Action Affordance Affects Proximal and Distal Goal-oriented Planning

Ashima Keshava[1], Nina Gottschewsky[1], Stefan Balle[1], Farbod Nezami[1], Thomas Schüler[2], and Peter Konig[1]

[1]University of Osnabrück
[2]Halocline GmbH & Co. KG

November 25, 2022

Abstract

Visual attention is mainly goal-directed and allocated based on the upcoming action to be performed. However, it is unclear how far this feature of gaze behavior generalizes 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 action affordance. In a between-subject design, a cohort of participants interacted with a VR controller in a low realism environment; another performed the task with an interaction setup that allowed differentiated hand and finger movements in a high-realism environment. We investigated the differences in odds of fixations and their eccentricity towards the tool parts before action initiation. The results show that participants fixate more on the tool's effector part before action initiation when asked to use the tool and during interaction with unfamiliar tools. The spatial viewing bias on the tool reveals early fixations are influenced by the task and the familiarity of the tools. Our findings suggest that fixations are made in a task-oriented way to plan the intended action well before action initiation. With more realistic action affordances, more fixations were allocated toward the tool handle. We hypothesize that these fixations are made towards the proximal goal of planning the grasp even though the perceived action on the tools is identical for both experimental setups. Taken together, proximal and distal goal-oriented planning is contextualized to the realism of action/interaction afforded by an environment.

Introduction

A longstanding goal of the cognitive sciences is understanding cognition, behavior, and experience as it unfolds in the natural world (Parada & 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 to new insights (Ladouce et al., 2016; Parada, 2018). More recently, a pragmatic turn has emerged in the field where there is a more significant push towards incorporating the body and bodily action to infer cognitive function (Engel et al., 2013). **These technological advancements and philosophical motivations have opened the door to further inquiry into the contribution of the acting body to cognitive processing.**

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 & 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 & 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 processing.

When using tools, a wealth of information is parsed to produce the relevant action. The semantic knowledge associated with the tool helps understand how it is used, and the mechanical knowledge maps the physical properties of the tool for potential usage. 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) necessitates planning any action associated with the tool. When this knowledge is not readily available, inferential processes must be deployed to deduce the relevant action.

Henderson(2017) proposed that gaze control in natural scenes can be characterized as the result of knowledge-driven predictions. These predictions for gaze control are based on the knowledge gained from past experiences. Predictive gaze control aims to sample informative and meaningful information in the environment relevant to the cognitive system's current needs. Furthermore, gaze control proactively samples information in anticipation of the following action (Hayhoe, 2004; Land & Hayhoe, 2001; Pelz & Canosa, 2001). Belardinelli et al. (2016b) showed that eye movements are goal-oriented and are modulated in anticipation of the upcoming object interaction. There is strong evidence that task plays a vital role in how the eyes scan the scene and are differentiated between passive viewing and actual interaction (Belardinelli et al., 2015). Similarly, Keshava et al. (2020) showed that rudimentary object interactions could be predicted using eye-movement data alone. Even without an active interaction, task relevance plays an important role (Castelhano et al., 2009; Henderson & Hayes, 2017). These studies point towards gaze control being the consequence of knowledge and task-driven predictions.

Similarly, Belardinelli et al. (2016a) investigated the role of anticipatory eye movements when interacting with familiar and unfamiliar tools in a controlled lab setting. These tools had differentiable parts: tool handle and effector. The results showed that in the case of unfamiliar tools, initial fixations are made on 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 action 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 as images on a screen, and participants pantomimed 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 and can be reproduced in more realistic environments.

Herbort and Butz (2011) further investigated the interaction of habitual and goal-directed processes that affect grasp selection while interacting with everyday objects. They presented objects in different orientations and showed that grasp selection depended on the overarching goal of the movement sequence dependent on the object's orientation. Belardinelli et al. (2016b) further showed that fixations have an anticipatory preference for the region where the index finger is placed. Consequently, these studies show that the location of fixations is predictive of both proximal goals of manual planning and task-related distal goals.

When studying anticipatory behaviors corresponding to an action, there must be a distinction between symbolized or pantomimed vs. actual actions. Króliczak et al. (2007) showed that brain areas typically involved in real actions are not driven by pantomimed actions, and that pantomimed grasps do not activate the object-related regions 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., a grasp) can be essential when studying anticipatory behavioral control.

As research steadily moves towards a more ecological view of cognitive processing with bodi-

ly actions and interactions with the environment, there is also a greater need to understand behavioral differences induced by varying degrees of action affordances. Virtual reality (VR) allows us to test these behavioral differences by varying the realism of interactions from less completely artificial (e.g., showing the stimuli on a 2D plane and responding with button presses) to high fidelity (e.g., interacting with highly realistic 3D models). Chalmers and Ferko (2008) posit that defining levels of realism is necessary to achieve a one-to-one mapping of an experience in the virtual environment with the same experience in the real environment. Moreover, a one-to-one mapping is necessary for research purposes to avoid misrepresenting the real environment. In virtual reality (VR), realistic actions can be produced by different kinds of interaction methods. Using interfaces such as VR controllers, ego-centric visual feedback of a hand can be simulated. These interfaces usually consist of hand-held devices tracked in space and through which different actions are controlled by pressing buttons. One advantage of controller-based VR interaction is the possibility of haptic feedback. However, a disadvantage is that the actual hand posture while holding the controller does not necessarily correspond to the user's simulated hand as they perform the action. Conversely, camera-based interaction interfaces, such as LeapMotion, capture the real-time movements of the user's hand and finger gestures, like wrap grasp or pinch grasp, to control different actions in the environment. These interfaces give the user a realistic simulation of their finer hand and finger movements, while they cannot give direct haptic feedback of the gripped object. Consequently, the chosen method of interaction in VR can afford different levels of realism and could elicit different behavioral responses.

We investigated anticipatory gaze control concerning tool interactions in two experiments in the present study. We were interested in the extent to which the action affordance and the environment modulated active inference processes exhibited by gaze behavior. In experiment-I, subjects performed the experiment in a low-realism environment. They interacted with the tool models using a VR controller, which produced a grasp in the virtual environment by pulling the index finger button. In experiment-II, subjects were immersed in a high-realism setting where they interacted with the tools using LeapMotion, which required natural hand and finger movements. Thus, the action affordance appeared closer to the real world. Furthermore, in both experiments, participants interacted with 3D models of tools by lifting or using them in VR. For the stimuli set, we used familiar or unfamiliar tools as described in Belardinelli et al. (2016a). Additionally, to differentiate between proximal and distal goal planning, we manipulated the spatial orientation of the tool handle so that they were either congruent or incongruent with the subjects' handedness. With this experimental design, we investigate the influence of task, tool familiarity, the spatial orientation of the tool, and, notably, the impact of the action affordance on anticipatory gaze behavior.

2. Methods

2.1 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. The time course of the trials is illustrated in Figure 1A. At the start of a trial, subjects saw the cued task for 2s after which the cue disappeared, and 0.5s later, a tool appeared on the virtual table. Subjects were given 3s 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 subsequent trial.

