REALISM MATTERS. EYE MOVEMENTS IN VR REVEAL PRIOR TOOL KNOWLEDGE, INTENDED INTERACTION AND END-STATE COMFORT PLANNING

A PREPRINT

Ashima Keshava
University of Osnabrück
Germany
akeshava@uos.de

Nina Gottschewsky
University of Osnabrück
Germany

Stefan Balle
University of Osnabrück
Germany

Farbod Nosrat Nezami
University of Osnabrück
Germany

Thomas Schüler
Halocline GmbH & Co. KG
Germany

Peter König
University of Osnabrück
University Medical Center,
Hamburg-Eppendorf
Germany

July 27, 2021

ABSTRACT

Seminal studies on human cognitive behavior have been conducted in controlled laboratory settings, demonstrating that visual attention is mainly goal-directed and allocated based on the action performed. However, it is unclear how far these results generalize to cognition in more naturalistic settings. The present study investigates active inference processes revealed by eye movements during interaction with familiar and novel tools with two levels of realism of the performed action. We presented participants with 3D tool models that were either familiar or unfamiliar, oriented congruent or incongruent to their handedness, and asked participants to interact with them by either lifting or using. Importantly, we used the same experimental design in two setups. In the first experiment, participants interacted with a VR controller; in the second, they performed the task with a more realistic interaction setup that allowed differentiated hand and finger movements. We used linear mixed models to determine the odds of fixations on the tool effector vs. handle before action initiation. The results show that participants fixate more on the tool's effector part before action initiation during the use task for unfamiliar tools. Furthermore, under more natural interaction conditions, subjects fixate more on the tool's handle as a function of the tool's orientation, well before the action was executed. Secondly, the spatial viewing bias on the tool reveals early fixations are associated with the task and familiarity of the tools. In contrast, later fixations are associated with the manual planning of the interaction. In sum, the findings from the experiments suggest that fixations are made in a task-oriented way to extract the mechanical properties of the tools to plan the intended action well before action initiation. Further, in more natural conditions, fixations are made towards proximal goals of optimally planning the grasp even when the perceived action on the tools is identical. Thus, the realism of the action in a virtual environment seems to be as important as the sensory stimulation.

Keywords anticipatory behavior, action-oriented attention, active inference, eye-tracking, virtual reality, natural cognition

1 Introduction

A longstanding goal of the cognitive sciences is to understand cognition, behavior, and experience as it unfolds in the natural world (Parada and Rossi, 2020). Given the technological advancements made in the last decade, there are few

methodological roadblocks to understanding natural cognition where laboratory studies can be extended to naturalistic settings and hopefully lead towards new insights (Ladouce et al., 2016; Parada, 2018). More recently, a pragmatic turn has emerged in the field where there is a greater push towards incorporating the body and bodily actions to infer cognitive function (Engel et al., 2013).

Human tool use is an explicitly natural cognitive function that involves the transfer of proximal goals (e.g., placement of grasp) to distal goals for the tool (Arbib et al., 2009). Moreover, simple tools fundamentally expand the body representations to include representations of the tool in the peripersonal space (Berti and Frassinetti, 2000; Farnè et al., 2005; Maravita et al., 2002). Furthermore, tool use is differentiated from other object-based actions where the tool is "acted with" (Johnson and Grafton, 2003) and requires semantic knowledge of the tool as well as the necessary skill to perform actions with it (Johnson-Frey, 2004). Hence, tool use involves complex behaviors ranging from cognitive and semantic reasoning to perceptual and motor processes.

When using tools, a wealth of information is parsed to produce the relevant action. The semantic knowledge associated with the tool helps the understanding of how it is used, the mechanical knowledge maps the physical properties of the tool for potential usage, and finally, sensorimotor knowledge helps decipher possible movements required to use the tool (Baumard et al., 2014). The amalgamation of these knowledge sources (which can be unique to a tool) is a necessitating factor to plan any action associated with the tool. When this knowledge is not readily available, inferential processes must be deployed to deduce the relevant action.

In naturalistic settings, studies have shown that eye movements are made to locations in the scene in anticipation of the next action (Hayhoe, 2004; Land and Hayhoe, 2001; Pelz and Canosa, 2001). Moreover, Belardinelli et al. (2016b) showed that eye movements are goal-oriented and are modulated in anticipation of the object interaction task. Similarly, there is strong evidence that task plays an important role in how the eyes scan the scene and are differentiated between passive viewing and pantomimed interaction (Belardinelli et al., 2015). Similarly, Keshava et al. (2020) showed that rudimentary object interactions can be decoded by using eye-movement data alone. Even in the absence of an interaction, task relevance plays an important role (Castelhano et al., 2009; Henderson and Hayes, 2017). These studies point towards gaze control being the consequence of knowledge and task-driven predictions (Henderson, 2017).

Moreover, Belardinelli et al. (2016a) investigated the role of anticipatory eye movements when interacting with familiar and unfamiliar tools in a classic lab setting. These tools had clearly differentiable parts: tool handle and effector. The results showed that in the case of unfamiliar tools, preparatory eye movements were made to the tool-effector to extract the mechanical properties of the tool as the semantic information was not readily available. This effect was enhanced when subjects were asked to perform tool-specific movements instead of a generic movement of lifting the tool by the handle. The authors, hence, concluded that eye movements are used to actively infer the appropriate usage of the tools from their mechanical properties. In the study, the tools were presented to the participants as images on a screen and they were asked to pantomime their actions of either lifting or using the tool. While the study revealed valuable insights into anticipatory gaze control, a question remains if these results are part of natural cognition where they can be reproduced in more realistic environments.

When studying anticipatory behaviors corresponding to an action, one must also ask, is symbolized action enough, and how real should the action actually be? Króliczak et al. (2007) showed brain areas typically involved in real actions are not driven by pantomimed actions, and that pantomimed grasps do not activate the object-related areas within the ventral stream. Similarly, Hermsdörfer et al. (2012) showed a weak correlation between the hand trajectories for pantomimed and actual tool interaction. These studies indicate that the realism of sensory and tactile feedback while acting (e.g., grasping) can be an essential factor when studying anticipatory behavioral control.

In virtual reality (VR), realistic actions can be studied by simulating an interaction within an environment. In these simulations, using interfaces such as VR controllers, ego-centric visual feedback of a body part, most often a hand and sometimes the whole body (Keshava et al., 2021) can be designed. These interfaces usually consist of a hand-held device that can be controlled the index finger or the thumb alone. In a typical VR setup, different interactions such as grasping, lifting, moving an object can be possible by simply pressing a button with the index finger. Moreover, while users only press a button, the virtual visual feedback they get is of a simulated hand that performs the action. Conversely, more naturalistic interfaces such as LeapMotion, capture the real-time movements of the user's hand as they interact in the environment. These interfaces give the user realistic visual feedback of their finer hand and finger movements. Consequently, the chosen method of interaction in VR can afford different levels of realism and could elicit different behavioral responses.

In the present study, we investigated anticipatory gaze control in two different experiments where we were interested in the extent to which the realism of the interaction and the environment modified the results shown by Belardinelli et al. (2016a). We asked participants to lift or use 3D models of tools in VR that we developed from the original study. These

tools were categorized as familiar or unfamiliar. Additionally, we extended the experimental design to include the tool handle's spatial orientation, either congruent or incongruent to the subjects' handedness.

