walk to it before reaching, results in different gaze strategies. Relocating before a reach is associated with gaze to additional sites, such as the floor and the table upon which the object was placed.

Keywords: predictive gaze, virtual reality, eye focus, grasping reach, action selection

Abstrakt

Den mänskliga blicken är avgörande för planering och kontroll av rörelse. Blicken riktas vanligtvis mot visuella mål före interaktion med dessa. Denna proaktiva blick (eng. 'proactive gaze'; PEG) har observerats i många olika slags fysiska situationer. Dock har inte PEG undersökts i virtual reality (VR). Identifiering av PEG i VR skulle kunna vara användbart för applikationer med digital mänsklig modellering och människo-robotinteraktioner. I denna studie instruerade vi 10 försöksdeltagare att utföra en s.k. pick-and-place-uppgift (PAP) i en VR-miljö medan vi registrerade deltagarnas blick. Våra resultat indikerar att PEG också förekommer i VR. Vidare leder handlingen att direkt sträcka sig efter objektet till ett annorlunda blickbeteende jämfört med att först förflytta sig innan man sträcker sig efter PAP-objektet. Vid förflyttning innan man sträcker sig efter objektet fästs blicken på ytterligare områden såsom golvet och bordet som objektet placerats på.

Nyckelord: prediktiv blick, virtuell verklighet, ögonfokus, greppande, handlingsväljande

Introduction

In many everyday tasks, prediction of the states of both one's own body and the world outside the body is necessary for success (Clark, 2013). While human abilities to anticipate these states is often seen as grounded in the brain, observing the states of the brain in real-time and ecologically valid contexts is not yet technologically feasible. As a result, one way to observe and study human prediction behaviors is through studying human eye-gaze. Proactive gaze behaviors are eye movements that focus on locations where one expects information to be extracted, even if the information is not there yet (Anderson et al., 1997). Proactive gaze enables individuals to collect information necessary to guide actions (Anderson et al., 1997; Johansson et al., 2001). Such proactive gazing behaviors have been observed in humans as young as a few months old (Ambrosini et al., 2013) and even animals.

One can directly observe proactive gaze in ordinary activities e.g. when picking up a coffee mug, using the force, playing with a ball or simply moving a mouse cursor to click on a screen icon. Studies have shown participants exhibiting proactive eye gaze (PEG) in various activities from reading music or text, playing ping pong, typing (Anderson et al., 1997), scoring in basketball (Vickers, 2007a) and golf (Vickers, 1992) to tea making (M. Land et al., 1999)and even simple item transfers (Johansson et al., 2001). One explanation of PEG behavior is that it provides individuals with time to plan and execute a response that is both controlled and accurate in each situation (Anderson et al., 1997; Johansson et al., 2001). Training techniques have also employed PEG practices, like the "Quiet Eye". According to the Quiet Eye technique, being aware of and practicing proactive gaze prior to taking action, can help focus visual attention and ultimately improve performance (Vickers, 2007b; Causer et al., 2014; Harle & Vickers, 2001).

Eye tracking is routinely employed but has not always been subtle, accurate or comfortable when investigating PEG. For instance, in a study by Johansson et al.(2001), participants' heads had to be stabilized. To do this, participants were biting down on a steel plate covered in wax while performing a task situated on a plane directly in front of them. An eye-tracking camera, an infrared light source and a mirror would provide the gaze data (*see Figure 1)*. Other studies have used head-mounted eye cameras not designed for research purposes(e.g. for driving) (M. F. Land, 2009; M. Land et al., 1999). Eye tracking discretion is also discussed as a potential bias factor in the more recent Gesierich et al. (2008) study. Even with a remote recording system such as the Tobii eye tracker, participants could still infer the point of interest was their gaze, which according to the authors could have generated extreme gaze behaviors as noted in their results.

Virtual reality (VR) has grown to be a commonly used research method. A plethora of scientific studies have used VR as a method (Gaggioli, 2001; Magora et al., 2006; Cushman et al., 2008; Sauzéon et al., 2012) and with good reasons. VR poses several advantages making it a valuable and preferred method, allowing for realistic, safe and highly controlled representations of multiple environments. In 2011, Wilson et al. demonstrated PEG behaviors in surgeons performing a laparoscopic simulation. However contrary to the terminology used in the paper, the equipment utilized did not support a VR setting. Regardless of the increasing VR practice, it appears there is a serious lack of studies regarding PEG activation.

In the present study we explore PEG behaviors during a pick-and-place task involving action selection within a VR environment. In a pick-and-place (PAP) task the individual picks up a given object and places it at a desired location. Complexity can vary, but even simple versions can offer valuable information about task constrained gaze behavior while providing opportunities for PEG. In this task participants pick up small boxes from a table and place them in a bin nearby. Similar VR PAP experiments have successfully been conducted (Lamb et al., 2017) but did not record gaze behavior.

The purpose of this study is to extend research and further our understanding regarding proactive gaze behaviors. However, this is not the only goal of this project. Since there is insufficient information about PEG activation within VR, one of our aims is to address this gap and hopefully contribute to a more informed research technique. Additionally, the motion tracking systems integrated in our VR system permit a range area of movement larger than most previous PEG studies. This allows for our final aim in this study which is to investigate gaze behaviors in a PAP task requiring whole body movement.

Figure 1 Eye tracking mechanisms have often been restrictive and static. Image from Johansson et al. 2001. (Reprinted with author's permission.)

Notably, this study is part of a larger project and thus future secondary aims have been set. It is our intention to further utilize the collected data in directions we will not explore within the boundaries of this thesis. However, a short description could provide a more holistic rationale behind our decision making. To begin with, a long-term aim for this project is to use accumulated data to advance dynamic models for action selection e.g. Lamb et al., 2017. Another long-term aim for this project is to direct our findings towards digital human modeling (Billing, Hanson, et al., 2019) and human-robot collaboration (Billing, Sciutti, et al., 2019) similarly to previous studies using PAP tasks and PEG (Perumaal & Jawahar, 2013; Huang et al., 2015).

Background

Some of the earliest studies on proactive gaze come from Ballard and colleagues in the 1990s with a block task being copied by the participants (Ballard et al., 1992). The results showed gaze fixation close to action sites preceding the motor actions by half to one second. Their findings sparked an interest in proactive gaze research and many studies gradually emerged. For example, a few years later Land et al. (1999) moved towards an unconstrained daily activity study, examining gaze during tea making. Two years later Johansson et al. (2001) followed with a more controlled block stacking task. Which in turn, inspired a similar study conducted on a computer(Gesierich et al., 2008). As we go through the background we will discuss some of those studies in more detail.
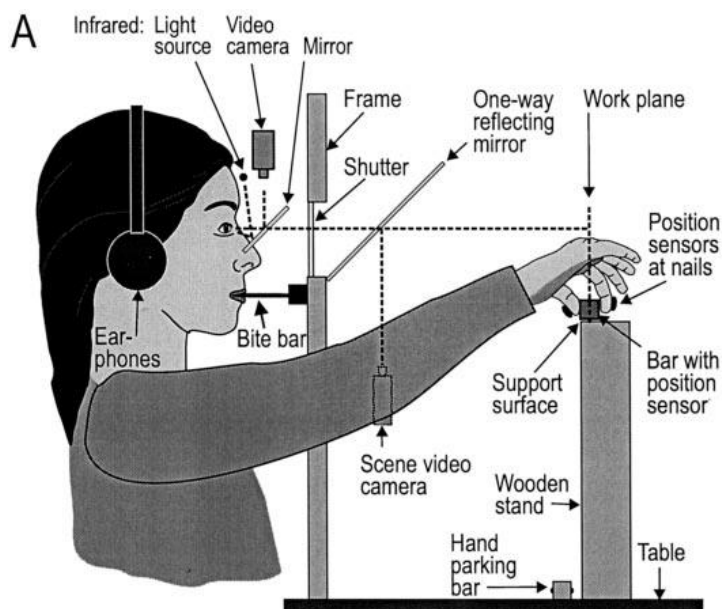
In a study of 2001 by Johansson and colleagues, the aim was to identify the critical landmarks our vision is drawn to when performing a pick and place stacking task with small blocks. Nine participants were asked to reach for a small bar and place it on a target area either directly or around obstacles. Their results demonstrated exclusive gaze fixations in landmarks crucial for better control. The fixations were directed on the grasp site of the bar, the final target area and around obstacles intermediate to their direct path; in other words informative locations which were critical for successful manipulation and task completion. The study concluded that eye-hand coordination is crucial for hand motion planning. As a result, the authors reflect on gaze behaviors as a predictor of motor manipulation occurrences.