2.2 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 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.

2.3 Experimental Design and Procedure

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

Figure 1C illustrates two exemplar trials from the experiments. We used a 2x2x2 experimental design for both experiments, 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 either simple, familiar tools or tools that are not seen in everyday contexts and are unfamiliar. Handle orientation corresponded to the position of the tool handle, which was presented to the participants either on the LEFT or the RIGHT side. Both experiments had 144 trials per participant, with an equal number of trials corresponding to the three factors. Subjects performed the trials over six blocks of 24 trials each. We simultaneously measured the eye and hand movements while the 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* 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 experiment 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 on a 5-point Likert-like scale where the lowest rating (1) corresponded to "I have never used it or heard about it," and the highest rating (5) referred to "I see it every day or every week."

2.4 Experimental Stimuli

The experimental setup consisted of a virtual table that mimicked the table in the real world. The table's height, width, and length were 86cm, 80cm, and 80cm, respectively. In experiment-I, subjects were 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 with the exact dimensions of the real-world table and a buzzer on the right to indicate the end of a trial. 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), the mean breadth was 7.60cm (SD=3.68), and the mean height= was 4.17cm (SD=2.13). To determine the tool effector and handle regions of interest, we halved the length of the 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-tracker11https://enterprise.vive.com/us/product/vive-pro-eye-office/. 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 the 5-point calibration function provided by the manufacturer. To ensure the calibration error was less than 1 * , we performed a 5-point validation after each calibration. As the experiments allowed a lot of natural head movements, the eye tracker was calibrated repeatedly during the experiment after each block of 48 trials. We designed the experiment using the Unity3D game engine22Unity, www.unity.com (v2019.2.14f1) and controlled the eye-tracking data recording using HTC VIVE Eye Tracking SDK SRanipal33SRanipal, developer.vive.com/resources/vive-sense/sdk/vive-eye-tracking-sdk-sranipal/ (v1.1.0.1).

In experiment-I, we used HTC Vive controller44SteamVR, https://valvesoftware.github.io/steamvr_unity_-plugin/articles/Quickstart.html (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. Subjects pulled the trigger button of the controller over the virtual tools to interact with the tools, and the rendered hand grasped the tool handle **(See SI Appendix Movie-1 for the time course of the trials from a participant's perspective)** .

Similarly, in experiment-II, we used LeapMotion55LeapMotion Unity modules, https://developer.leapmotion.com/unity (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

(See SI Appendix Movie-2 for the time course of the trials from a participant's perspective).

2.6 Data pre-processing

2.6.1 Data Rejection

For both experiments, we rejected trials based on two criteria. Firstly, we rejected trials where the hand position was not recorded. Secondly, we rejected trials where the gaze position and direction vectors were not recorded or recorded as invalid. In experiment-I, we rejected 29.8% (SD= $+\text{-}\ 9.\ 6$**) of trials over 18 participants. In experiment-II, we rejected a mean of 35.5% (SD=** $+\text{-}\ 19.\ 24$**) of trials over 30 participants. Experiment II showed a greater rejection rate as the LeapMotion camera lost hand-tracking more often.**

2.6.2 Gaze Data

As a first step, using the eye-in-head 3D gaze direction vector for the cyclopean eye, we calculated the gaze angles in degrees for the horizontal ϑ_{h} and verticalϑ_{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 coordinates. In our setup, the x coordinate corresponds to the left-right direction, y in the up-down direction, and z in the forward-backward direction. The formulas used for computing the gaze angles are as follows:

$$\theta_h \ = \ \frac{180}{\pi} \arctan\left(\frac{x}{z}\right)$$

$$\theta_v \ = \ \frac{180}{\pi} \arctan\left(\frac{y}{z}\right)$$

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:

$\omega_{\mathrm{h}} \vartheta_{\mathrm{h}}/t$

$\omega_{\mathrm{v}} \vartheta_{\mathrm{v}}/t$

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

$\omega =$

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 * /s. 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 step was done until the difference between ϑ_{n} and $\vartheta_{\mathrm{n+1}}$ was less than or equal to 1 * /s. 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 of less than 50 ms or were larger than 3.5 times the median absolute deviation of the fixation duration. For further data analysis, we only considered the fixations 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.

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. We analyzed the fixations in the 3 seconds from when the tool was presented until the go cue. 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 tool handle (LEFT, RIGHT) and the experiment interaction method (CONTROLLER, LEAP MOTION). All within-subject factors were also modeled with random intercepts and slopes for each subject.

We used effect coding (Schad et al., 2020) to construct the design matrix for the linear model, where we coded the categorical variables LIFT, FAMILIAR, RIGHT, CONTROLLER to -0.5 and USE, UNFAMILIAR, LEFT, LEAP MOTION 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 & 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 degrees of freedom of the fixed effects. For both experiments, the Wilkinson notation (Wilkinson & Rogers, 1973) of the model was:

$$\log\left(\frac{p(fixation\ on\ effector)}{p(fixation\ on\ handle)}\right) = 1 + task * \text{familiarity} * \text{handle orientation} * \text{interaction method}$$
$$+ (1 + task * \text{familiarity} * handle\ orientation \mid Subject)$$

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 and the effect of the interaction method used in the experiment groups.

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 when the tool was visible on the table (3s from the start of the trial) until the go cue indicated when subjects could start manipulating the tool. We divided this 3s period into 20 equal bins of 150ms each. We calculated the median distance of the fixations from the tool center for each trial and time bin. Next, we normalized the distance with the length of the tool so that we could compare the fixation eccentricity across different tools.

We used the cluster permutation method to find the time points for significant differences between the three conditions and their interactions. Here, we use the t-statistic as a test statistic for each time-bin, where t is defined as

$$t \ = \ * \frac{x}{\sigma}$$

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 corresponding to the t-value where the p-value is 0.05. 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 representing the cluster's mass. 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. The permutations 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 three different conditions, their interactions, and the corresponding p-values.

2.7.3 Subjective Rating of Tool Familiarity

To validate that the tool familiarity categorization in our experiment design aligned with participants' subjective rating, we analyzed the questionnaire completed by participants at the end of each experiment. We calculated the mean subjective rating of familiar and unfamiliar tools. We performed a mixed-ANOVA with familiarity as a within-subject factor, the experiment group as the between-subject factor, and the subjective rating of the tool as the dependent variable.

2.7.4 Learning Effects

To quantify the learning effects on fixation patterns due to the repeated presentation of familiar and unfamiliar tools, we computed the relative change in fixations on the tool effector. We computed the mean proportion of fixations on tool effector for each participant in the first five versus the last five familiar and unfamiliar trials. We subsequently calculated the percent change (C) from early trials to late trials for each tool familiarity using the following formula:

$$C \ = \ 100 \ * \ \frac{X_f \ - \ X_i}{X_i}$$

where X_f denotes the mean proportion of fixations on tool effector in the last five trials, and X_i denotes the mean proportion of fixations on tool effector in first five trials for a given subject. To statistically assess the differences between the experimental groups and the tool familiarity, we performed a mixed-ANOVA with C as the dependent variable, tool familiarity as a within-subject factor, and experiment interaction method as a between-subject factor.

