JUST-IN-TIME: GAZE GUIDANCE BEHAVIOR WHILE ACTION PLANNING AND EXECUTION IN VR

A PREPRINT

Ashima Keshava
University of Osnabrück
Germany
akeshava@uos.de

Farbod Nosrat Nezami
University of Osnabrück
Germany

Henri Neumann
SALT AND PEPPER
Software GmbH & Co.KG.
Germany

Krzysztof Izdebski
SALT AND PEPPER
Software GmbH & Co.KG.
Germany

Thomas Schüler
SALT AND PEPPER
Software GmbH & Co.KG.
Germany

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

January 29, 2021

ABSTRACT

Eye movements in the natural environment have primarily been studied for over-learned everyday activities such as tea-making, sandwich making, driving that have a fixed sequence of actions associated with them. These studies indicate a just-in-time strategy of fixations i.e. the fixation provides the information for a particular action immediately precedes that action. However, it is unclear if this strategy is also in play when the task is novel and a sequence of actions must be planned in the moment. To study attention mechanisms in a novel task in a natural environment, we recorded gaze and body movement data in a virtual environment while subjects performed a sorting task where they sorted objects based on object features on a life-size shelf. To study the action planning and execution related gaze guidance behavior we also controlled the complexity of the sorting task by introducing EASY and HARD tasks. We show that subjects are close to optimal while performing EASY trials and are more sub-optimal while performing HARD tasks. Based on the scan-paths as well as latency of first fixations on the task-related ROIs during action planning and execution we show that subjects use a just-in-time strategy of fixating on the task relevant objects. From our findings, we can conclude that subjects use the just-in-time strategy in a way that sacrifices optimality by offloading cognitive task demands on the environment. These findings also lend further support to the embodied cognitive framework of cognitive processing in natural environments.

Keywords top-down attention, embodied cognition, gaze guided planning, virtual reality

1 Introduction

Humans actively use vision during everyday activities to gather and refine information about the environment.Since the seminal works of Yarbus (1967) and Buswell (1935) there has been consistent evidence that eye movements depend on the viewing task the observer is performing. This is further supplemented by studies that have revealed that overt attention during natural viewing conditions is driven by scene semantics i.e. meaning and not by salience Henderson and Hayes (2017). Kollmorgen et al. (2010) have also shown that spatial viewing biases and task dependent features contribute highly to attention in a visual scene. We have growing evidence that eye movements are driven by top-down factors of task relevance and less by bottom-up factors of salience in the environment.

While studying eye movements profiles on static images in a controlled lab environment has offered illuminating insights into task-relevant modulation of attention, mobile subjects can give us a richer picture of cognitive processing in more naturalistic settings. In a pragmatic turn in cognitive science, there is a greater push towards incorporating the study of cognitive processes while interacting with the external world Parada and Rossi (2020). Moreover, Engel et al. (2013) proposed that cognition encompasses the body and in turn, bodily action can be used to infer cognition. To this effect, understanding the control of eye movements in natural, real-life situations requires a mobile setup that allows for a subject to be recorded in tandem with volitional actions in a controlled yet unconstrained environment. In recent years, virtual reality (VR) and mobile sensing has offered great opportunity to create controlled, natural environments where subjects' eye and body movements can be measured reliably along with their interactions with the environment (Keshava et al., 2020; Clay et al., 2019; Mann et al., 2019). Experiments in virtual environments have grown popular in recent years and have shown promise towards studying cognition in naturalistic and controlled environments.

Seminal studies have already investigated eye movement behavior in natural environment with fully mobile participants. For example, eye movements have been studied under a plethora of natural condition while walking (Matthis et al., 2018), tea-making (Land et al., 1999) , sandwich making (Hayhoe et al., 2003) , driving (Mars and Navarro, 2012; Sullivan et al., 2012), hand-washing (Pelz and Canosa, 2001) , hitting a ball (Land and McLeod, 2000). For these studies head-mounted eye-tracking devices were used as participants performed these tasks that allowed for fully mobile interaction with the world. Although the head mounted camera and eye-tracker record the subjects' ego-centric view of the environment, there are still some deficiencies in precisely recording the changes in the environment simultaneously with bodily interactions in a unified space. Here, VR based studies can provide a unique solution to monitor time-resolved changes in the environment and the participant.