In 2008, Gesierich et al. inspired by the study of Johansson et al., 2001 proceeded to a similar study using a computer screen. Tactile information plays an important role in manual manipulation tasks (J. R. Flanagan et al., 2006), however PEG activation is not solely limited to physical tasks. Participants of this study were instructed to move bars from one location portrayed on the screen to another using only a computer mouse for object manipulation. PEG activation while performing this computerized task was similar to the results of Johansson et. al. (2001). Gesierich et al. demonstrate that proactive gaze can occur without a direct physical interaction with the object. Analogous findings come from Sailer et al., 2005 who used a novel bimanual controlling device for navigating a on screen task with targets. Their results show that after several trials, participants started predicting their on screen target sites and proactively prepare to mark them.

Physical and tactile information, like weight factors are also important in object manipulation and contribute to the overall experience of the actor. The objects used in our study have no weight due to being virtual but will be "manually" handled and not guided by a mouse on a screen. Studies have shown that after a few trials, participants learn that an item can be much lighter or heavier than it appears, and adjust the applied force for the next trials (J. R. Flanagan et al., 2001). According to the authors, this demonstrates that sensorimotor memory can prevail over controversial visual size-weight stimuli. As a result, individuals can access their newly built internal models about an object's actual weight and adjust their actions accordingly.

Because gaze is proactive, intentional gaze shifts while the hand is still moving towards the target, are rare to observe. There are however exceptions, like when our attention needs to be divided. For example, attention was divided while participants made tea by Land et al., 1999. The study is also discussed in combination with another involving food preparation from Hayhoe in Land, 2009, because their results indicate similar patterns. In both tasks, gaze rarely diverted from the relevant targets. Fixations appeared to have monitoring functions, with a head start of about half a second before the hand arrived. The gaze monitoring functions identified were object *locating*, hand or object *directing* and *guiding,* as well as *checking* for information updates and feedback (Land, 2009). Notably, it was also found that parallel sub-tasks led to gaze transitions prior to establishing physical contact with an upcoming site in both studies (Land, 2009).

Similar early gaze shifts were noted in a bimanual study (Srinivasan & Martin, 2010) in which participants simultaneously moved two cylindrical objects to various target locations. In-depth analysis identified four eye-hand coordination strategies used almost exclusively when specific conditions were met e.g. depending on the size or distance of the targeting location. Most importantly participants in this study also exhibited a rapid gaze shift before finalizing their motions. The aforementioned studies required the use of two

hands as well as visual attention division for the completion of their respective tasks. It is believed that gaze shifts just prior to physical contact can occur in order to attend another task, pending for visual guidance. It is possible that the hand is sufficiently directed to no longer require continuous gaze guidance in order to succeed. Early gaze transitions are generally not as secure as visually guided motions and thus more rarely observed in cases where the attention is not elsewhere needed.

Current Study

Having established an informed grasp of the background, we can direct our attention to the study at hand. Our first objective in this study is to determine the possibility PEG occurrence in an immersive VR setting. Therefore our first hypothesis is as follows.

H1: PEG behaviors will occur in VR.

H01: PEG behaviors will not occur in VR.

We are also interested in the gaze differences between actions involving whole body movement, in this case direct reach trials and trials involving walking to the target before reaching. If indeed participants exhibit PEG behaviors, significant differences between the two actions could indicate the first steps towards a predictive model for action selection solely though eye gaze. Different gazing strategies and areas of interest, depending on the accessibility of the target object are expected according to PEG literature; leading to the formulation of our second hypothesis.

H2: PEG behaviors will differentiate between trials requiring different actions for PAP task completion.

H02: PEG behaviors will not differentiate between trials requiring different actions for PAP task completion.

This study is reported according to the following structure. Initially we start with a recount of the methods, including details about the sample, the equipment and the procedure. Following is a description of the analysis employed before presenting our results. Finally, we address our initial hypotheses and discuss our study findings.

Methods

The present study was conducted in the Interaction Lab at the University of Skövde. The lab is 60m2 of which about half is an open space dedicated to virtual reality (VR). This experiment took place in a VR setting where participants were asked to complete a pick-and-place (PAP) task. Participants were instructed to pick up virtual boxes from a table and place them in a bin located behind the starting location. In each trial a single box would appear in various distances for each of the 150 total trials. As a result participants would sometimes reach directly or need to relocate in order to reach the box. During the task, eye tracking and kinematic data were collected for each of the 10 participants. The project was submitted to the Swedish ethical board of Umea (2020-00677) for approval where it was decided to not review the application since there are no ethical concerns.

Figure 2

The VR room. The starting location is indicated by the bright yellow marks on the table and the gray bin. Participants could pick up boxes like the one on the table by either direct reaches or relocating.

Recruitment

Initially we started with a small number of colleagues for piloting and then moved forward to completely naive participants. Participants were recruited through variety of means, such as social media posts, flyers around campus and participants informing their social circles about an ongoing study. Participation was voluntary with only inclusion requirement being a right handed adult with normal or corrected vision. Participants of the experiment were compensated with an open cinema ticket.

Instruments and Equipment

VR headset: The headset used in this experiment was an HTC VIVE Pro Eye[1] with an imbedded eye-tracker used to follow participant gaze with 120 frames per second. The headset would also inform us of the participant's head position (temporal resolution 90Hz) and was tracked through Steam VR 2.0 tracking system, similarly for all our sensors.

Sensors: There were several wireless sensors used for kinematic data collection *(see Figure 3)*. First of all, a VIVE hand Controller (2018) would track the participants' right hand as well as allow them to interact with the box. Second, a VIVE Tracker was placed with

---

[1] Specification for the HTC Vive Pro Eye VR headset can be found at https://www.vive.com/eu/product/vive-pro-eye/

adjustable straps on the participant's right shoulder blade. Lastly, in order to indicate their steps, an additional tracker was worn at the participants left ankle.