2.7.5 Difference between experiment groups

Next, we assessed whether participants differentially allocated fixations to the overall environment and the tools in the two experiments. We calculated the percentage of fixations allocated to the environment vs. the tool during the 3s viewing period in each trial. To statistically assess the difference in the percentage of fixations allocated to the tools vs. environment, we performed a mixed-ANOVA with fixation location (tools vs. environment) as a within-subject factor, the two experiments as a between-subject factor, and the percentage of fixations as the dependent variable.

3. Results

In the present study, we investigated the differences in gaze-based planning strategies dependent on task, tool familiarity, and handle orientation. We investigated two anticipatory gaze-based strategies 3s before action initiation. Firstly, the odds of fixations in favor of the tool effector. Secondly, the dynamic changes in the fixations' eccentricity relative to the tool's center. We further compared the behavioral differences in two experiments that had the same experimental conditions but differed based on the realism of sensory input and interaction in VR.

3.1 Odds of Fixations in favor of tool effector

First, we compared the log-odds of fixations in favor of the tool effector across the three conditions: task, tool familiarity, and handle orientation in the 3s period when the subjects studied the tool. Figure 3A shows the log-odds of the fixations on the tool effector for experiment-I (with VR Controllers) and experiment-II (with LeapMotion). In experiment-I (Figure 2A, left panel), subjects showed a mean log odds of 0.04 (95%CI = [- 0.04, 0.11]) for the LIFT task, and for the USE task, the mean log-odds were 0.17 (95%CI = [0.08, 0.26]). For the FAMILIAR tools, the mean log-odds in favor of the tool effector were -0.14 (95%CI = [-0.21, -0.07]) and for UNFAMILIAR 0.34 (95%CI = [0.23, 0.45]). For the RIGHT-oriented tool handle, the mean log-odds were 0.14 (95%CI = [0.06, 0.22]), and for the LEFT-oriented tool handle, the mean log-odds were 0.08 (95%CI = [-0.09, 0.26]). In experiment-II (Figure 2A, right panel), subjects showed a mean log odds of 0.16 (95%CI = [-0.01, 0.31]) of fixations on the tool effector for the LIFT task. In the USE task, the mean log-odds were 0.24 (95%CI = [0.09, 0.38]). For the FAMILIAR tools, the mean log-odds in favor of the tool effector were 0.08 (95%CI = [-0.06, 0.23]) and for UNFAMILIAR 0.30 (95%CI = [0.15, 0.44]). For the RIGHT-oriented tool handle, the mean log-odds were 1.31 (95%CI = [1.21, 1.42]), and for the LEFT-oriented tool handle, the mean log-odds were -0.32 (95%CI = [-0.45, -0.19]).

To statistically assess the significant differences between the experimental factors, we used linear mixed models. We used effect coding for the linear model to directly interpret the regression coefficients as main effects. Figure 2B shows the estimated regression coefficients and their associated effect sizes. Irrespective of the experiment groups, there was a main effect of factor task (USE - LIFT) $\beta = 0.1$ (95%CI = [0.02, 0.182], t(50.32)=2.35, p = 0.02, $\eta^2 = 0.1$, 90%CI = [0.01, 0.24])). There was also a significant main effect of familiarity (UNFAMILIAR - FAMILIAR) $\beta = 0.40$ (95%CI = [0.33, 0.46], t(89.14)=11.52, $p < 0.001$,$\eta^2 = 0.6$, 90%CI = [0.49, 0.68]). There was a main effect of handle orientation (LEFT - RIGHT) β= -0.85, (95%CI = [-1.01, -0.70], t(45.75)=-10.52, $p < 0.001$, $\eta^2 = 0.71$, 90%CI = [0.59, 0.78]). There was an overall significant interaction of task and familiarity with $\beta = 0.16$, 95%CI = [0.04, 0.29], t(184.17) = 2.53, p = 0.01, η^2= 0.03, 90%CI = [0.0, 0.09]). The was a somewhat significant interaction of task and handle orientation β= 0.15, 95%CI = [0.01, 0.29], t(71.32) = 2.11, p =

0.04,η^2 = 0.06, 90%CI = [0.0, 0.17]). The interaction of familiarity and orientation was not significantβ = 0.05, 95%CI = [-0.1, 0.2], t(51.03) = 0.63, p = 0.53, η^2 = 0.01, 90%CI = [0.0, 0.09]). In sum, irrespective of the interaction methods employed in the two experiments, there were significantly higher odds of fixations on effector in the USE task, and for UNFAMILIAR tools. Moreover, there were significantly lower odds of fixations on the effector when the tool handle was oriented to the LEFT and incongruent with the subjects' handedness.

The linear model also showed the effects of the interaction method used in the two experiments. There was no significant difference between the experiments for the factor of task (β = -0.09, 95%CI = [-0.26, 0.07], t(50.32) = -1.12, p = 0.27,η^2 = 0.02, 90%CI = [0.0, 0.13]). There was a significant difference in the effects of familiarity between the experiment groups (β = -0.19, 95%CI = [-0.33, -0.06], t(89.14) = -2.78, p = 0.01, η^2= 0.08, 90%CI = [0.01, 0.18]). There was a significant difference between the experimental groups for the factor handle orientation β = -1.62 (95%CI = [-1.94, -1.31], t(45.75) = -10.0, p = 0.0, η^2 = 0.69, 90%CI = [0.56, 0.77]), where the LeapMotion experiment biased fixations towards the tool handle as compared to the controller experiment when the tool handle was oriented to the LEFT. The interaction of task and familiarity between the experimental groups was also not significantly different (β = 0.04, 95%CI = [-0.22, 0.29], t(184.17) = 0.28, p = 0.78,η^2 = 0.0, 90%CI = [0.0, 0.02]). There was a significant difference between groups in the interaction effects of familiarity and handle orientation (β = 0.5, 95%CI = [0.2, 0.79], t(51.03) = 3.24, p = 0.0,η^2 = 0.17, 90%CI = [0.04, 0.32]). There was no significant difference between groups for interaction between task and handle orientation (β = 0.04, 95%CI = [-0.24, 0.32], t(71.32) = 0.26, p = 0.8,η^2 = 0.0, 90%CI = [0.0, 0.04]). The largest difference between the experimental groups was seen when the tool handle was oriented to the LEFT and incongruent with the subjects' handedness.

3.2 Spatial bias of fixations on the tools

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 periods where the experimental conditions elicited significant differences. As shown in Figure 3A, in experiment-I, the differences in task (USE - LIFT) were significant from 1.06s to 2.6s after tool presentation, $p < 0.001$. Differences in tool familiarity (UNFAMILIAR - FAMILIAR) were significant from 0.4s to 3s with a p-value <0.001. Moreover, the differences in the handle orientation (LEFT - RIGHT) were not significant. The interaction of task and familiarity was significant from 0.4s to 1.8s, p=0.02. In sum, while interacting with the VR controller, there were effects of USE task and UNFAMILIAR tools that biased the fixations towards the tool effector and the orientations of the tool handle did not affect this bias significantly.