Experiments in naturalistic settings have revealed several distinct functions of the eye movements during everyday tasks. In the pioneering studies of Land et al. (1999) and Hayhoe et al. (2003), subjects performed everyday activities of tea-making and sandwich-making, respectively. These studies required a sequence of actions which involved manipulating objects one at a time to achieve the goal. Both studies showed that nearly all the fixations were task-related. Furthermore, there was a systematic relative timing of visual fixation on object and manipulation where fixations were made to target objects about 600ms before manipulation. More importantly, Ballard et al. (1995) proposed a *"just-in-time"* strategy that universally applies to this relative timing of fixations and actions. In other words, they state that fixations that provide information for a particular action immediately precede that action and is crucial for a fast and economical execution of the task.

Land and Hayhoe (2001) posit that fixations can be broadly categorized into four functional groups; 'locating' fixations which retrieve visual information; 'directing' fixations which acquire the position information of an object and accompany a manipulation action and facilitate reaching movements; 'guiding' fixations which alternate between two objects being manipulated e.g. knife, bread, and butter; and 'checking' fixations which monitor where the task-constraints in the scene have been met. These findings have also been corroborated by Pelz and Canosa (2001); Mennie et al. (2007) .

Pelz and Canosa (2001) showed similar just-in-time strategy of gaze allocation while performing a hand-washing task. They also reported a small number of fixations 5% that did not serve the immediate sub-task but rather provided information that would be needed for a future action. The authors hypothesize these "look-ahead" fixations provide a mechanism to stabilize the visual input stream that result from a sequence of actions, facilitate task-switching, and reduce conscious effort required to complete the actions in a sequence. Hence, look-ahead fixations can be explained as a perceptual strategy to ease the cognitive load attending to complex tasks in the real world. In sum, the wide-ranging functions of eye movements are well documented in natural and routine tasks.

Land and Hayhoe (2001) have also proposed a framework that outlines that flow of visual and motor control during task execution as shown in Figure 1. The process summarizes the various operations that *must* occur during an 'object-related action' i.e., individual actions performed on separate objects to achieve the desired goal. As described in Land (2006) each schema " specifies the object to be dealt with, the action to be performed on it, and the monitoring required to establish that the action has been satisfactorily completed." Further, the gaze control samples the information about the location and identity of the object and directs the hands to it. Subsequently, the motor control system of the arms and hands implement the desired actions. Here, vision provides the information of where to direct the body, which state to monitor, and determine when the action must be terminated. Taken together, a 'script' of instructions is sequentially implemented where the eye movements earmark the task-relevant locations in the environment that demand attentional resources for at action.

The common theme in the above studies is that the tasks (tea-making, sandwich-making, hand-washing) have an organized task structure. These tasks involve specific object-related actions such as picking up the knife, picking up the teapot, etc. and have a predefined 'script' for the execution of the tasks. The studies, therefore, study eye movements that are under strict control of a task sequence. Moreover, these tasks are over-learned and over-generalized as they are part of a habitual action repertoire for an adult human being. As discussed by Land (2006) these low level schemas defined above are likely not executed under deliberate conscious control. This distinction corresponds to James (2007)'s