The virtual environment: The virtual room (*see Figure 2*) resembled a simplified version of the actual laboratory room. There was a physical table 95x160cm and 87cm height precisely matched to a counterpart in the virtual world, so participants could accurately and safely interact with it. In regards to the PAP task, a virtual bin was located behind the participant's starting location. The distance between the table and the bin was 120cm. A door was also created and placed in the position of the physical door. Even though they would not interact with it, it was created to impede tension or possible anxiety due to being in a closed room alone, regardless of it being virtual. Finally for physical safety, in case a participant moved away from the central area of the interaction, a discrete blue line would appear on the floor indicating the borders.

Figure 3 The equipment and trackers placement.

PAP task: A pick and place task was specifically designed for this experiment and run in Unity3D 2019. The object was a dark yellow box 10x10x20cm (*see Figure 2*) that would appear randomly on the left table side in one of three distance regions (*see Figure 4*). The box appeared for 50 trials in each region on randomized positions within them. A box presented in the first region would be close enough for all participants to reach directly. The second and third regions were further away and thus participants' action selection could be to relocate before picking up the box.
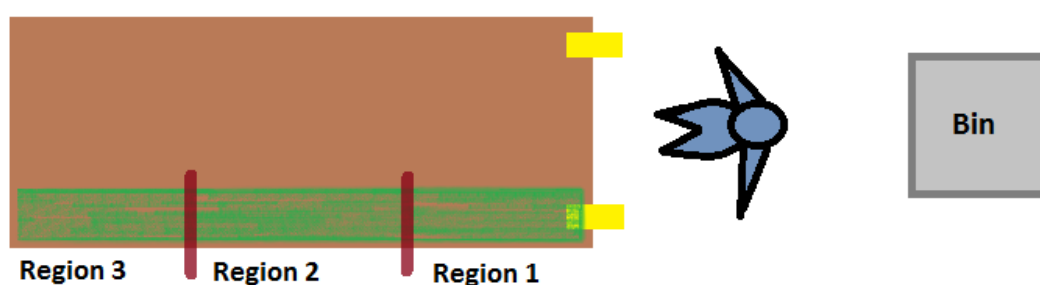

Figure 4 Panoramic representation of experimental set up. The table was divided in 3 equal distance regions, each represented with 50 random trials. The object would appear on the table within the green area.

In order to pick up the object the participant had to press a button on their hand controller, similar to a trigger (*see Figure 3*). After the box had been picked up, it was to be placed in the bin located behind the starting area. The starting area was located between the table and the bin, indicated by yellow marks on Figure 2 and 4. To drop the box, participants simply released the trigger. Dropping the box in the bin would signify the end of one trial.

However, before the next trial was initiated, participants had to stand within the starting location. A total of 150 trials took place for each participant, which lasted on average 16 minutes in VR. If a participant dropped the box, the same trial was repeated until the box was successfully placed in the bin.

Procedure

A detailed script had been constructed and would be followed for each participant in order to ensure similar treatment. After agreeing on a convenient time, individuals would be met by the experimenter at the entrance of the building where the lab is located. First of all, a consent form was be supplied and verbally explained to each participant. After they read and signed the form, a more detailed description of what their task entailed followed prior to starting the experiment. This included a time estimate of the overall procedure, visual representations of the room and the task from the experimenter's computer screen as well as instructions about the controller and the trackers. Participants were also reminded that they could take breaks or leave the experiment at any time without having to explain themselves and without losing their compensation.

In the next step participants were lead to the table and assisted in wearing the equipment. An identification code was assigned to ensure data de-identification before proceeding to a momentary eye tracking calibration check. Upon entering the virtual room, participants were given the instruction "*You will be picking up boxes and placing them into the bin. Some of the boxes might be too far to reach and so you may move around the table to grasp them*". Further instructions and assistance would be provided during the procedure if necessary.

After all the trials were finished, participants were asked about their experience, and could share any thoughts or feedback about the task. The experimenter took notes and filled out a form (*Appendix 1*) regarding demographics and daily life style information. Previous VR experience was also collected along with some physical measurements such as height or arm length. Participants then received a cinema ticket before the debriefing, and were provided the opportunity to ask any study related questions if they wished.

Sample

Our sample originally comprised 12 participants with an average age of 22 years. Two participants' data was excluded. The first set was excluded because the participant was too playful, exhibiting behaviors of throwing the box to the bin behind their back etc. The second participant's data was excluded because one of the motion trackers was misplaced on the right ankle instead of the left. In the end, our final sample consisted of 10 participants with a mean age of 21.7 years. In the final sample, 4 identified as female, 5 as male and 1 as non binary. All of them were studying at the University of Skövde at the time and all but one, had minimal VR exposure in daily life.

Analysis

Data pre-processing was conducted through functions built and executed in Python and R in order to select specific trial sections. For the purpose of this thesis, the data from each trial is divided into two phases. We will focus only on the first phase, which begins with the box appearance and ends with the participant grasping it. The second phase begins from

the grasp until the box is released in the bin. Final analysis and graphical representations were conducted using MATLAB.

The first priority was to determine a definition of action initiation. Based on previous works (Lamb et al., 2017, 2019; Fitts, 1954), a human reaches motion peak velocity around midway to the target. In this study we defined PEG when it occurred prior to the 5% of peak reaching velocity of the right hand. In other words, PEG was measured when the hand motion for contact towards the front of the participant had just started. Several studies have followed a similar approach to define the initiation of a reaching action (Valevicius et al., 2018; Le & Niemeier, 2014).

Due to this definition some trials were excluded (around 2/3rd) from the descriptive analysis (Results 1.1 and 1.2) but all 1500 trials have been used for the results presented in 1.3 and 1.4 sections. In the excluded trials, fixations to the object had not been registered prior to the 5% velocity peak. This could be the case for several reasons. To name a few, peripheral vision may also be sufficient for a person to spot an object without the need to fixate it. Furthermore, our participants were in constant motion and the 5% could sometimes be difficult to detect against the constant background motion noise. Overall, exclusion does not mean that there were no object fixations at all before contact with the object.

To analyze where participants were looking when reaching for the object, everything in the virtual room was categorized in one out of 8 eye-focus points we will refer to as areas of interest (AOI). The first AOI defined was the "Hand". Participants could track its position and rotation in VR in the form of their hand controller. The next AOI was the "Object", which contained fixations directed within a 15 cm radius from the box's epicenter. Further fixation points were the "Bin", the "Table", the "Walls" including the door, the "Floor" and the "Ceiling". Lastly, a final AOI was added collecting all unaccounted gazes e.g. while blinking or when eye-tracking data was not recorded properly, which we will call "Unknown".