Similarly, Figure 3B shows the eccentricity of fixations from the center of the tool during the 3s period when the subjects studied the tool in experiment-II with LeapMotion as the interaction method. The differences in task were significant from 0.6s to 3s, p< 0.001. The differences in familiarity were significant from 0.2s to 3s, $p < 0.001$. Furthermore, the differences in handle orientation were significant from 0.2s to 3s, $p < 0.001$. There was a significant interaction of task and familiarity from 0.6s to 1.6s with p = 0.008 and from 2.2s to 2.8s, p = 0.03. There were also significant clusters in the interaction of tool familiarity and handle orientation from 0.4s to 1.4s with p = 0.007 and from 1.8s to 3s, p = 0.007. In sum, while interacting with LeapMotion, the use task and unfamiliar tools biased the anticipatory gaze towards the tool's effector. Importantly, the left-oriented tool handle biased the fixations towards the tool's handle.

3.3 Subjective Rating of Tool Familiarity

Next, we analyzed how the participants subjectively assessed the familiarity of the 12 tools and if there were any differences between the subjective ratings between the participants in experiment-I and II. Figure 4A shows the mean subjective familiarity ratings for the familiar and unfamiliar tools used in the study. The mean familiarity rating for familiar tools in experiment-I was 4.55 (SD=0.60) and for unfamiliar tools 1.81 (SD=1.17). In experiment-II, the mean familiarity rating for familiar tools was 4.48 (SD=0.52) and for unfamiliar tools 1.56 (SD=1.04). A mixed-ANOVA showed no differences in the familiarity ratings between the two experiments (F(1, 44)=3.08, p=0.06, $\eta^2 = 0.07$). Furthermore, there was a significant difference in the subjective rating of the tools (F(1, 44)=3094.05, p<0.001, $\eta^2 = 0.98$), i.e. the familiar tools were rated as subjectively familiar and the unfamiliar tools were rated subjectively unfamiliar. There were also no significant interactions between the two factors (F(1, 44)=2.52, p=0.11, $\eta^2 = 0.069$). In sum, our experimental condition of familiarity was consistent with the participants' subjective rating as well.

3.4 Learning Effects

To determine if participants fixated on the tools' effector differently in later trials compared to earlier trials, we computed the percentage change of fixations on tool effector per tool familiarity. Figure 4B shows the change in fixations on tool effector from first five and last five trials of familiar and unfamiliar tools in the two experiments. The mean percentage change in fixations on tool effector in experiment-I was 4.02 (SEM=7.27) for familiar tools and 12.32 (SEM=6.98) for unfamiliar tools. The mean percentage change in fixations allocated to tool effector in experiment-II was 4.47 (SEM=8.72) for familiar tools and 14.60 (SEM=6.32) for unfamiliar tools. The mixed-ANOVA showed no significant effect of the interaction method on the percent change in fixations on tool effector (F(1, 46)=0.03, p=0.85, $\eta^2 = 0.0007$). There was also no significant effect of tool familiarity on percent change in fixations on the tool effector (F(1, 46)=1.26, p=0.26 $\eta^2 = 0.02$). There was also no significant interaction between the experiment groups and tool familiarity (F(1,46)=0.01, p=0.91 $\eta^2 = 0.0002$). Hence, we can conclude that participants did not allocate fixations differently on the tool effector in later trials compared to earlier trials.

3.5 Difference between experiment groups

Lastly, we wanted to ensure that the visual differences in the virtual environments did not affect the subjects' allocation of attention to the experimental task. We calculated the percentage of fixations on the tool vs. anywhere else in the environment in the two experiments. Figure 4C shows the percentage of fixations allocated to the tools vs. the environment for the two experiments during the 3s viewing period. For experiment-I with the interaction method of VR controller, the mean percentage of fixations on the environment was 0.29 (SD=0.06) and on the tools 0.80 (SD=0.09). Conversely, in experiment-II with LeapMotion as the interaction method, the mean percentage of fixations allocated to the environment was 0.30 (SD=0.15) and on the tools 0.80 (SD=0.14). The mixed-ANOVA showed no significant difference in the percentage of fixations between the two experiments (F(1, 47)=2.86, p=0.09,$\eta^2 = 0.05$). There were significant differences in the percentage of fixations located on the tool vs. the environment (F(1, 47)=217.47, p<0.001,$\eta^2 = 0.82$). We did not find any interactions between the two factors (F(1, 47)=0.02, p=0.87,$\eta^2 = 0.0004$).

These results show that fixations were primarily made in a task-oriented manner, and there is no evidence that they were affected by the visual differences in the virtual environment of the two experiments.

4. Discussion

The primary aim of this study is to investigate how gaze-based strategies vary for a given task, tool familiarity, and tool handle orientation in naturalistic settings and how the action affordance of the environment can affect this gaze behavior. The cued task required the production of tool-specific movements in the case of the use task and generic movements in the case of the lift task. We manipulated the factor of tool familiarity by presenting either familiar or unfamiliar tools. We further controlled the tool orientation with the handle oriented to the right or the left and were congruent and incongruent to the participants' handedness, respectively. Our experiment design differentiated gaze-based planning into proximal planning related to grasping the tools and distal goal-oriented planning of ultimately acting with the tools. Our study successfully added to the current body of research in two important ways. Firstly, irrespective of the interaction methods in the virtual environment, the number of fixations and their eccentricity were modulated by distal goal-oriented factors of task and tool familiarity. When cued to use the tools, there were higher odds of fixating on the effector of the tool as compared to lifting the tool. Similarly, when presented with unfamiliar tools, there were higher odds of fixations on the tool effector than familiar tools. This effect was more pronounced when participants were instructed to use an unfamiliar tool. Secondly, with a naturalistic interaction method that allowed for finer hand and finger movements, preparatory fixations were biased toward the tool handle when it was oriented incongruent to the subjects' handedness. In sum, our study shows that the action affordance of the virtual environment affects the anticipatory fixations related to proximal goal-oriented planning but not distal goal-oriented planning.

As behavioral studies in virtual reality become popular, there is a need to understand how the interaction methods in the virtual environment can affect behavioral outcomes. In our study, we conducted two experiments to disentangle the role of action affordance for goal-oriented planning. While in experiment-I, participants interacted with 3D tool models using VR controllers, which produced a virtual grasp by pressing the trigger button with their index fingers, in experiment-II, participants interacted by producing an actual grasp. Our results show that the action affordance provided in experiment-II greatly biased the fixations towards the tool handle. When the tool handle was incongruent with the subjects' handedness, this bias appeared as a greater number of fixations on the tool handle. These results can be interpreted as a bias in fixations to plan the grasp on an incongruently placed tool handle. This effect indicates an end-state comfort planning (Rosenbaum et al., 1996; Herbort & Butz, 2012) where preference is given to the final hand posture rather than a comfortable initial posture. While using VR controllers, such planning is not required as the hand posture is primarily fixed. In line with Pezzulo et al. (2008), our study shows that anticipatory eye movements can reveal motor planning that selects appropriate movements to achieve the current goal.