In experiment-I, subjects performed the experiment in a less natural/realistic environment and interacted with the tool models using a VR controller, which mimicked grasp in the virtual environment. In experiment-II, subjects were immersed in a more realistic environment where they interacted with the tools using LeapMotion, which required natural hand and finger movements, and thus, made the interactions closer to the real world. With this experimental design, we investigate the influence of task, tool familiarity, the spatial orientation of the tool, and, notably, the impact of the realism of the performed movements.

2 Methods

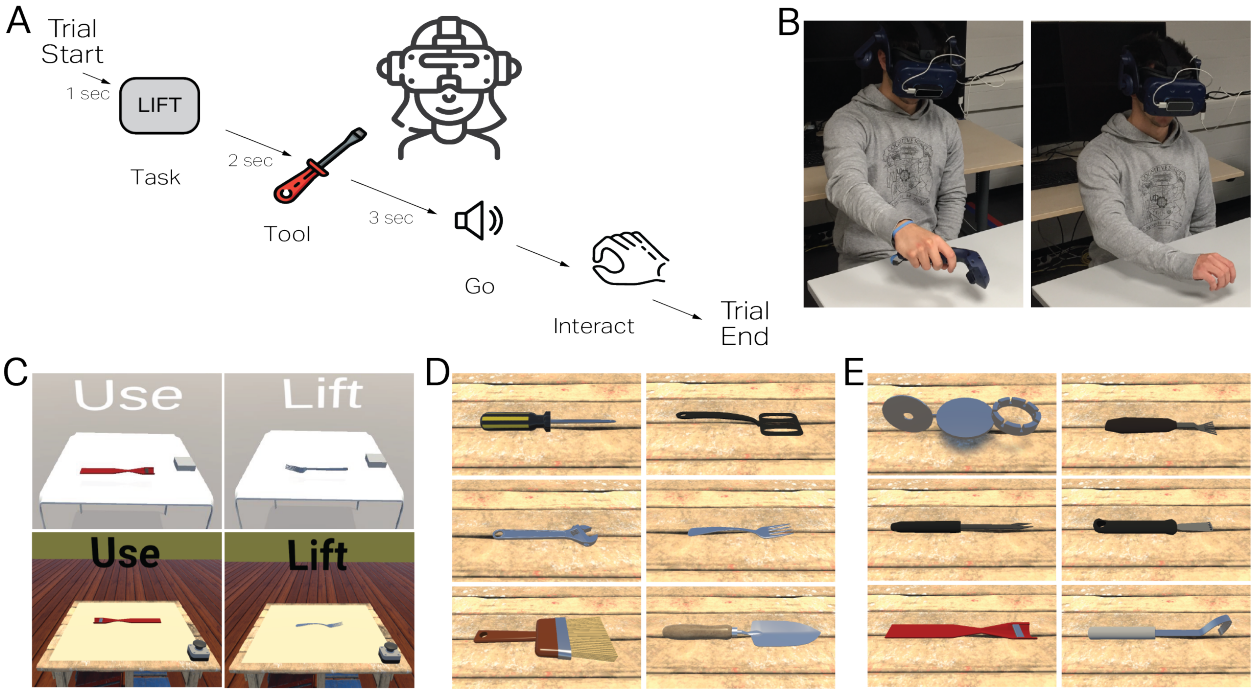

Figure 1: Experimental Task. In two virtual environments (VEs) participants interacted with tools in two ways (LIFT, USE). The tools were categorized based on familiarity (FAMILIAR, UNFAMILIAR) and were presented to the participants in two different orientations (HANDLE LEFT, HANDLE RIGHT). The two VEs differed based on the mode of interaction and perceived realism where in one experiment, subjects' hand movements were rendered virtually using the HTC-VIVE controllers and in the other the hands were rendered using LeapMotion allowing for more natural view of the hands. Panel **A** shows the timeline of a trial. Panel **B** shows a subject in real life performing the task in the two experiments. Panel **C** shows the differences in realism in the two experiments, TOP panels correspond to experiment with the controllers, the USE and LIFT conditions for an UNFAMILIAR and FAMILIAR tool respectively. The tools are presented in two different orientations. BOTTOM panels illustrate the three different conditions in the more realistic experiment with LeapMotion as the interaction method. Panel **D** Familiar tools, from top-left: screwdriver, spatuala, wrench, fork, paintbrush, trowel. Panel **E** Unfamiliar tools, from top-left: spoke-wrench, palette knife, daisy grubber, lemon zester, flower cutter, fish scaler.

2.1 Participants

For experiment-I with the HTC Vive controller's interaction method, we recruited 18 participants (14 females, mean age = 23.68, SD = 4.05 years). For experiment-II with the interaction method of LeapMotion, we recruited 30 participants (14 female, mean age=22.7, SD = 2.42 years). All participants were recruited from the University of Osnabrück and the University of Applied Sciences Osnabrück. Participants had a normal or corrected-to-normal vision and no history of neurological or psychological impairments. All of the participants were right-handed. They either received a monetary reward of C10 or one participation credit per hour. Before each experimental session, subjects gave their informed consent in writing. They also filled out a questionnaire regarding their medical history to ascertain they did not suffer from any disorder/impairments which could affect them in the virtual environment. Once we obtained their informed consent, we briefed them on the experimental setup and task.

2.2 Experimental Task

Subjects were seated in a virtual environment where they had to interact with the presented tool by either lifting or pretending its use. At the start of a trial, subjects saw the cued task for 2 sec after which the cue disappeared, and a tool appeared on the virtual table. Subjects were given 3 sec to view the tool, after which there was a beep (go cue) which indicated that they could start manipulating the tool based on the cued task. After interacting with the tool, subjects pressed a button on the table to start the next trial. The time course of the trials is illustrated in Figure 1A.

2.3 Experimental Design and Procedure

The two experiments differed based on the realism of the interaction method and the environment. In experiment-I, subjects interacted with the tool models using the HTC Vive VR controllers. While in experiment II, subjects' hand movements were captured by LeapMotion. Figure 1B illustrates the physical setup of the participants for the two experiments.

For both experiments, we used a 2x2x2 experiment design, with factors task, tool familiarity, and handle orientation. Factor task had two levels: LIFT and USE. In the LIFT conditions, we instructed subjects to lift the tool to their eye level and place it back on the table. In the USE task, they had to pantomime using the tool to the best of their knowledge. Factor familiarity had two levels FAMILIAR and UNFAMILIAR, which corresponded to tools either being everyday familiar tools or tools that are not seen in everyday contexts and are unfamiliar. The factor handle orientation corresponded to the tool handle which was presented to the participants either on the LEFT or the RIGHT. Both experiments had 144 trials per participant with an equal number of trials corresponding to the 3 factors. Subjects performed the trials over six blocks of 24 trials each. Figure 1C illustrates two exemplar trials from the experiments.

We measured the eye movements and hand movements simultaneously while subjects performed the experiment. We calibrated the eye-trackers at the beginning of each block and ensured that the calibration error was less than 1 degree of the visual angle. At the beginning of the experiment, subjects performed three practice trials with a hammer to familiarize themselves with the experimental setup and the interaction method. Each experimental session lasted for approximately an hour. After that, subjects filled out a questionnaire to indicate their familiarity with the 12 tools used in the experiment. They responded to the questionnaire based on a scale of 5-point Likert-like scale where 1 corresponded to "I have never used it or heard about it," and 5 referred to "I see it every day or every week."