For defining the action taken between direct reach (DR) and relocate and reach (RAR) trials, we used the ankle's position. The ankle tracker's transformation during each trial was recorded and each trial categorized as DR or RAR depending on how far left the foot was placed. An example of that function is demonstrated in Figure 5 which shows 150 trials of one participant and how we can differentiate the DR from the RAR trials regardless of region.
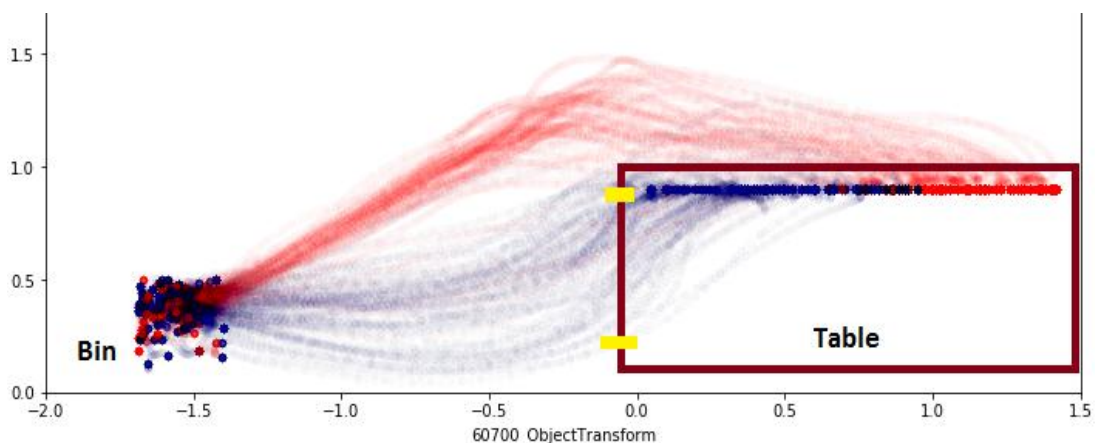

Figure 5 Example of ankle tracking sideways movements during DR (blue) and RAR (red) trials.

Results

1.1 PEG occurrence in VR

Trial sections from the moment a participant first realized a new box had appeared on the table until the reaching action initiation, demonstrate a collective gaze interest in the object as shown in *Figure 6*. Regardless the object being the smallest AOI in size, it still gets a significant portion of the total gaze fixations. With the term eye fixations in this study we refer to registered frames in the AOI.

Further analysis shows large differences between the three table regions (regions shown on Figure 4) where the object was generated. To begin with, trials in region 1 (R1), the nearest and fastest of all, are demonstrated in *Figure 7*. The first thing to notice is that since the trials are short the gaze spread in extremely low for all AOI with the exception of the target object.
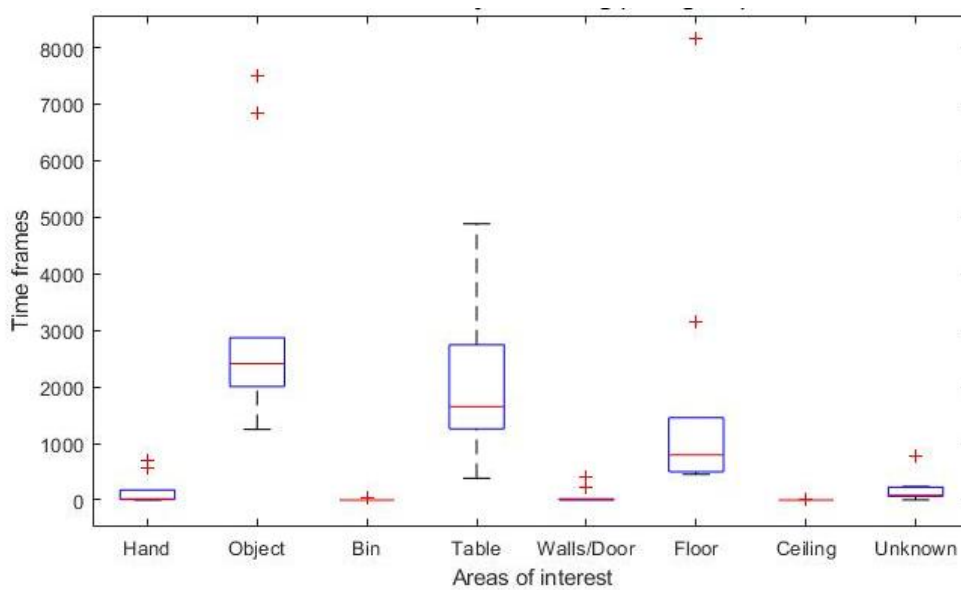

Figure 6 All participant fixations of AOI from the moment the box was spotted up until a reach was initiated (all regions included).
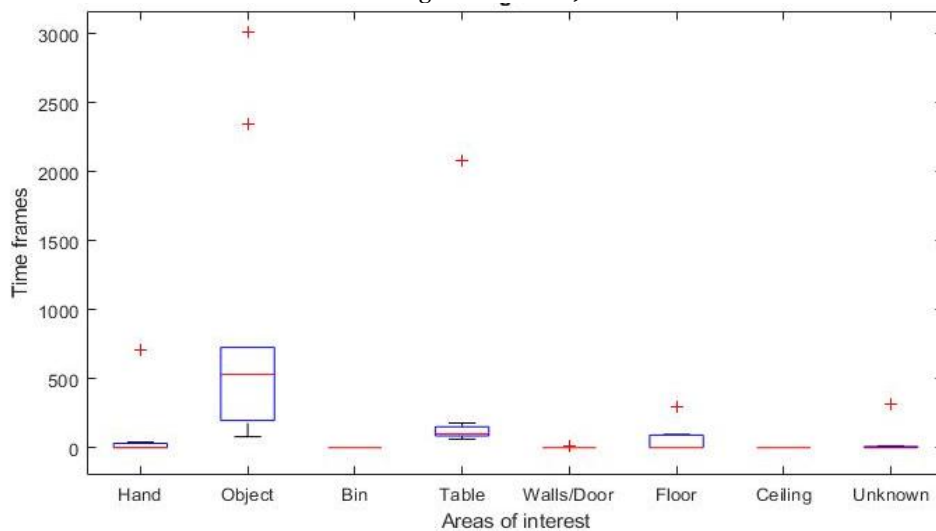

Figure 7 R1 trials and fixations from the moment the box was spotted until the reach action. Gaze is strongly focused on the target object.

However, the behavioral patterns look different in trials from region 2 and 3 conditions (R2 & R3 respectively). As it is obvious from *Figure 8 and 9* different AOI are gradually on the rise, presumably due to the increase in the distance. Moreover, we see that most gaze fixations away from the object are directed towards the table and floor.
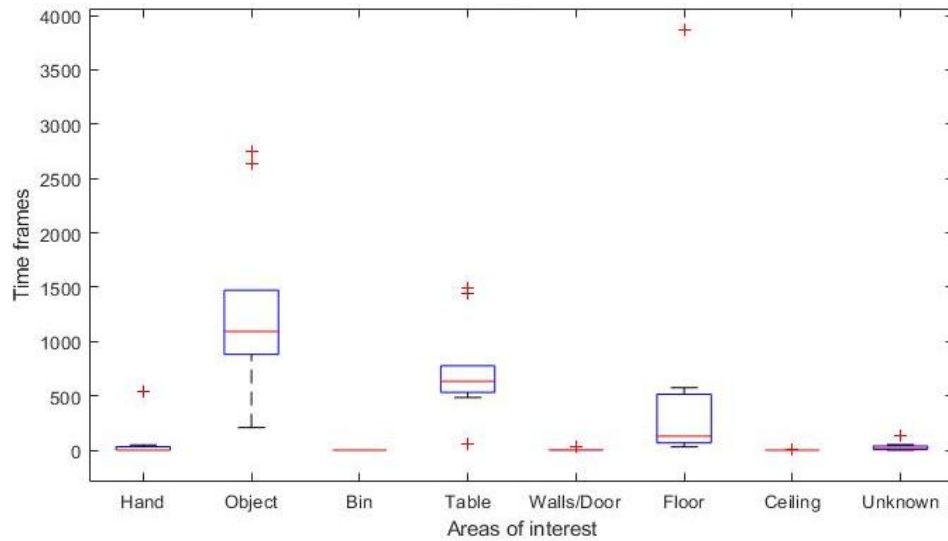