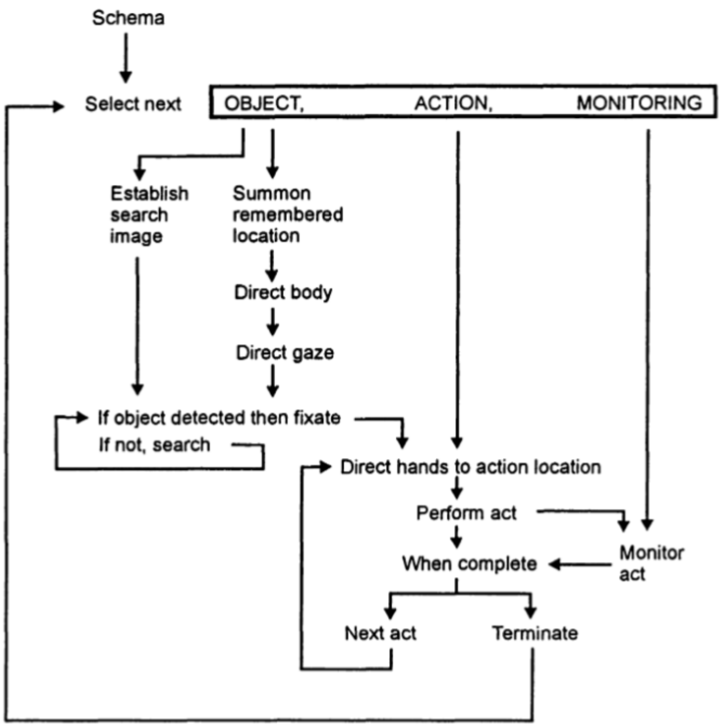


Figure 1: Schematic of motor and gaze control systems during performance of natural tasks. From Land and Hayhoe (2001)

distinction between "ideo-motor" and "willed" acts. As James described, ideo-motor actions correspond to movements where we are "aware of nothing between the conception and execution" of the said action. In contrast, the willed actions require "an additional conscious element in the shape of a fiat, mandate, or expressed consent." Hence, it is unclear whether these low-level schemas of gaze control operate similarly for deliberate actions where an internal task 'script' is not already known.

Norman and Shallice (1986) proposed a theoretical framework for the components of attentional mechanisms that govern deliberate/planned actions. In comparison to the low-level schema proposed by Land & Hayhoe which can account for routine, well-learned tasks, the Norman and Shallice (1986) model suggests another supervisory module that selects a schema to implement. In well-learned tasks, a schema is triggered automatically without conscious control. However, when a task is fairly complex and requires planning, multiple low-level schemas might compete for resources at the same time and require *contention scheduling*. For example, contentions can arise on whether to monitor the current action with respect to previous actions or future planned actions to fulfill the task-relevant goals. Such a scheduling mechanism is then required to provide conflict resolution for potentially relevant schemas either by inhibition or activation. Additionally, the model proposes a Supervisory Attentional System (SAS) that modulates the selection of a schema. The model suggests that attentional resources are deployed only at the specific points in a task where schema selection is required. Taken together, the model predicts that a failure of the supervisory control can lead to an instability of attention and heightened distraction.

In the present study, we investigate the mechanisms of allocation of attention while performing a novel task in a naturalistic environment. We created two types of tasks that varied in complexity and required performing a sequence of actions to accomplish the cued goal. We asked subjects to sort objects on a life-size shelf based on the object features. The complexity of the task depended on sorting based on one object feature or both. We designed the tasks to be novel in a way that subjects had to plan their action sequences on-the-fly and in absence of a pre-defined action "script". We concurrently measured the eye and body movements while subjects performed the tasks.

Based on the models proposed by Land and Hayhoe (2001) and Norman and Shallice (1986), we predicted four behavioral outcomes:

1. if the actions are deliberately and optimally planned, subjects would exhibit targeted gaze guidance towards task-relevant objects. Conversely, eye movement would be more random when actions are produced randomly and without deliberation.

2. optimal planning of action would require gaze allocation to target locations of previous actions to monitor task requirements are validated by current actions.
3. with optimal selection of relevant action schemas, subjects would allocate gaze to task-locations relevant to the actions in the future action sequence.
4. given optimal gaze allocation towards current task-relevant locations would not be just-in-time i.e. fixations on the target objects would immediately precede an action.

Overall, these predictions generalize the flow of low-level schemas proposed by Land and Hayhoe (2001) by taking into account a novel task which forces deliberate action planning.

2 Methods

2.1 Participants

A total of 60 participants (39 females, mean age = 23.9 ± 4.6 years) 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. They either received a monetary reward of €7.50 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. The Ethics Committee of the University of Osnabrück approved the study.