2.4 Experimental Stimuli

In both experiments, the experimental setups consisted of a virtual table that mimicked the table in the real world. The height, width, and length of the table were 86cm, 80cm, 80cm respectively. In experiment-I, subjects were present in a bare room with grey walls and constant illumination. They sat before a light grey table, with a dark grey button on their right side to indicate the end of the trial. Similarly, In experiment-II, subjects were present in a more immersive, realistic room. They sat in front of a wooden workbench that had the exact dimensions as the real-world table. We displayed the task (USE or LIFT) over the desk 2m away from the participants for both experiments. For both experiments, we used the tool models as presented in Belardinelli et al. (2016a). Six of the tools were categorized as familiar (Figure 1D) and the other six as unfamiliar (Figure 1E). We further created bounding box colliders that encapsulated the tools to capture the gaze position on the tool models. The mean length of the bounding box was 34.04cm (SD=5.73), mean breadth=7.60cm (SD=3.68) and mean height= 4.17cm (SD=2.13). To determine the tool effector and tool handle regions of interest, we halved the length bounding box colliders from the center of the tool and took one half as the effector and the other half as the handle. This way we refrained from making arbitrary-sized regions-of-interest for the different tool models.

2.5 Apparatus

For both experiments, we used an HTC Vive head-mounted display (HMD)(110° field of view, 90Hz, resolution 1080 x 1200 px per eye) with a built-in Tobii eye-tracker [1]. The HTC Vive Lighthouse tracking system provided positional and rotational tracking and was calibrated for a 4m x 4m space. For calibration of the gaze parameters, we used 5-point calibration function provided by the manufacturer. To make sure the calibration error was less than 1°, we performed a 5-point validation after each calibration. Due to the nature of the experiments, which allowed a lot of natural head movements, the eye tracker was calibrated repeatedly during the experiment after each block of 36 trials. We designed

[1] https://enterprise.vive.com/us/product/vive-pro-eye-office/

the experiment using the Unity3D game engine [2] (v2019.2.14f1) and controlled the eye-tracking data recording using HTC VIVE Eye Tracking SDK SRanipal[3] (v1.1.0.1)

For experiment-I, we used HTC Vive controller[4] (version 2.5) to interact with the tool. The controller in the virtual environment was rendered as a gloved hand. When participants pulled the trigger button of the controller with their right index finger, their right virtual hand made a power grasp action. To interact with the tools, subjects pulled the trigger button of the controller over the virtual tools and the rendered hand grasped the tool handle.

Similarly, in experiment-II, we used LeapMotion[5] (version 4.4.0) to render the hand in the virtual environment. Here, subjects could see the finer hand and finger movements of their real-world movements rendered in the virtual environment. When participants made a grasping action with their hand over the virtual tool handle, the rendered hand grasped the tool handle in the virtual environment.

2.6 Data pre-processing

2.6.1 Gaze Data

As a first step, using eye-in-head 3d gaze direction vector for the cyclopean eye we calculated the gaze angles in degrees for the horizontal θ_{h} and vertical θ_{v} directions. All of the gaze data was sorted by the timestamps of the collected gaze samples. The 3d gaze normals are represented as (x, y, z) a unit vector that defines the direction of the gaze in VR world space coordinates. In our setup, the x coordinate corresponds to the left-right direction, y in the up-down direction, z in the forward-backward direction. The formulas used for computing the gaze angles are as follows:

$$\theta_{\mathrm{h}} = \frac{180}{\pi} \arctan \frac{x}{z}$$

$$\theta_{\mathrm{v}} = \frac{180}{\pi} \arctan \frac{y}{z}$$

Next, we calculated the angular velocity of the eye in both the horizontal and vertical coordinates by taking a first difference of the angular velocity and dividing by the difference between the timestamp of the samples using the formula below:

$$\omega_{\mathrm{h}} = \Delta\theta_{\mathrm{h}}/\Delta t$$

$$\omega_{\mathrm{v}} = \Delta\theta_{\mathrm{v}}/\Delta t$$

Finally, we calculated the magnitude of the angular velocity (ω) at every timestamp from the horizontal and vertical components using:

$$\omega = \sqrt{\omega_h^2 + \omega_v^2}$$

To classify the fixation and saccade-based samples, we used an adaptive threshold method for saccade detection described by Voloh et al. (2020). We selected an initial saccade velocity threshold θ_0 of 200 ∘/sec. All eye movement samples with an angular velocity of less than θ_0 were used to compute a new threshold θ_1. θ_1 was three times the median absolute deviation of the selected samples. If the difference between θ_1 and θ_0 was less than 1 ∘/sec θ_1 was selected as the saccade threshold else, θ_1 was used as the new saccade threshold and the above process was repeated. This was done until the difference between θ_{n} and $\theta_{\mathrm{n+1}}$ was less than or equal to 1 ∘/sec. This way we arrived at the cluster of samples that belonged to fixations and the rest were classified as saccades.

After this, we calculated the duration of the fixations and removed those fixations that had a duration less than 50ms or were larger than 3.5 times the median absolute deviation of the fixation duration. For further data analysis, we only considered those fixations that were positioned on the 3D tool models. We further categorized the fixations based on their position on the tool, i.e., whether they were located on the effector or handle of the tool.

[2] Unity, www.unity.com

[3] SRanipal, developer.vive.com/resources/vive-sense/sdk/vive-eye-tracking-sdk-sranipal/

[4] SteamVR, https://valvesoftware.github.io/steamvr_unity_plugin/articles/Quickstart.html

[5] LeapMotion Unity modules, https://developer.leapmotion.com/unity

2.7 Data Analysis

2.7.1 Odds of Fixations in favor of tool effector

After cleaning the dataset, we were left with 2174 trials from 18 subjects in experiment-I and 3633 trials from 30 subjects in experiment-II. For the two experiments, we modeled the linear relationship of the log of odds of fixations on the effector of the tools and the task cue (LIFT, USE), the familiarity of the tool (FAMILIAR, UNFAMILIAR), and orientation of the handle (LEFT, RIGHT). All within-subject effects were also modeled with random intercepts and slopes based on the subjects. We were also interested in modeling the random effects based on the tool, to assess the differential effects on the individual tools. As we did not have enough data to estimate both random intercept and slope of the stimuli, we fitted a random intercept for each of the 12 tools.

We used effect coding (Schad et al., 2018) to construct the design matrix for the linear model, where we coded the categorical variables LIFT, FAMILIAR, RIGHT to -0.5 and USE, UNFAMILIAR, LEFT to 0.5. This way, we could directly interpret the regression coefficients as main effects. The model fit was performed using restricted maximum likelihood (REML) estimation (Corbeil and Searle, 1976) using the lme4 package (v1.1-26) in R 3.6.1. We used the L-BFGS-B optimizer to find the best fit using 10000 iterations. Using the Satterthwaite method (Luke, 2017), we approximated the fixed effects degrees of freedom. For both experiments, the following model was used:

$$log\frac{p(fixations \quad on \quad effector)}{p(fixations \quad on \quad handle)} \sim 1 + task * familiarity * handle_orientation$$
$$+(1 + task * familiarity * handle_orientation|Subject)$$
$$+(1|tool)$$

As we used effects coding, we can directly compare the regression coefficients of the two models. The fixed-effect regression coefficients of the two models would describe the differences in log-odds of fixations in favor of tool effector for the categorical variables task, familiarity, and handle orientation.