Furthermore, the different interaction methods did not affect distal goal-oriented planning. In both experiments, the odds of sampling visual information from 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 tool effector increased for unfamiliar tools. This effect was more pronounced when subjects were instructed to produce tool-specific movements for unfamiliar tools. These results of distal goal-oriented planning are in line with the findings reported by Belardinelli et al. (2016a). In our study, we show that well before action initiation, subjects attended to the relevant tool parts to produce the tool-specific actions irrespective of the hand posture afforded by the interaction method.

VR has been poised as a viable method to probe cognitive processing in ecologically valid settings without sacrificing experimental control (Parsons, 2015). In our study, we probe how the action affordance of the interaction methods in VR affects gaze-based planning behavior.

Our findings give a fuller view of gaze-based planning strategies needed to produce relevant action. Our study shows that irrespective of the interaction method, semantic and sensorimotor knowledge are inferred from the mechanical properties of the tools as revealed by anticipatory fixations. However, a natural interaction method can bias the anticipatory fixations towards the proximal goals of grasping the tool. Specifically, multiple factors such as semantic, sensorimotor knowledge, and end-state comfort planning contribute to anticipatory goal planning. Notably, our study shows that different constraints on the interaction method can also result in different anticipatory behavioral responses. From the perspective of Gibson (1977), the affordances of the environment are tightly linked to the actions one can perform in it. Similarly, O'Regan and Noë (2001) posited that actions constitute the cognitive processes that govern relevant sensorimotor contingencies. Hence, our study offers a veridical and ecologically valid context to aspects of anticipatory behavior control.

Studies in eye-hand coordination (Johansson et al., 2001; Lohmann et al., 2019; Belardinelli et al., 2018) have shown that eye movements are proactively made towards the grasp contact points. Furthermore, Flanagan et al. (2006) proposed that predictions are event-oriented and are at the heart of successful control strategies for object manipulations. They propose that predicted sensory events are compared with actual events like grasping, lifting and moving the object to monitor task progression. With our study, we make the case that gaze-based predictions are made for action outcomes at different time scales. Eye movements are used to simultaneously plan the proximal goal of grasping the tools and the distal task-specific requirements well before action initiation.

Our study adds to the growing body of evidence that anticipation and prediction are at the core of cognition (Pezzulo et al., 2007). Motor theories of cognition have proposed that simulations of actions reuse internal models of motor commands to effect multiple predictions (Jeannerod, 2006). The simulation of action theory has been used to explain numerous phenomena of planning, prediction of external events, visual perception, and imitation. Hoffmann (2003) introduced anticipatory behavior control as the mechanism by which action-effect representations are activated by the need for an effect-related goal and contingent stimuli. Furthermore, Pezzulo et al. (2021) recently proposed that generative models provide top-down predictive signals for perception, cognition, and action during active tasks. These signals are otherwise weak or absent when the brain is at rest, or the stimuli are weak. Our study shows that anticipatory behavior is tightly linked to the production of relevant actions and is contextualized to the action affordance of the environment.

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. First, the naturalistic setting of experiments I and II allowed natural head movements. To maintain the optimal quality of 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 variable sampling rates (Ehinger et al., 2019) or calibration errors due to slippage (Niehorster et al., 2020). We also ensured we calibrated the eye trackers at regular intervals to mitigate any calibration errors. Thirdly, both controller-based and camera-based VR interaction methods are still new technology. It could have been challenging for participants to get used to, even though we ensured they practiced the interaction before the experiment. While we simulated grasping the tool using LeapMotion's gesture recognition, it could still be an inadequate substitute for a real grasp where additional tactile feedback might change the behavior. In this vein, Itaguchi (2021) showed that grasping movements within a virtual environment do not differ quantitatively from grasping kinematics in the real world. Lastly, while there are visible differences between the environments, we did not see systematic differences between the percentage of fixations allocated to the tool and the rest of the environment in both experimental settings. Hence, we contend that the differences in the eye movement behavior reported in the study result from the differences in the action affordance and much less because of visual differences between the environments.

There are still some open questions about anticipatory behavior elicited by tool interactions. Our study provides a first step towards distinctly investigating proximal and distal goal-oriented planning. Firstly, while our study distinguishes between levels of action affordances, future work can look at goal-oriented planning for passive observers at both proximal and distal levels. Secondly, with mobile imaging, we can also probe into the predictive neural signals that give rise to such oculomotor behaviors.

5. Conclusion

The present study gives a veridical and ecologically valid context to planning and anticipatory gaze behavior. Our results support the hypothesis that eye movements serve the cognitive function of actively sampling information from the environment to produce relevant actions. When semantic information about the object is not readily available, eye movements seek information about its mechanical properties from specific locations. Furthermore, we show that fixations are made in a goal-oriented way in anticipation of the relevant action. When considering the realism of the action affordance, our results show that eye movements are affected by both proximal goals of manually grasping objects and the distal task-based demands. Lastly, our study is at the frontier of naturalistic vision research, where novel technologies can be harnessed to answer previously open questions.

Author Contributions

AK, PK: conceived and designed the study. TS, PK: Procurement of funding. 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, TS, PK: revision and finalizing the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgment

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).

The graphical abstract and the experimental setup figures in this paper have been designed using resources from Flaticon.com.

Conflict of Interest

The authors declare no conflict of interest.

Data Accessibility

The authors have made the data associated with this study available on request.

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. Belardinelli, J. Lohmann, A. Farnè, and M. V. Butz. Mental space maps into the future.*Cognition*, 176:65–73, July 2018. URL http://dx.doi.org/10.1016/j.cognition.2018.03.007. 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. A. Chalmers and A. Ferko. Levels of realism: from virtual reality to real virtuality. In*Proceedings of the 24th Spring Conference on Computer Graphics*, SCCG '08, pages 19–25, New York, NY, USA, Apr. 2008. Association for Computing Machinery. ISBN 9781605589572. URL https://doi.org/10.1145/1921264.1921272. 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. J. R. Flanagan, M. C. Bowman, and R. S. Johansson. Control strategies in object manipulation tasks. *Current opinion in neurobiology*, 16(6):650–659, Dec. 2006. URL http://dx.doi.org/10.1016/j.conb.2006.10.005. J. J. Gibson. The theory of affordances. *Hilldale, USA*, 1(2):67–82, 1977. 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. Habitual and goal-directed factors in (everyday) object handling.*Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, 213(4):371–382, Sept. 2011. URL http: //dx.doi.org/10.1007/s00221-011-2787-8. 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. J. Hoffmann. Anticipatory behavioral control. In M. V. Butz, O. Sigaud, and P. Gérard, editors,*Anticipatory Behavior in Adaptive Learning Systems: Foundations, Theories, and Systems*, pages 44–65. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003. URL https://doi.org/10.1007/978-3-