Figure 8 R2 trials and fixations from the moment the box was spotted until the reach action. Gradually more AOI are present.
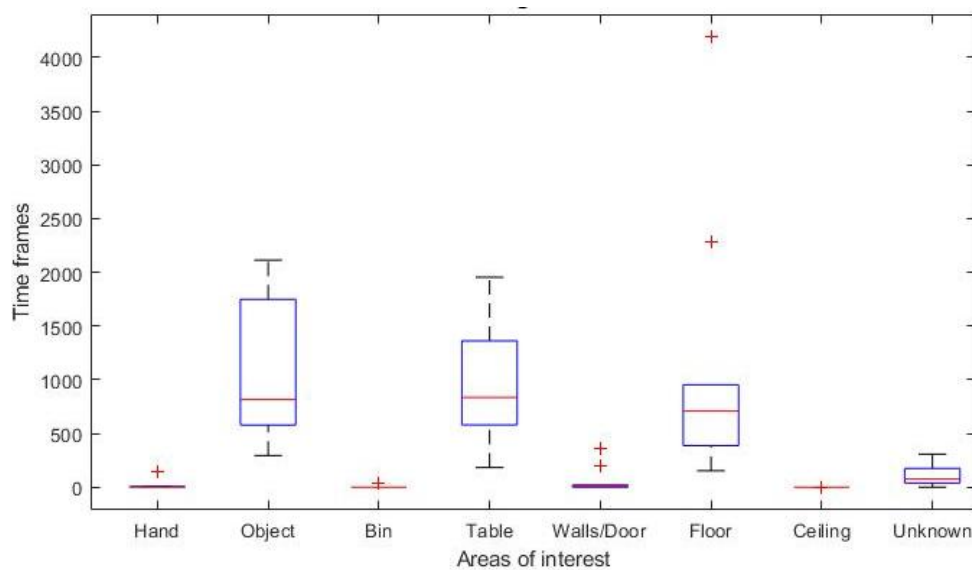

Figure 9 R3 trials and fixations from the moment the box was spotted until the reach action. Target object is the prevalent but surrounding AOI increase as well.

1.2 AOI and final frames of reaching action

Limiting the time frames to the last 300 (1 frame= 0.12ms) before the reach action begins shows in more detail the amount of fixations devoted on individual AOI. In the following graphs the closer to 1 in the Y axis indicates more fixations aimed on the AOI in question (1 represents 100% of gazes). For the X axis, 0 indicates the reach action initiation (5% of peak velocity).

Starting with the target object, we see once again that gazes in the final frames of R1 trials were almost exclusively aimed at the box (*Figure 10*) with a high average near 70%.

This number decreases dramatically for R2 and R3 as shown in *Figure 11 and 12*. Averages of object fixations in R2 and R3 plummet near 50% and 30% respectively.
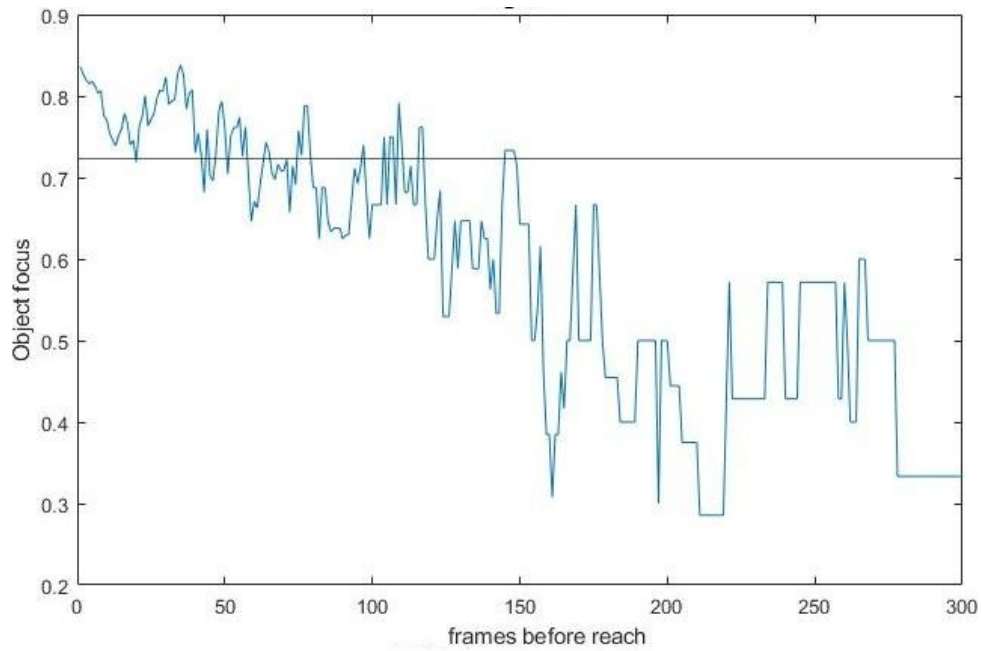

Figure 10 Average proportion of object gaze preceding 5% peak velocity of reaching action in R1. Horizontal bar indicates average proportion of object gaze.
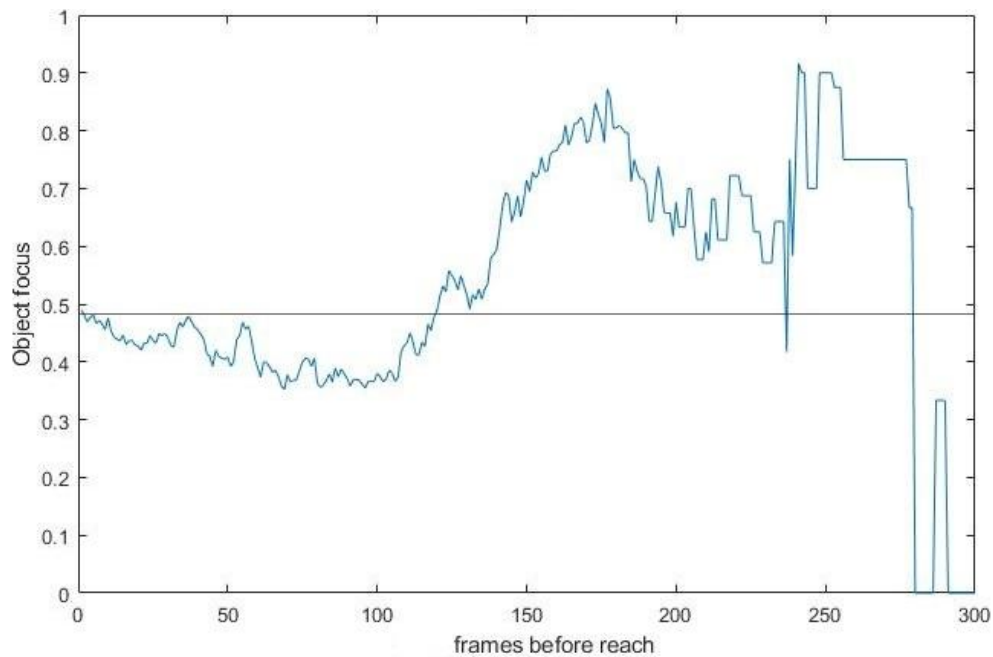