2.7.2 Spatial bias of fixations on the tools

In this analysis, we wanted to assess the effects of task, tool familiarity, and handle orientation on the eccentricity of fixations on the tools. To do this, we studied the fixations from the time when the tool was visible on the table (3s from the start of trial) till the go cue indicated when subjects could start manipulating the tool. We divided this 3s period into 20 equal bins of 150ms each. For each trial and time bin, we calculated the median distance of the fixations from the tool center. Next, we normalized the distance with the length of the tool so that we could compare the fixation eccentricity across different tools.

To find the time-points where there were significant differences for the 3 conditions and their interactions, we used the cluster permutation method. Here, we use the t-statistic as a test statistic for each time-bin, where t is defined as:

$$t = \frac{x}{\frac{\sigma}{\sqrt{N}}}$$

and, x is the mean difference between conditions, and σ is the standard deviation of the mean and N is the number of subjects. We used a threshold for t at 2.14 which corresponds to the t-value at which the p-value is 0.05 in the t-distribution. We first found the time-bins where the t-value was greater than the threshold. Then, we computed the sum of the t-values for these clustered time-bins which gave a single value that represented the mass of the cluster. Next, to assess the significance of the cluster, we permuted all the time-bins across trials and subjects and computed the t-values and cluster mass for 1000 different permutations. This gave us the null distribution over which we compared the cluster mass shown by the real data. We considered the significant clusters to have a p-value less than 0.05. In the results, we report the range of the significant time-bins for the 3 different conditions and their interactions and the corresponding p-values.

3 Results

The present study investigated the differences in gaze-based strategies dependent on task, tool familiarity, and handle orientation. Here, we investigated two anticipatory gaze-based strategies before action initiation; the odds of fixations

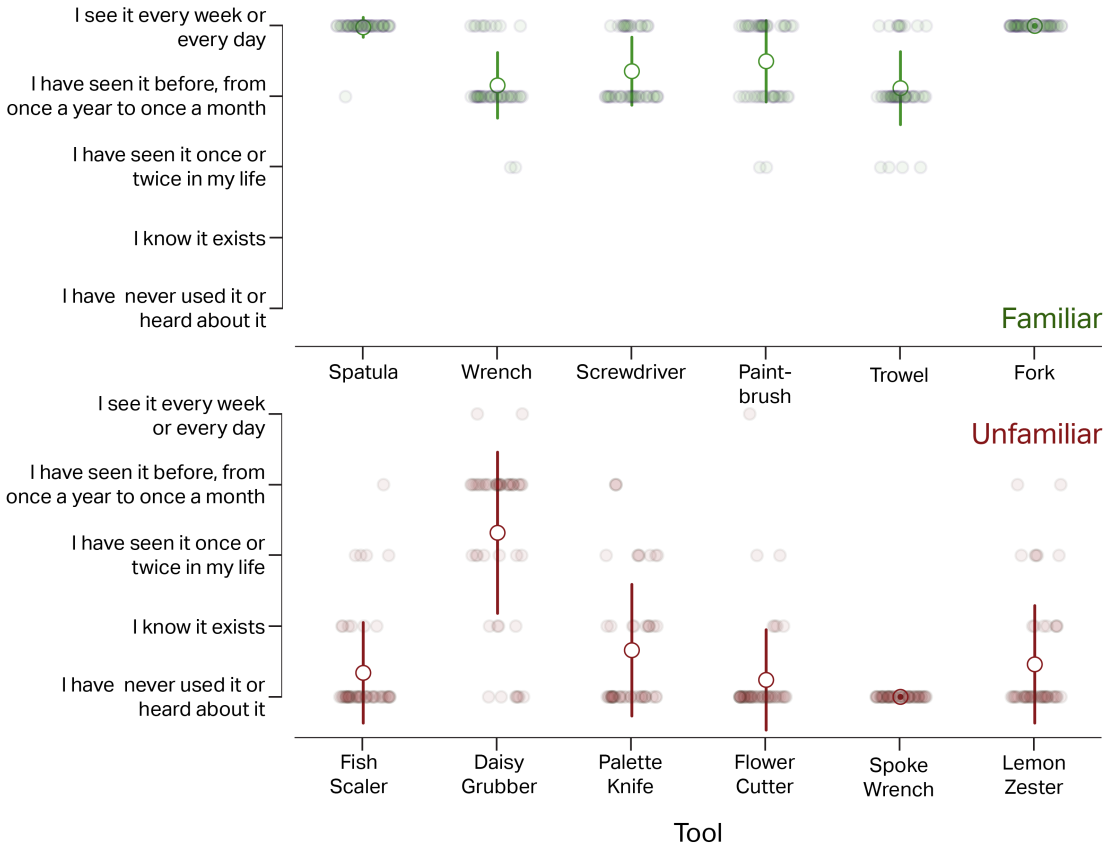

Figure 2: Participants' familiarity rating of the tools. Participants provided their subjective rating of familiarity with the 12 tool stimuli on a 5-point Likert scale. The small circles correspond to ratings from individual subjects. The larger circles correspond to the mean rating for each tool and error bars represent the standard deviation.

in favor of the tool effector and the eccentricity of the fixations through time towards the tool effector or handle. We further compared the differences in two experiments that had the same experimental design but differed in the realism of the interaction and environment.

First, we were interested in how the subjects assessed the subjective familiarity of the 12 tools. The mean familiarity rating for familiar tools was 4.51 (SD=0.55) and for unfamiliar tools 1.65 (SD=1.10). Figure 2 shows the familiarity ratings for all the familiar and unfamiliar tools used in the study. To test the significance of the differences in the familiarity ratings, we performed a pairwise t-test. The results show that the subject ratings for the familiar and unfamiliar tools differed significantly with t=38.69 and p-value<0.001. In sum, our experimental condition of familiarity was consistent with the participants' subjective rating as well.

Next, we compared the log-odds of fixations in favor of the tool effector across the three conditions: task, tool familiarity, and handle orientation. Figure 3A shows the log-odds of the fixations on the effector for experiment-I (with HTC VIVE Controllers) and experiment-II (with LeapMotion).