540-45002-3_4. Y. Itaguchi. Size perception bias and Reach-to-Grasp kinematics: An exploratory study on the virtual hand with a consumer immersive Virtual-Reality device. *Frontiers in Virtual Reality*, 2, 2021. ISSN 2673-4192. URL https://www.frontiersin.org/article/10.3389/frvir.2021.712378. M. Jeannerod. *Motor Cognition: What Actions Tell the Self*. OUP Oxford, June 2006. 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. 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. J. Lohmann, A. Belardinelli, and M. V. Butz. Hands ahead in mind and motion: Active inference in peripersonal hand space. *Vision (Basel, Switzerland)*, 3(2), Apr. 2019. URL http://dx.doi.org/10.3390/vision3020015. 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 vi- sion 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. J. K. O'Regan and A. Noë. A sensorimotor account of vision and visual consciousness. *The Behavioral and brain sciences*, 24(5):939–73; discussion 973–1031, Oct. 2001. URL http://dx.doi.org/10.1017/ s0140525x01000115. 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. *The European journal of Neuroscience*, May 2020. URL http://dx.doi.org/10.1111/ejn.14783. T. D. Parsons. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in human neuroscience*, 9:660, Dec. 2015. URL http://dx.doi.org/10.3389/ fnhum.2015.00660. 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. G. Pezzulo, J. Hoffmann, and R. Falcone. Anticipation and anticipatory behavior. *Cognitive processing*, 8(2):67–70, June 2007. URL https://doi.org/10.1007/s10339-007-0173-z. G. Pezzulo, M. V. Butz, and C. Castelfranchi. The anticipatory approach: Definitions and taxonomies. In G. Pezzulo, M. V. Butz, C. Castelfranchi, and R. Falcone, editors, *The Challenge of Anticipation: A Unifying Framework for the Analysis and Design of Artificial Cognitive Systems*, pages 23–43. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. URL https://doi.org/10.1007/978-3-540-87702-8_2. G. Pezzulo, M. Zorzi, and M. Corbetta. The secret life of predictive brains: what's spontaneous activity for? *Trends in cognitive sciences*, 2021. URL https://www.sciencedirect.com/science/article/pii/ S1364661321001285. D. A. Rosenbaum, C. M. van Heugten, and G. E. Caldwell. From cognition to biomechanics and back: the end- state comfort effect and the middle-is-faster effect.*Acta psychologica*, 94(1):59–85, Oct. 1996. URL http: //dx.doi.org/10.1016/0001-6918(95)00062-3. D. J. Schad, S. Vasishth, S. Hohenstein, and R. Kliegl. How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, 110:104038, 2020. ISSN 0749-596X. URL https: //doi.org/10.1016/j.jml.2019.104038. 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. G. N. Wilkinson and C. E. Rogers. Symbolic description of factorial models for analysis of variance. *Journal of the Royal Statistical Society. Series C, Applied statistics*, 22(3):392, 1973. URL https://www.jstor.org/stable/10.2307/2346786?origin=crossref.

Figures

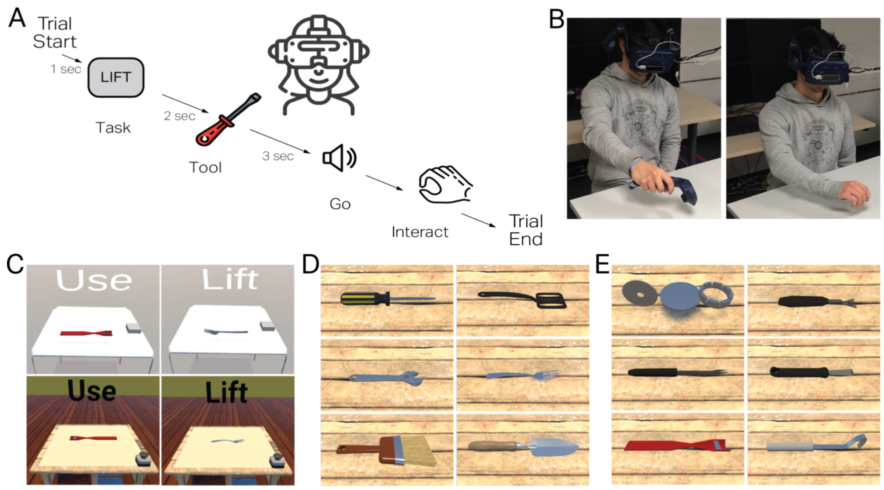


Figure 1: Experimental Task. In two virtual environments participants interacted with tools in two ways (LIFT, USE). The tools were categorized based on familiarity (FAMILIAR, UNFAMILIAR) and presented to the participants in two orientations (HANDLE LEFT, HANDLE RIGHT). The two virtual environments differed based on the mode of interaction and perceived realism, wherein in one experiment, subjects' hand movements were rendered virtually using the HTC-VIVE controllers. In the other experiment, the hands were rendered using LeapMotion, allowing finer hand and finger movements. **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 with the tool handles presented in two different orientations. BOTTOM panels illustrate the three different conditions in a more realistic environment with LeapMotion as the interaction method. **Panel D** Familiar tools, from top-left: screwdriver, spatula, wrench, fork, paintbrush, trowel. **Panel E**Unfamiliar tools, from top-left: spoke-wrench, palette knife, daisy grubber, lemon zester, flower cutter, fish scaler.