Figure 11 Average proportion of object gaze of reaching action in R2. Horizontal bar indicates average proportion of object gaze.
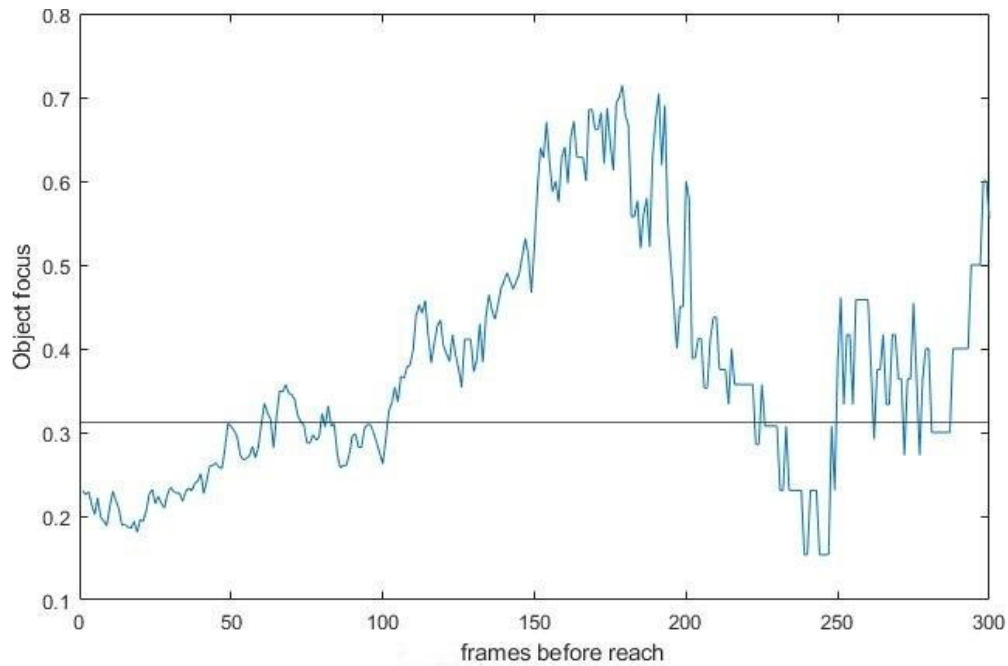
Figure 12 Average proportion of object gaze of reaching action in R3. Horizontal bar indicates average proportion of object gaze being near 30%.

Further analysis on R2 and R3 shows higher gaze fixations on AOI other than the object. More specifically, R2 demonstrated averages near 30% for the table and for 20% the floor. Notably, R3 showed averages around 30%for the table and 35% for the floor (*Figure 13*) antagonizing R3 object fixations during the final frames before the reach.
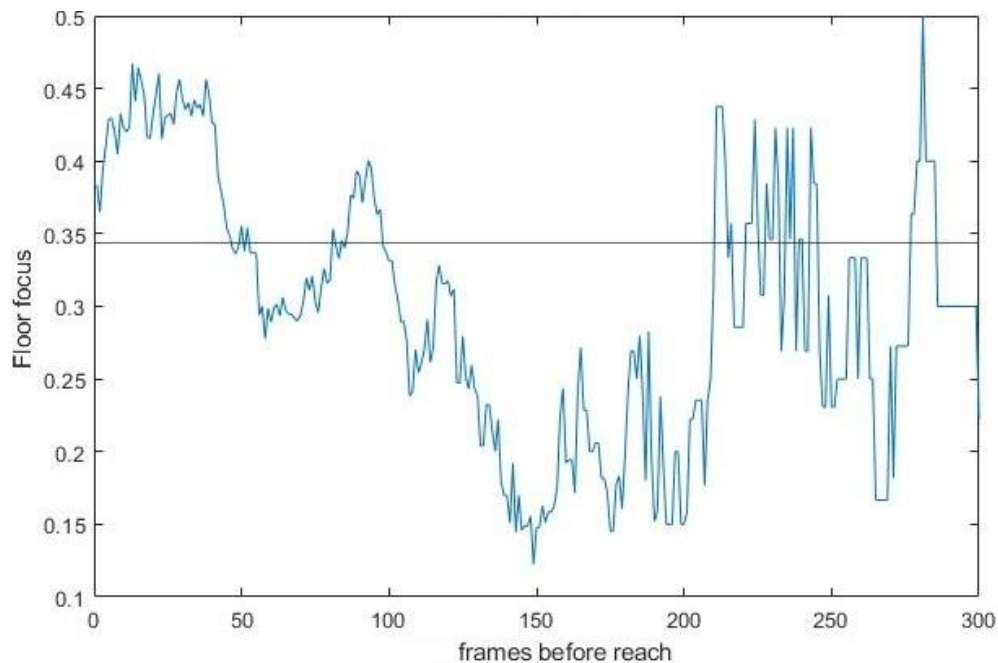

Figure 13 Average proportion of floor gaze preceding 5% peak velocity of reaching action in R3. Horizontal bar indicates average proportion of object gaze.

1.3 Selected action t-test

To address our second hypothesis, crude independent t-tests were conducted between DR and RAR trials, and duration of gaze spent on individual AOI, again during the final 300 frames before the reach. Out of 1500 trials, 880 were DR and 620 were RAR, which prevented a paired t-test comparison. The first AOI t-tested was the object, which showed no significant differences in gaze between the two actions with $p > 0.05$ (sd 0.24, df 1498).

Since our earlier data indicated a rise of table and floor gazes towards the farther regions, they were the next AOI tested. The t-test for the time spent gazing on the table was statistically significant $p < 0.01$ (sd 0.20, df 1498). Similarly, floor gazes were tested and too came out significantly different between the DR and RAR actions ($p<0.01$, sd 0.24, df 1498).

1.4 Selected action and "floor" logistic regression

Further analysis regarding the second hypothesis involved individual logistic regression analysis for each participant of the floor AOI and the action taken (DR and RAR). All participants showed statistically significant difference when gazing at the floor in relation to their action of DR or RAR ($p< 0.05$) apart from one who showed no difference in their gazing behavior (*Table 1*).

Table 1 Individual logistic regressions for AOI floor and action DR and RAR. All p1 values are statistically significant apart from participant 1.

Participant	Beta0	Beta1	p0	p1
1	-0.19	0.12	0.25	0.86
2	-1.52	42.72	0.00	0.00
3	-1.44	11.34	0.00	0.00
4	-0.78	51.09	0.00	0.00
5	-0.27	3.55	0.11	0.02
6	-0.81	7.22	0.00	0.00
7	-2.38	10.52	0.00	0.00
8	-1.51	3.09	0.00	0.00
9	-1.62	4.40	0.00	0.00
10	-0.81	11.59	0.00	0.00

Discussion

Our H1 appears to be confirmed, as PEG behaviors were universally present from all participants (even if not in all trials). The average gazes prior to the reaching action show that the target object was fixated. Nonetheless, it is important to point out that our data come from averages across trials and subjects and thus may be a poor indicator for individual trial behavior. This is the first time PEG is examined and demonstrated in VR, and our results are in agreement with the extensive literature on PEG which has been demonstrated in several different settings.