In experiment-I (Figure 3A, left panel), subjects showed a mean log odds of 0.01 (95%CI = [-0.04, 0.08]) for the LIFT task and for the USE task the mean log-odds were 0.19 (95%CI = [0.11, 0.28]). For the FAMILIAR tools, the mean log-odds in favor of the tool effector were -0.16 (95%CI = [-0.23, -0.09]) and for UNFAMILIAR 0.35 (95%CI = [0.25, 0.45]). For the RIGHT oriented tool handle, the mean log-odds were 0.14 (95%CI = [0.06, 0.21]) and for the LEFT oriented tool handle, the mean log-odds were 0.08 (95%CI = [-0.09, 0.26]). To assess the significance of the factors, we used linear mixed models. For the linear model, we used effect coding so the regression coefficients can be directly interpreted as main effects. There was a significant main effect of factor task (USE - LIFT) β = 0.18 (95%CI = [0.08, 0.27], t(70.09)=3.7), with a p-value < 0.001. There was a significant main effect of familiarity (UNFAMILIAR - FAMILIAR) β = 0.58 (95%CI = [0.09, 1.08], t(10.79)=2.33), with p-value = 0.04. The main effect of handle orientation was not significant (LEFT - RIGHT) β = -0.04, (95%CI = [-0.29, 0.21], t(16.94)=-0.32), p-value = 0.75. We found a significant interaction of task and familiarity with β = 0.24 (95%CI = [0.03, 0.45], t(25.88)=2.21), p-value = 0.036. The interaction of task and handle orientation was not significant, β = 0.19 (95%CI = [-0.05, 0.43], t(21.26)=1.55),

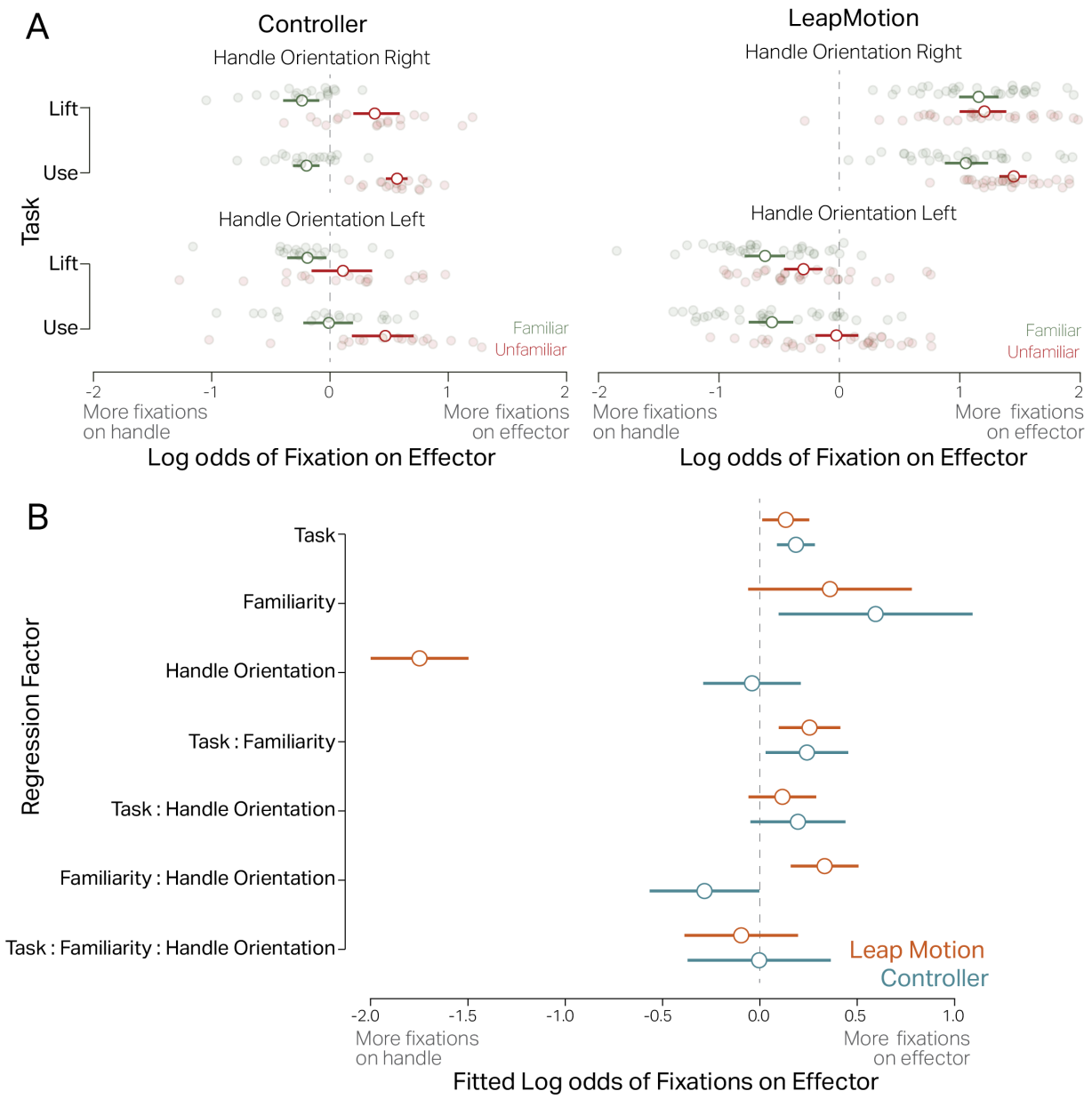

Figure 3: Experiment results. A) top-left: shows the log-odds of fixation on effector in the controller study when the tool handle is oriented to the right. The log odds on fixations are higher for unfamiliar tools (red) than the familiar tools (green) for both the LIFT and the USE tasks. Bottom-left: log odds of fixation on effector when the tool handle is oriented to the left and is incongruent to the subjects' handedness. The plot shows that the orientation of the tool does not affect the log-odds fixation on the effector. Top-right: the log-odds of fixation on effector in the LeapMotion study when the tool handle is oriented to the right. The log odds of fixations on the effector are higher for unfamiliar tools (red) than the familiar tools (green) and the USE task. Bottom-right: log odds of fixation on effector when the tool handle is oriented to the left and is incongruent to the subjects' handedness. The plot shows that the orientation of the tool results in significant log-odds of fixations over the handle. B) The linear regression coefficients for the two experiments. The effect of the task is significant for both experiments with higher log-odds of fixations on the effector. For the factor orientation, the log-odds are significant in the LeapMotion experiment and not for the controller experiment. Similarly, the interaction between task and familiarity is significant for both experiments.

p-value = 0.13. The interaction of familiarity and orientation was not significant , β = -0.28 (95%CI = [-0.56, -0.004], t(17.07)=-1.99), p-value = 0.06. The 3-way interaction was also not significant, β = -0.005 (95%CI = [-0.37, 0.36], t(1695)=-0.02), p-value=0.97.

In experiment-II (Figure 3A, right panel), subjects showed a mean log odds of 0.08 (95%CI = [-0.04, 0.22]) of fixations on the tool effector for the LIFT task and for the USE task the mean log-odds were 0.22 (95%CI = [0.09, 0.36]). For the FAMILIAR tools, the mean log-odds in favor of the tool effector were 0.04 (95%CI = [-0.08, 0.16]) and for UNFAMILIAR 0.25 (95%CI = [0.12, 0.39]). For the RIGHT oriented tool handle, the mean log-odds were 1.18 (95%CI