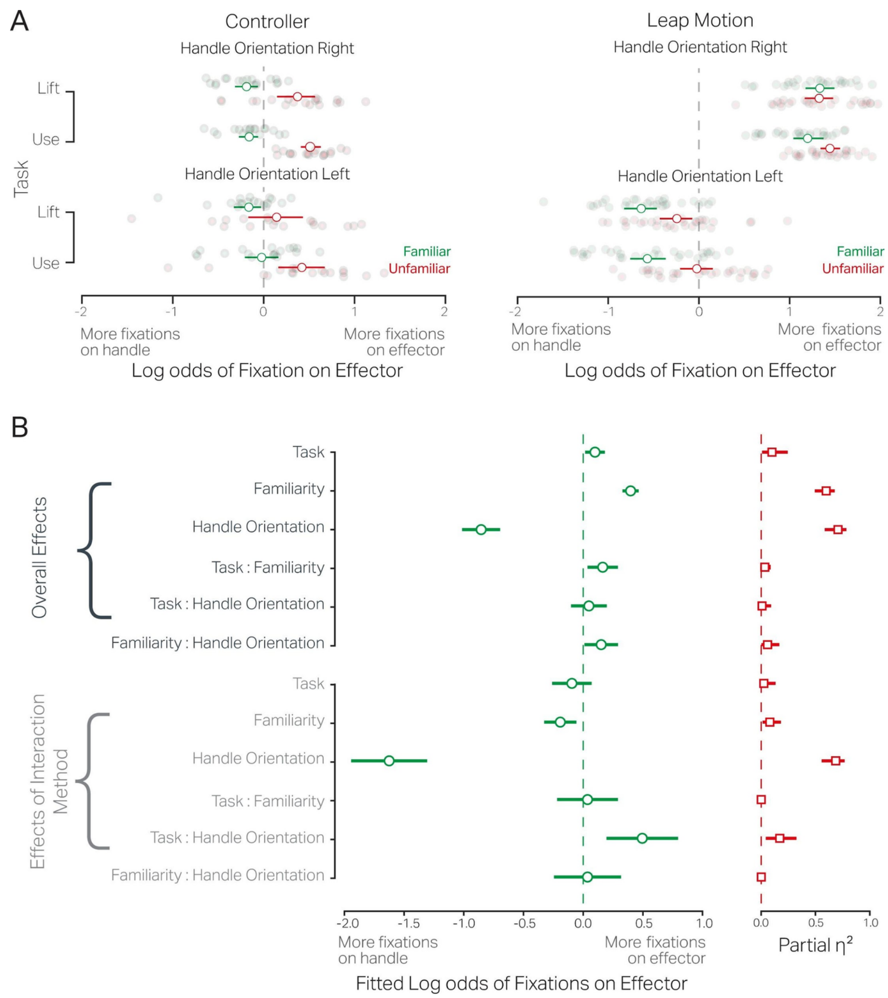


Figure 2: Effect of task, tool familiarity, handle orientation, interaction method on log-odds of fixations . **A) top-left: shows the log-odds of fixation on effector vs. handle in the controller study when the tool handle is oriented to the right. The log odds on fixations are higher on the effector 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 incongruent with the subjects' handedness. The plot shows that the orientation of the tool does not significantly 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 incongruent with the subjects' handedness. The plot shows that the orientation of the tool results in significant log-odds of fixations over the handle in the LIFT task, while in the USE task and with unfamiliar tools (red), significantly more fixations were on the effector. B.) The regression coefficients (green) and their 95%confidence and associated effect size (red) from the fitted linear mixed model. The regression coefficients that do not include null in the 95% confidence intervals are significant.**

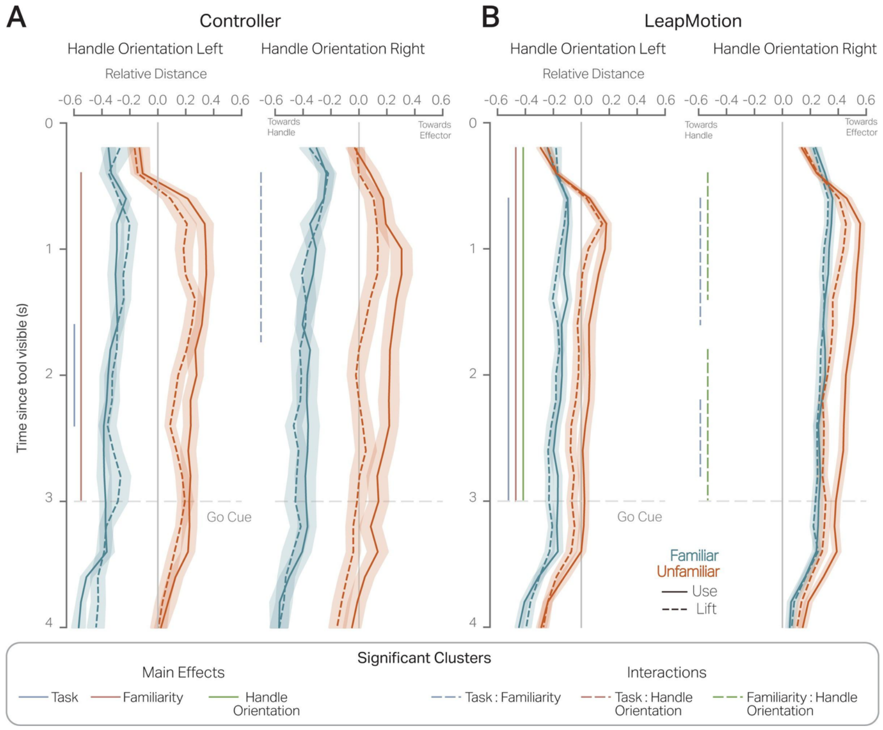
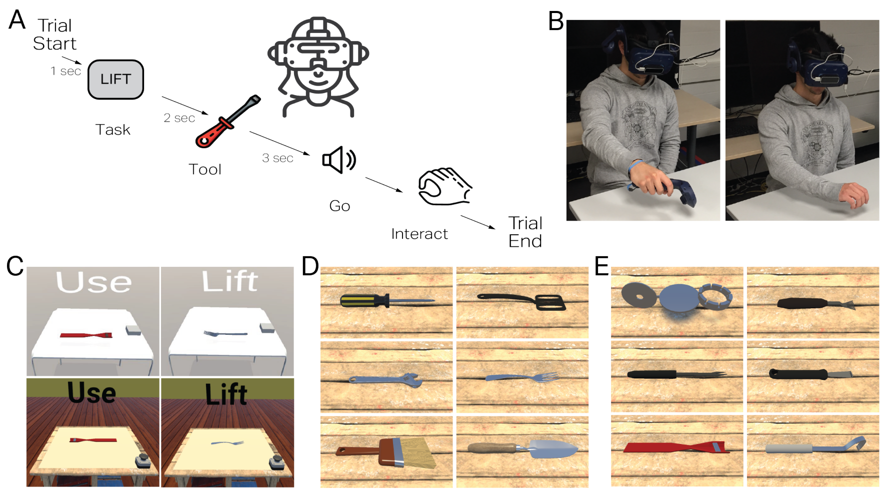
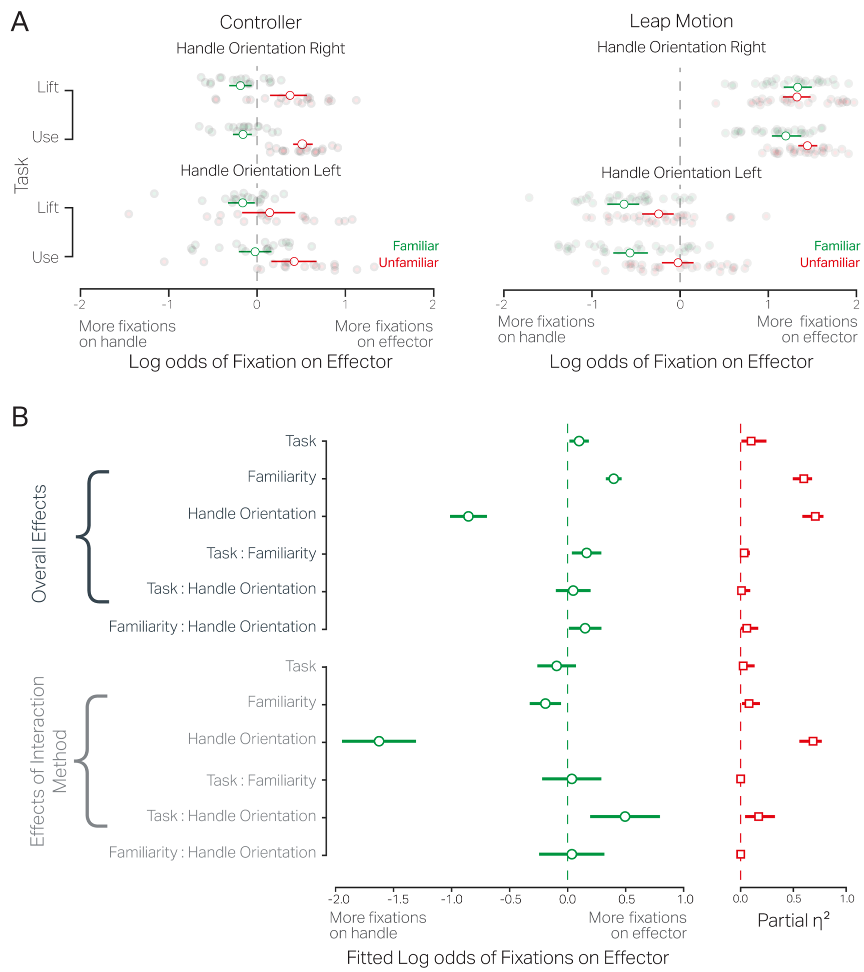
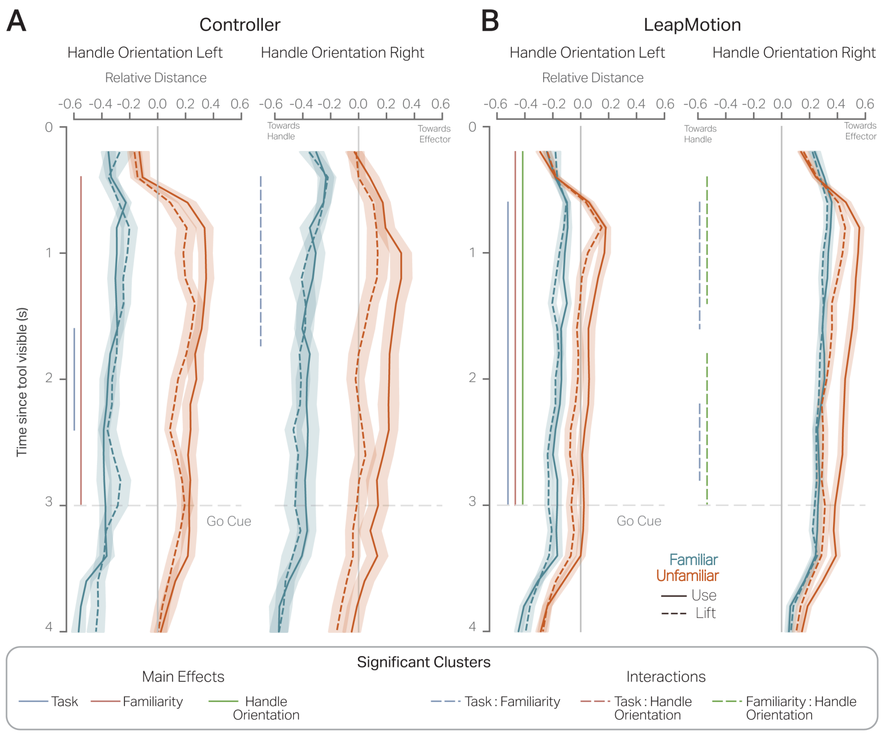
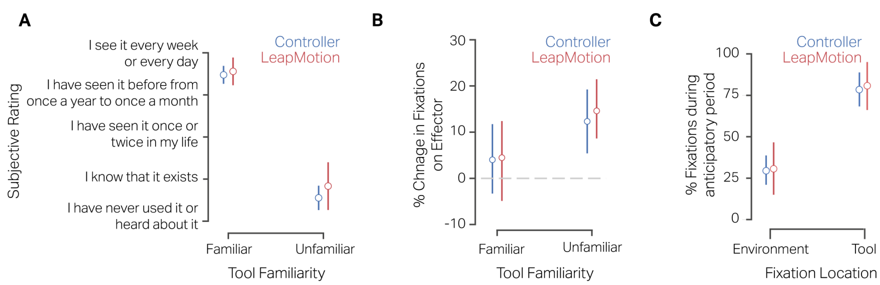