In *Figure 14,* we can also see a PEG example of a simultaneous physical and virtual representation. The participant has walked to a region 3 (R3 trial) and is just about to grab the object. On the screen on the left, we can see her gaze being projected as a brown axis,

extending from the white cube (indicating her head position) and piercing through her target, the dark yellow box.

Figure 14 A participant exhibiting PEG moments before picking up a box. The gaze is indicated on the screen by a brown line directed on the box, while the hand controller is about to arrive. The screen is portraying a view point from the right side, in relation to the starting location (indicates by yellow marks), and in relation to the photographer's perspective. (Photographic segment from a video recording requested by the participant and reprinted with their permission)

Our results indicate a substantial difference in AOI depending on the region of the object's appearance. Region 1 (R1) trials are characterized by a static element that arguably makes them simpler in comparison to the other region trials. For trials so close to the participant, one does not need to divert gaze from the target in order to reach it and there is minimal effort for these direct grasps. On the other hand, trials in R2 and even more so in R3 are a bit more complex. They involve potentially several sub-tasks before an individual can grab the object. A likely explanation of why gazes in R2 and R3 trials divert from the main target is because they are drawn towards additional targets integral to completing the task. Alternative targets can provide context information regarding the participants' next steps. In this case, gazes at the table or the floor likely offer rich information such as how close to the table one must stand to reach directly, whether they could reach by leaning on the table or do need to relocate. In case of relocation, they need to be aware of how far they are from the object and how to avoid obstacles in their path e.g. hitting their hip on the table (Johansson et al., 2001).

The regions offer valuable insight but they are only indicators of the actual action selection, which leads us to our second hypothesis regarding the gaze differences depending on the action section. According to the t-test analysis, it appears that there were no significant gaze differences towards the object between the people performing a DR and a RAR action. Although fixation distribution in RAR trials appears to be directed towards other AOI, there was no significant fixation difference to the main object during the final frames before the reach.

However, H2 appears to be confirmed because there were significant differences between DR and RAR trials for the table and floor areas. Even though our t-tests do not show which of the two actions lead to more gaze duration on alternative targets, it is safe to infer from our data and prior research, that RAR trials were the ones guided by those. Similarly the action selection logistic regression shows that there is indeed a significant difference in gazing strategy in 9/10 participants. The one participant that did not show gazing difference in his behavior was noted to RAR in almost all trials, regardless of having some of the larger anthropometrics (e.g. 1.90 height and overall longest measures). This could be an indicator of insecurity in performing the task, meaning he gazed at additional location even on the few DR trials he did. Alternatively, compared to the 9 other participants perhaps he employed a different strategy involving gazes at other visual sites not involving the floor AOI at all.

These results indicate that relocating before grasping an object employs looking at sites additional to the main target. Our findings were not unexpected and are in line with results from Johansson et al., 2001 demonstrating that gaze will be drawn to areas crucial for control and successful task execution e.g. to obstacles. In our study these sites appear to be the table and even more so the floor demonstrating a relationship between fixation patterns and RAR. As a result these AOI are naturally invoked during the more complex relocating trials.

Another finding consistent with previous studies is that our results indicate the subjects were focused on task relevant targets. Despite the lack of familiarity with VR environment and technology, we see that if there was a level of distraction present, it was not overwhelming. This is both visible from the lack of unrelated gazes to e.g. walls, door or ceiling and the fact that they were not disturbed by the lack of an actual hand. Hand gazes were consistently very low regardless the trial. Gaze not diverting from task related AOI and not focusing on the hand motions is a consistent finding in the vast majority of PEG studies (M. F. Land, 2009; Johansson et al., 2001).

In conclusion, our findings demonstrate that PEG behaviors can occur and be observed in VR environments. Moreover, farther trials requiring more mobility tend to include more gaze fixations at AOI instead of focusing only on the main target. This is reinforced by our action selection analysis supporting that RAR actions require more gazes for environment related information compared to DR in order to complete a PAP task .This appears to be a strategy aiming to obtain and evaluate information crucial for motion navigation and control (Anderson et al., 1997; Johansson et al., 2001). Yet, according to the independent t-test, both RAR and DR actions maintained focus to the target object regardless of RAR fixations to other AOI.

Regarding future studies, it would be prolific to see them directed towards gaze identification patterns for action prediction, to develop better models for machine interactions. Accurate gaze predictive models could help humans, especially those in need of support, to better communicate their needs and desires (e.g. assisting robots, Huang et al., 2015). Another direction that would be interesting to explore concerns the impact of tactile experience and eye-hand coordination in VR. For example, during preliminary analysis of the second phase, an interesting gaze pattern emerged. While dropping the box in the bin, some fixations shifted towards the "hand". This could be a behavior solely observed in VR, serving as an indication for the lack of tactile feedback. Sadly in this thesis, we did not explore gaze behaviors during the second phase, involving the object transport to the target location. It would therefore be interesting to see if transport activities generate different or novel findings compared to reaching behaviors.

Lastly, there were several unexpected difficulties that arose during this thesis and ought to be at least briefly addressed. Being an international student and moving for the sole purpose of this study was sufficiently challenging provided the brief timeframe of a thesis. Designing as well as conducting the experiments and reports amidst a pandemic outbreak was definitely an unforeseen turn of events. COVID-19's effects extend to societal, personal, professional and academic consequences that often remain unseen.

One example of how plans for this study changed was another experiment, briefly mentioned in the thesis plan. PEG behaviors can also occur when observing another person's actions. This suggests the activation of a mirror neuron system specialized at execution-observation matching which is thought to be involved in action recognition (Buccino et al., 2004) both while performing and observing an action (Rizzolatti et al., 2001). Therefore, an observer may exhibit gaze fixations similar to those of the actor, resulting in similar PEG behaviors (J. Randall Flanagan & Johansson, 2003; Rotman et al., 2006). Unfortunately this interesting direction in PEG was not addressed within this study as originally planned.

Bibliography

Ambrosini, E., Reddy, V., de Looper, A., Costantini, M., Lopez, B., & Sinigaglia, C. (2013). Looking Ahead: Anticipatory Gaze and Motor Ability in Infancy. *PLoS ONE*, *8*(7). https://doi.org/10.1371/journal.pone.0067916

Anderson, J., Barlow, H. B., Gregory, R. L., Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *352*(1358), 1231–1239. https://doi.org/10.1098/rstb.1997.0105

Ballard, D. H., Hayhoe, M. M., Li, F., & Whitehead, S. D. (1992). Hand-eye coordination during sequential tasks. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *337*(1281), 331–338; discussion 338-339. https://doi.org/10.1098/rstb.1992.0111

Billing, E., Hanson, L., Lamb, M., & Högberg, D. (2019). *Digital Human Modelling in Action*. 25–28. http://urn.kb.se/resolve?urn=urn:nbn:se:his:diva-17882

Billing, E., Sciutti, A., & Sandini, G. (2019). *Proactive eye-gaze in human-robot interaction*. Anticipation and Anticipatory Systems: Humans Meet Artificial Intelligence, Örebro, Sweden, June 10-13, 2019. http://urn.kb.se/resolve?urn=urn:nbn:se:his:diva-17797

Buccino, G., Binkofski, F., & Riggio, L. (2004). The mirror neuron system and action recognition. *Brain and Language*, *89*(2), 370–376. https://doi.org/10.1016/S0093-934X(03)00356-0