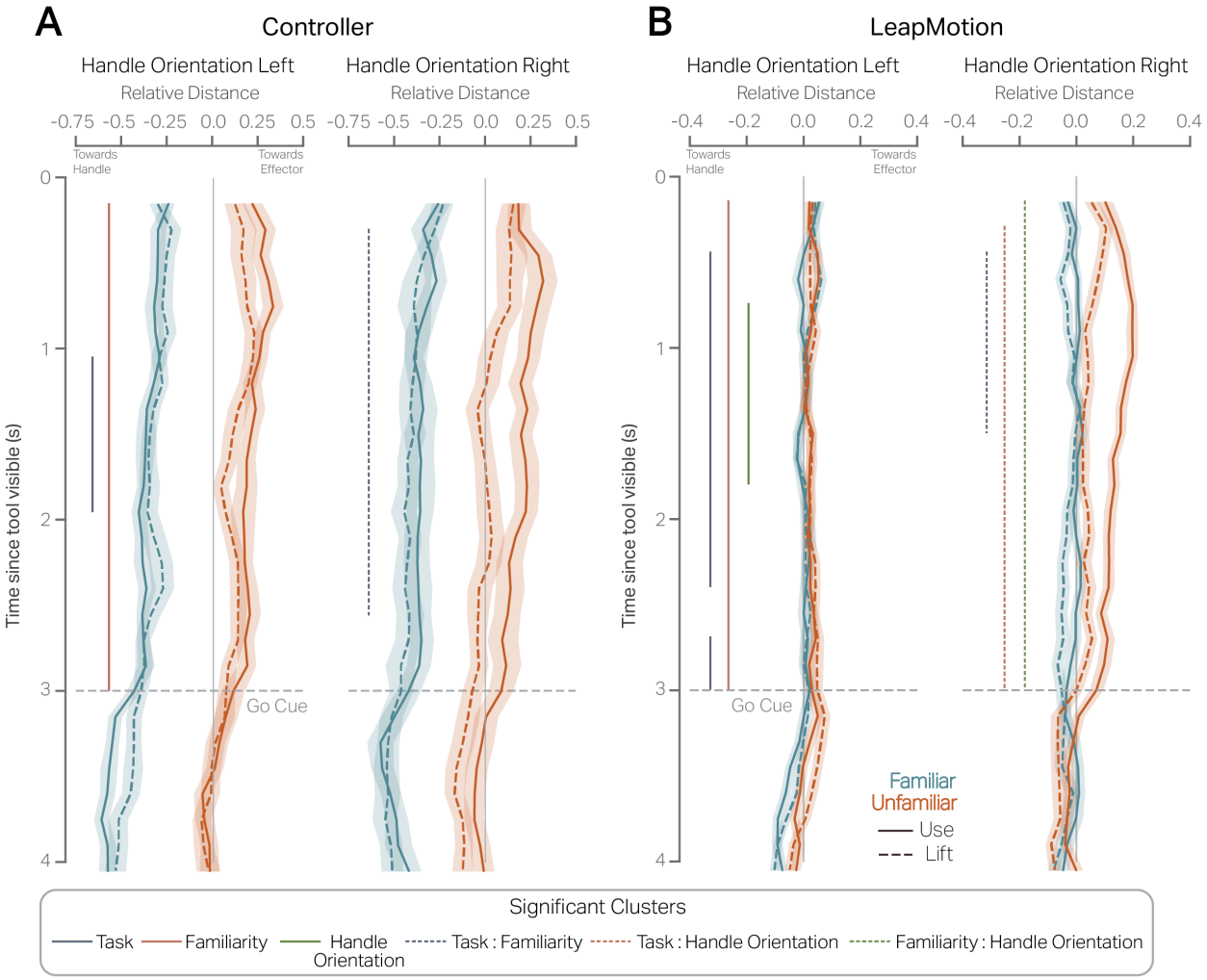

Figure 4: Eccentricity of fixations on the tool models. The negative values of the abscissa correspond to fixations towards the handle, the positive values refer to fixations towards the tool effector, and zero represents the center of the tool. The ordinate axis refers to the time elapsed since the tool is visible on the virtual table. The go cue is given to participants at 3s after which they can start interacting with the tool. The blue lines correspond to the FAMILIAR tool and red to the UNFAMILIAR tools. The error bars represent the standard error of the mean across subjects. The vertical solid lines correspond to the significant time clusters for main effects and the vertical dashed lines to the interactions. Panel A shows the findings from experiment-I and the two handle orientations. Panel B shows the findings from experiment-II and the two handle orientations.

= [1.06, 1.29]) and for the LEFT oriented tool handle, the mean log-odds were -0.32 (95%CI = [-0.45, -0.20]). Using the linear mixed model, we assessed the significance of the three factors. There was a significant main effect of factor task (USE - LIFT) β = 0.13 (95%CI = [0.01, 0.25], t(27.28)=2.13) with a p-value = 0.04. The main effect of familiarity (UNFAMILIAR - FAMILIAR) was not significant β = 0.35 (95%CI = [-0.06, 0.77], t(10.02)=1.67) with p-value = 0.12. The main effect of handle orientation (LEFT - RIGHT) was significant, β = -1.74 (95%CI = [-1.99, -1.48], t(28)=-13.65), p-value <0.001. We found a significant interaction of task and familiarity with β = 0.25 (95%CI = [0.09, 0.41], t(43.64)=3.13), p-value = 0.003. The interaction of task and handle orientation was not significant, β = 0.11 (95%CI = [-0.06, 0.28], t(31.90)=1.29), p-value = 0.20. Similarly, the interaction of familiarity and orientation was significant, β = 0.33 (95%CI = [0.15, 0.50]), p-value = 0.001. The 3-way interaction was also not significant, β = -0.09 (95%CI = [-0.39, 0.19], t(2201)=-0.65), p-value=0.51.

Figure 3B summarizes the regression coefficients of the linear model from both experiments pictorially. Importantly, we see that the main effect of the task is significant for both experiments. Similarly, the interaction of task and familiarity is significant for both experiments. However, the effect of handle orientation is only significant in experiment-II with the LeapMotion interaction method.

Next, we were interested in the effect of task, tool familiarity, and handle orientation on the eccentricity of the fixations on the tool before action initiation. We calculated the relative distance of fixations from the center of the tool in the

3s period when the subjects studied the tool. We used cluster permutation tests to evaluate the time clusters when the effects of the different conditions were significant. As shown in Figure 4A, in experiment-I, the differences in task (USE - LIFT) were significant from 1.05s to 1.95s period, p-value<0.001. Differences in tool familiarity (FAMILIAR - UNFAMILIAR) were significant from 0.15s to 3s with a p-value <0.001. Moreover, the differences in the two orientations (LEFT - RIGHT) were not significant. The interaction of task and familiarity were significant from 0.3s to 2.55s, p-value=0.006. The interactions of task and handle orientation, and the interaction of handle orientation and tool familiarity were not significant in any time period.

Similarly, figure 4B shows the eccentricity of fixations from the center of the tool during the 3s period when the subjects studied the tool in experiment-II. The differences in task were significant in two time periods 0.45s to 2.4s, p-value< 0.001 and from 2.7s to 3s, p-value=0.03. The differences in familiarity were significant from 0.15s to 3s, p-value < 0.001. Furthermore, the differences in handle orientation were significant from 0.75s to 1.8s, p-value < 0.001. A significant interaction of task and familiarity from 0.45s to 1.5s with p-value <0.001. There were also significant clusters in the interaction of task and handle orientation, from 0.3s to 3s with p-value<0.001 and for tool familiarity and handle orientation from 0.15s to 3s with p-value < 0.001.

4 Discussion

4.1 Replication and the ecological validity of the results

The primary aim of this study is to investigate how gaze-based strategies vary for tasks, tool familiarity, and manual planning in naturalistic settings. We replicated and extended the study introduced by Belardinelli et al. (2016a) in two separate experiments. The original study investigated behavioral responses to tool affordances in a lab setting where subjects responded to stimuli images on a computer screen and pantomimed their manual actions. With our study, we successfully replicated the main findings of the original study and added to the ecological validity of the results in two important ways.