Figure 3: Eccentricity of fixations on the tool models. The negative values of the abscissa correspond to fixations towards the handle, the positive values refer to fixations towards the tool effector, and zero represents the center of the tool. The ordinate axis refers to the time elapsed since the 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 trace 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. **A** Eccentricity of fixations from the tool center in and the two handle orientations. **B.** Eccentricity of fixations from the tool center in experiment-II and the two handle orientations.**Figure 4: A. Participants' subjective rating of tool familiarity. Participants provided their subjective rating of familiarity with the 12 tool stimuli on a 5- point Likert-like scale. The circles correspond to the mean rating for each tool category, and error bars represent the standard deviation across subjects. B. Percentage change in fixations on the effector for the two different experiments. The circles correspond to the mean percentage change in across participants, and the error bars represent the standard error or mean. C. Percentage of fixations allocated to the environment vs. the tool during the 3s viewing period. The circles correspond to the mean percentage of fixations across participants, and the error denotes the standard deviation.**


A
Trial Start
1 sec
LIFT
Task
2 sec
Tool
3 sec
Go
Interact
Trial End
B
C
Use
Lift
Use
Lift
D
E


A
Controller
Handle Orientation Right
Handle Orientation Left
Leap Motion
Handle Orientation Right
Handle Orientation Left
Task
Lift
Use
Lift
Use
Familiar
Unfamiliar
-2
-1
0
1
2
More fixations on handle
More fixations on effector
Log odds of Fixation on Effector
B
Overall Effects
Effects of Interaction Method
Task
Familiarity
Handle Orientation
Task : Familiarity
Task : Handle Orientation
Familiarity : Handle Orientation
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
More fixations on handle
More fixations on effector
Fitted Log odds of Fixations on Effector
Partial η²


A
Controller
B
LeapMotion
Handle Orientation Left
Handle Orientation Right
Relative Distance
Towards Handle
Towards Effector
Time since tool visible (s)
Go Cue
Familiar
Unfamiliar
Use
Lift
Significant Clusters
Main Effects
Task
Familiarity
Handle Orientation
Interactions
Task : Familiarity
Task : Handle Orientation
Familiarity : Handle Orientation


A
B
C
Controller
LeapMotion
Subjective Rating
I see it every week or every day
I have seen it before from once a year to once a month
I have seen it once or twice in my life
I know that it exists
I have never used it or heard about it
Familiar
Unfamiliar
Tool Familiarity
% Chnage in Fixations on Effector
% Fixations during anticipatory period
Environment
Tool
Fixation Location