Causer, J., Harvey, A., Snelgrove, R., Arsenault, G., & Vickers, J. N. (2014). Quiet eye training improves surgical knot tying more than traditional technical training: A randomized controlled study. *The American Journal of Surgery*, *208*(2), 171–177. https://doi.org/10.1016/j.amjsurg.2013.12.042

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Cushman, L. A., Stein, K., & Duffy, C. J. (2008). Detecting navigational deficits in cognitive aging and Alzheimer disease using virtual reality. *Neurology*, *71*(12), 888–895. https://doi.org/10.1212/01.wnl.0000326262.67613.fe

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(6), 381–391. https://doi.org/10.1037/h0055392

Flanagan, J. R., Bowman, M. C., & Johansson, R. S. (2006). Control strategies in object manipulation tasks. *Current Opinion in Neurobiology*, *16*(6), 650–659. https://doi.org/10.1016/j.conb.2006.10.005

Flanagan, J. R., King, S., Wolpert, D. M., & Johansson, R. S. (2001). Sensorimotor prediction and memory in object manipulation. *Canadian Journal of Experimental Psychology = Revue Canadienne De Psychologie Experimentale*, *55*(2), 87–95. https://doi.org/10.1037/h0087355

Flanagan, J. Randall, & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769–771. https://doi.org/10.1038/nature01861

Gaggioli, A. (2001). Using virtual reality in experimental psychology. In *Towards cyberpsychology: Mind, cognition and society in the internet age* (pp. 157–174). IOS Press.

Gesierich, B., Bruzzo, A., Ottoboni, G., & Finos, L. (2008). Human gaze behaviour during action execution and observation. *Acta Psychologica*, *128*(2), 324–330. https://doi.org/10.1016/j.actpsy.2008.03.006

Harle, S. K., & Vickers, J. N. (2001). Training Quiet Eye Improves Accuracy in the Basketball Free Throw. *The Sport Psychologist*, *15*(3), 289–305. https://doi.org/10.1123/tsp.15.3.289

Huang, C.-M., Andrist, S., Sauppé, A., & Mutlu, B. (2015). Using gaze patterns to predict task intent in collaboration. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01049

Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye–Hand Coordination in Object Manipulation. *The Journal of Neuroscience*, *21*(17), 6917–6932. https://doi.org/10.1523/JNEUROSCI.21-17-06917.2001

Lamb, M., Kallen, R. W., Harrison, S. J., Di Bernardo, M., Minai, A., & Richardson, M. J. (2017). To Pass or Not to Pass: Modeling the Movement and Affordance Dynamics of a Pick and Place Task. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01061

Lamb, M., Nalepka, P., Kallen, R. W., Lorenz, T., Harrison, S. J., Minai, A. A., & Richardson, M. J. (2019). *A Hierarchical Behavioral Dynamic Approach for Naturally Adaptive Human-Agent Pick-and-Place Interactions* [Research Article]. Complexity. https://doi.org/10.1155/2019/5964632

Land, M. F. (2009). Vision, eye movements, and natural behavior. *Visual Neuroscience*, *26*(1), 51–62. https://doi.org/10.1017/S0952523808080899

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328. https://doi.org/10.1068/p2935

Le, A., & Niemeier, M. (2014). Visual field preferences of object analysis for grasping with one hand. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00782

Magora, F., Cohen, S., Shochina, M., & Dayan, E. (2006). Virtual reality immersion method of distraction to control experimental ischemic pain. *The Israel Medical Association Journal: IMAJ*, *8*(4), 261–265.

Perumaal, S. S., & Jawahar, N. (2013). Automated Trajectory Planner of Industrial Robot for Pick-and-Place Task. *International Journal of Advanced Robotic Systems*, *10*(2), 100. https://doi.org/10.5772/53940

Rizzolatti, G., Fogassi, L., Gallese, V., & Rizolatti, G. (2001). *Neurophysiological mechanisms underlying the understanding and imitation of action.*

https://www.scienceopen.com/document?vid=f6c85f94-ea50-4c78-8393-b8d3f4269978

Rotman, G., Troje, N. F., Johansson, R. S., & Flanagan, J. R. (2006). Eye Movements When Observing Predictable and Unpredictable Actions. *Journal of Neurophysiology*, *96*(3), 1358–1369. https://doi.org/10.1152/jn.00227.2006

Sailer, U., Flanagan, J. R., & Johansson, R. S. (2005). Eye-hand coordination during learning of a novel visuomotor task. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *25*(39), 8833–8842. https://doi.org/10.1523/JNEUROSCI.2658-05.2005

Sauzéon, H., Arvind Pala, P., Larrue, F., Wallet, G., Déjos, M., Zheng, X., Guitton, P., & N'Kaoua, B. (2012). The Use of Virtual Reality for Episodic Memory Assessment. *Experimental Psychology*, *59*(2), 99–108. https://doi.org/10.1027/1618-3169/a000131

Srinivasan, D., & Martin, B. J. (2010). Eye-hand coordination of symmetric bimanual reaching tasks: Temporal aspects. *Experimental Brain Research*, *203*(2), 391–405. https://doi.org/10.1007/s00221-010-2241-3

Valevicius, A. M., Boser, Q. A., Lavoie, E. B., Murgatroyd, G. S., Pilarski, P. M., Chapman, C. S., Vette, A. H., & Hebert, J. S. (2018). Characterization of normative hand movements during two functional upper limb tasks. *PLOS ONE*, *13*(6), e0199549. https://doi.org/10.1371/journal.pone.0199549

Vickers, J. N. (1992). Gaze Control in Putting. *Perception*, *21*(1), 117–132. https://doi.org/10.1068/p210117

Vickers, J. N. (2007a). *Chapter 5—Gaze Control To a Fixed Target*. Quiet Eye Solutions. http://www.quieteyesolutions.com/content/Vickers_2007/Chapter5.htm

Vickers, J. N. (2007b). *Perception, cognition, and decision training: The quiet eye in action* (pp. xii, 267). Human Kinetics.

Wilson, M. R., McGrath, J. S., Vine, S. J., Brewer, J., Defriend, D., & Masters, R. S. W.

(2011). Perceptual impairment and psychomotor control in virtual laparoscopic

surgery. *Surgical Endoscopy, 25*(7), 2268–2274. https://doi.org/10.1007/s00464-

010-1546-4

Appendix

1.

Post Questionnaire, verbally asked and filled by the researcher

Questionnaire form filled by the researcher

Participant # _____

Date_____

Open questions

How was the experience? Any thoughts or comments about the procedure?

Demographics & General daily life information

Age_____

Gender (Do you identify as male, female, non binary, other or no answer) _____

Occupation/s (if student, what field) /or what is their daily occupation? :

Hours per week							
0	1-5	5-10	10-20	20-30	30-40	40-50	50+

Hobbies/ how do they spend their free time:

Video games (VG) & VR

Hours per week							
0	1-5	5-10	10-20	20-30	30-40	40-50	50+

Do you play **VG**? **Any** type, any platform.

 How many hours**/w** do you play games in total, from the following options?

Game titles you play the most in the last 12 months:

Do you ever use **VR** in your life?

How many hours/w from the following options?

What do you use it for?

Physical measurements

Height_____

Right Arm length

- Shoulder to middle finger tip_____
- Shoulder to elbow_____
- Elbow to wrist_____

Leg length

- Hip to floor_____
- Hip to knee_____