In experiment-I, we extended the original study to VR where subjects interacted with 3D tool models in a naturalistic environment. They grasped the tool models using a VR controller and performed generic and tool-specific movements with familiar and unfamiliar tools. We also manipulated the orientation of the tool handles to be congruent or incongruent to the subject's handedness. Here, because of the use of VR controllers, subjects mimicked a grasp by actually only pulling their index fingers. Our main findings suggest that the odds of sampling visual information about the mechanical properties of a tool are different based on the specificity of the task. Moreover, given tool familiarity, the odds of fixating on the effector increased for unfamiliar tools. Tool-specific knowledge also played a major role when subjects were instructed to produce tool-specific movements. Interestingly, the spatial orientation of the tool did not affect the odds of fixations in favor of the tool effector. These results indicate that although the task and tool familiarity were important factors to infer and produce the relevant action, precise manual planning was not required while using the VR controllers.

Here, we reproduced the results from the Belardinelli et al. study as well as from experiment-I. In experiment-II, we investigated oculomotor behavior within the same experimental paradigm but changed the realism of the experimental setup. In this case, subjects had to produce actual generic and tool-specific hand movements to interact with the 3D tool models. Additionally, we found a large effect of spatial orientation of the tool where the odds of fixations in favor of the tool effector decreased significantly when the tool handle was presented incongruent to the subjects' handedness. These results suggest that well before action initiation, subjects had to plan substantially the placement of their hand on the tool to interact with it. This would also be indicative of an end-state comfort planning where participants planned the grasp in anticipation of the final movement well in advance (Herbort and Butz, 2012). Within a more realistic context, task, tool familiarity, the spatial orientation of the tool affect the planning and production of relevant actions.

4.2 Spatial viewing bias

In this study, we also investigated the spatial viewing bias of the fixations on the tools across the anticipatory period. The time periods for significant differences in task somewhat overlap in the two experiments and give a view of the timing when goal-oriented factors can significantly bias fixations towards the tool effector. Importantly, this relative timing shows when fixations are directed towards the tool effector to make inferences about which tool-specific actions need to be produced. Moreover, irrespective of the task, tool familiarity biased the fixations towards the effector for the entire 3s viewing period for both experiments I and II. These results reinforce the notion that with limited tool-specific knowledge, attention is allocated to specific locations to extract information about the mechanical properties of the tool and is again necessary for the production of the relevant actions. In contrast, for familiar tools, the gaze is guided parsimoniously towards the tool effector as new information is not needed in anticipation of the action. Furthermore,

the spatial orientation of the tool only elicited significant differences in experiment-II from 0.75s to 1.8s. Here, the realism of the interaction played a key role in planning the relevant action where the planning of the end-state comfort of the hand occurred relatively early in the viewing period and well in advance of any action initiation.

The interactions of task and tool familiarity were found in both experiments. In experiment-I, the interaction was in the period from 0.3s to 2.55s and in experiment-II from 0.45s to 1.5s. These clusters show that gaze is differentially guided more towards the tool effector when task-specific tool movements are planned for unfamiliar tools. Moreover, the planning based on task and tool familiarity is relevant early in the anticipatory period. The interaction of task and tool orientation was again seen only in experiment-II in the 0.3s to 3s time period and interaction of tool familiarity and orientation from 0.15s to 3s. These results further suggest that fixations are directed towards the handle of the tool in anticipation of planning the proximal goal of an optimal grasp. Interestingly, these results indicate that the optimal planning of grasp is initiated from the beginning of the viewing time window till the end and might be more critical than inspecting the tool effector to produce the correct action. (Johansson et al., 2001) showed similar findings where eye movements are predominantly made towards the grasp contact points. Moreover, one could also posit that the anticipatory gaze behavior related to task and familiarity is seen only when additional planning of grasp control is not needed. Further, the planning of optimal motor control might supersede planning based on other variables. Taken together, our study expands on the findings of Belardinelli et al. by showing that anticipatory gaze is highly contextualized to the realism of the environment.

4.3 Limitations of the present study

We conducted the present study in virtual reality, which is still a burgeoning technology for vision research. While VR environments pose an exciting avenue of research, there are still limitations that practitioners must face while conducting experiments in these scenarios. First, the realistic setup of both experiments I and II afforded natural head movements. To maintain optimal quality over the data, we asked the participants in the study to make limited head movements. Additionally, we presented the tools and the task cues not to cause extreme pitch head movements. Secondly, mobile eye-trackers can be error-prone and might suffer from erratic sampling rates (Ehinger et al., 2019) or calibration errors due to slippage (Niehorster et al., 2020). To mitigate any calibration errors, we also made sure that we calibrated the eye-trackers at regular intervals. Thirdly, LeapMotion is still a new technology. Although we made sure subjects practiced the interaction method before the experiment, it could have been challenging for participants to get used to. Finally, although we simulated grasping actions with the tool using LeapMotion, it is an inadequate substitute for a real grasp action where the tactile feedback of the tool in hand might elicit more accurate responses. In light of these technological limitations, we are aware that our study must be considered from a nuanced perspective. Furthermore, there is still room for replicating our study with novel and more realistic means.

5 Conclusion

In the present study, we successfully replicated and extended a lab-based study to a more veridical and ecologically valid context. Our results support the hypothesis that eye movements serve the cognitive function of actively sampling information from the environment to produce actions adequately. When semantic information about the object is not readily available, eye movements seek information from its mechanical properties. Furthermore, fixations are made in a goal-oriented way in anticipation of the relevant action. When taking into account the realism of the interaction, our results show that eye movements prioritize proximal goals of optimal grasp over task-based demands. Furthermore, the gaze dynamics before the initiation of action reveal an unfolding sensorimotor process. Hence, future research can focus on whether these gaze dynamics might still exist for a passive observer who does not have to produce task-relevant actions. Lastly, our study is at the frontiers of naturalistic vision research where novel technologies can be harnessed to answer questions that were previously far-fetched.

Author Contributions

AK, PK: conceived and designed the study. NG, AK, FNN: programmed the controller study. SB, AK, FNN: programmed the LeapMotion study. NG, SB: data collection. AK: data analysis. AK: initial draft of the manuscript. AK, SB, NG, FNN, PK: revision and finalizing the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgement

We are grateful for the financial support by: the German Federal Ministry of Education and Research for the project ErgoVR (Entwicklung eines Ergonomie-Analyse-Tools in der virtuellen Realität zur Planung von Arbeitsplätzen in der industriellen Fertigung)-16SV8052; German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) project number GRK2340/1 (DFG Research Training Group Computational Cognition).

References

M. A. Arbib, J. B. Bonaiuto, S. Jacobs, and S. H. Frey. Tool use and the distalization of the end-effector. *Psychological research*, 73(4):441–462, July 2009. URL http://dx.doi.org/10.1007/s00426-009-0242-2.

J. Baumard, F. Osiurak, M. Lesourd, and D. Le Gall. Tool use disorders after left brain damage. *Frontiers in psychology*, 5:473, May 2014. URL http://dx.doi.org/10.3389/fpsyg.2014.00473.

A. Belardinelli, O. Herbort, and M. V. Butz. Goal-oriented gaze strategies afforded by object interaction. *Vision research*, 106:47–57, Jan. 2015. URL http://dx.doi.org/10.1016/j.visres.2014.11.003.

A. Belardinelli, M. Barabas, M. Himmelbach, and M. V. Butz. Anticipatory eye fixations reveal tool knowledge for tool interaction. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, 234(8): 2415–2431, Aug. 2016a. URL http://dx.doi.org/10.1007/s00221-016-4646-0.

A. Belardinelli, M. Y. Stepper, and M. V. Butz. It's in the eyes: Planning precise manual actions before execution. *Journal of vision*, 16(1):18, 2016b. URL http://dx.doi.org/10.1167/16.1.18.

A. Berti and F. Frassinetti. When far becomes near: remapping of space by tool use. *Journal of cognitive neuroscience*, 12(3):415–420, May 2000. URL http://dx.doi.org/10.1162/089892900562237.

M. S. Castelhano, M. L. Mack, and J. M. Henderson. Viewing task influences eye movement control during active scene perception. *Journal of vision*, 9(3):6.1–15, Mar. 2009. URL http://dx.doi.org/10.1167/9.3.6.

R. R. Corbeil and S. R. Searle. Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics: a journal of statistics for the physical, chemical, and engineering sciences*, 18(1): 31–38, Feb. 1976. URL https://www.tandfonline.com/doi/abs/10.1080/00401706.1976.10489397.

B. V. Ehinger, K. Groß, I. Ibs, and P. König. A new comprehensive eye-tracking test battery concurrently evaluating the pupil labs glasses and the EyeLink 1000. *PeerJ*, 7:e7086, July 2019. URL http://dx.doi.org/10.7717/peerj.7086.

A. K. Engel, A. Maye, M. Kurthen, and P. König. Where's the action? the pragmatic turn in cognitive science. *Trends in cognitive sciences*, 17(5):202–209, May 2013. URL http://dx.doi.org/10.1016/j.tics.2013.03.006.

A. Farnè, A. Iriki, and E. Làdavas. Shaping multisensory action-space with tools: evidence from patients with cross-modal extinction. *Neuropsychologia*, 43(2):238–248, Jan. 2005. URL http://dx.doi.org/10.1016/j.neuropsychologia.2004.11.010.

M. M. Hayhoe. Advances in relating eye movements and cognition. *Infancy: the official journal of the International Society on Infant Studies*, 6(2):267–274, Sept. 2004. URL http://dx.doi.org/10.1207/s15327078in0602_7.

J. M. Henderson. Gaze control as prediction. *Trends in cognitive sciences*, 21(1):15–23, Jan. 2017. URL http://dx.doi.org/10.1016/j.tics.2016.11.003.

J. M. Henderson and T. R. Hayes. Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature human behaviour*, 1(10):743–747, Oct. 2017. URL http://dx.doi.org/10.1038/s41562-017-0208-0.

O. Herbort and M. V. Butz. The continuous end-state comfort effect: weighted integration of multiple biases. *Psychological research*, 76(3):345–363, May 2012. URL http://dx.doi.org/10.1007/s00426-011-0334-7.

J. Hermsdörfer, Y. Li, J. Randerath, G. Goldenberg, and L. Johannsen. Tool use without a tool: kinematic characteristics of pantomiming as compared to actual use and the effect of brain damage. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, 218(2):201–214, Apr. 2012. URL http://dx.doi.org/10.1007/s00221-012-3021-z.

R. S. Johansson, G. Westling, A. Bäckström, and J. R. Flanagan. Eye-hand coordination in object manipulation. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 21(17):6917–6932, Sept. 2001. URL https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.21-17-06917.2001.

S. H. Johnson and S. T. Grafton. From 'acting on' to 'acting with': the functional anatomy of object-oriented action schemata. In *Progress in Brain Research*, volume 142, pages 127–139. Elsevier, Jan. 2003. URL https://www.sciencedirect.com/science/article/pii/S0079612303420104.

S. H. Johnson-Frey. The neural bases of complex tool use in humans. *Trends in cognitive sciences*, 8(2):71–78, Feb. 2004. URL http://dx.doi.org/10.1016/j.tics.2003.12.002.

A. Keshava, A. Aumeistere, K. Izdebski, and P. Konig. Decoding task from oculomotor behavior in virtual reality. In *ACM Symposium on Eye Tracking Research and Applications*, number Article 30 in ETRA '20 Short Papers, pages 1–5, New York, NY, USA, June 2020. Association for Computing Machinery. URL https://doi.org/10.1145/3379156.3391338.

A. Keshava, F. N. Nezami, H. Neumann, K. Izdebski, T. Schüler, and P. König. Just-in-time: gaze guidance behavior while action planning and execution in VR. Feb. 2021. URL https://www.biorxiv.org/content/10.1101/2021.01.29.428782v1.abstract.

G. Króliczak, C. Cavina-Pratesi, D. A. Goodman, and J. C. Culham. What does the brain do when you fake it? an FMRI study of pantomimed and real grasping. *Journal of neurophysiology*, 97(3):2410–2422, Mar. 2007. URL http://dx.doi.org/10.1152/jn.00778.2006.

S. Ladouce, D. I. Donaldson, P. A. Dudchenko, and M. Ietswaart. Understanding minds in Real-World environments: Toward a mobile cognition approach. *Frontiers in human neuroscience*, 10:694, 2016. URL http://dx.doi.org/10.3389/fnhum.2016.00694.

M. F. Land and M. Hayhoe. In what ways do eye movements contribute to everyday activities? *Vision research*, 41 (25-26):3559–3565, 2001. URL http://dx.doi.org/10.1016/s0042-6989(01)00102-x.

S. G. Luke. Evaluating significance in linear mixed-effects models in R. *Behavior research methods*, 49(4):1494–1502, Aug. 2017. URL http://dx.doi.org/10.3758/s13428-016-0809-y.

A. Maravita, C. Spence, S. Kennett, and J. Driver. Tool-use changes multimodal spatial interactions between vision and touch in normal humans. *Cognition*, 83(2):B25–34, Mar. 2002. URL http://dx.doi.org/10.1016/s0010-0277(02)00003-3.

D. C. Niehorster, T. Santini, R. S. Hessels, I. T. C. Hooge, E. Kasneci, and M. Nyström. The impact of slippage on the data quality of head-worn eye trackers. *Behavior research methods*, 52(3):1140–1160, June 2020. URL http://dx.doi.org/10.3758/s13428-019-01307-0.

F. J. Parada. Understanding natural cognition in everyday settings: 3 pressing challenges. *Frontiers in human neuroscience*, 12:386, Sept. 2018. URL http://dx.doi.org/10.3389/fnhum.2018.00386.

F. J. Parada and A. Rossi. Perfect timing: Mobile Brain/Body imaging scaffolds the 4e-cognition research program (published version). Mar. 2020. URL https://osf.io/wru4j.

J. B. Pelz and R. Canosa. Oculomotor behavior and perceptual strategies in complex tasks. *Vision research*, 41(25-26): 3587–3596, 2001. URL http://dx.doi.org/10.1016/s0042-6989(01)00245-0.

D. J. Schad, S. Vasishth, S. Hohenstein, and R. Kliegl. How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. July 2018. URL http://arxiv.org/abs/1807.10451.

B. Voloh, M. R. Watson, S. König, and T. Womelsdorf. MAD saccade: statistically robust saccade threshold estimation via the median absolute deviation. *Journal of eye movement research*, 12(8), May 2020. URL http://dx.doi.org/10.16910/jemr.12.8.3.