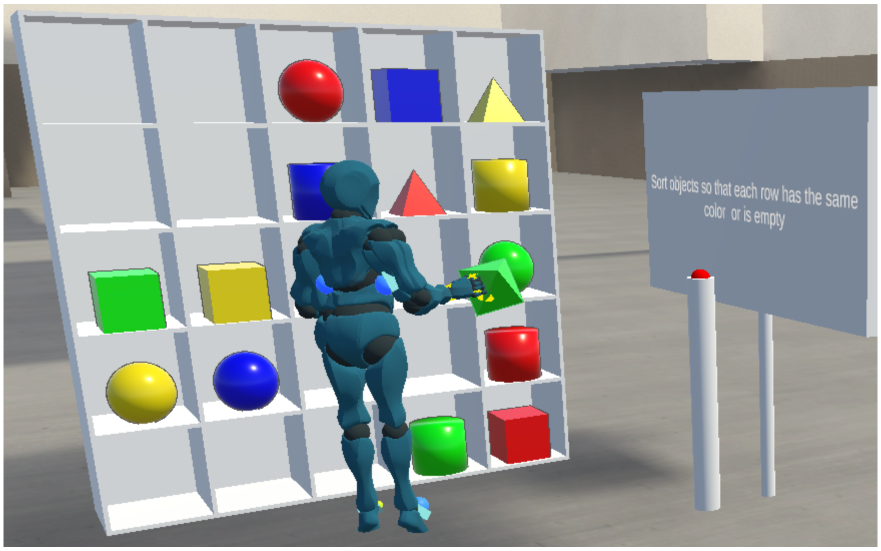

Sort objects so that each row has the same color or is empty

Figure 2: Experimental Task. In a virtual environment 60 participants sorted 16 different objects based on 2 features color or shape while we measured their eye and body movements. The objects were randomly presented on a 5x5 shelf at the beginning of each trial and were cued to sort objects by shape and/or color. Trials where objects were sorted based on just one object feature (color or shape) were categorized as EASY trials. Conversely, in the trials where sorting was based on both features (color and shape) were categorized as HARD trials. All participants performed 24 trials in total (16 easy trials and 8 hard trials) with no time limit.

2.2 Apparatus & Procedure

For the experiment, we used an HTC Vive Pro Eye head-mounted display (HMD)(110° field of view, 90Hz, resolution 1080 x 1200 px per eye) with a built-in Tobii eye-tracker [1]. Participants used an HTC Vive controller to manipulate the objects during the experiment with their right hand. The HTC Vive Lighthouse tracking system provided positional and rotational tracking and was calibrated for 4m x 4m space. For calibration of the gaze parameters, we used 5-point calibration function provided by the SRanipal SDK. To make sure the calibration error was less than 1°, we performed a 5-point validation after each calibration. Due to the study design, which allowed a lot of natural body movements, the eye tracker was calibrated repeatedly during the experiment after every 3 trials. Furthermore, subjects were fitted with HTC Vive trackers on both ankles, both elbows and, one on the midriff.The body trackers were also calibrated

[1] https://www.vive.com/eu/product/vive-pro-eye/overview/

subsequently to give a reliable pose estimation using inverse kinematics of the subject in the virtual environment. We designed the experiment using the Unity3D 2018.x.x (version) and SteamVR game engine and controlled the eye-tracking data recording using SRanipal SDK v0.7.2.1.

The experimental setup consisted of 16 different objects placed on a shelf of 5x5 grid. The objects were differentiated based on two features: color and shape. We used four high contrast colors (red, blue, green and yellow) and four 3D shapes (cube, sphere, pyramid and cylinder). The objects had an average height of 20cm and width of 20cm. The shelf was designed with a height and width of 2m with 5 rows and columns of equal height, width and, depth. Participants were presented with a display board on the right side of the shelf where the trial instructions were displayed. Subjects were also presented with a red buzzer that they could use to end the trial once they finished the task.

2.3 Experimental Task

Subjects performed two practice trials where they familiarized themselves with handling the VR controller and the general aspects of the setup. In these practice trials they were free to explore the virtual environment and handle the objects.

After the practice trials, subjects were asked to sort object based on the one and/or two features of the object. There were two types of trials: EASY and HARD. Subjects were not limited by time to complete the task. Each subject performed 24 trials with each trial type (as listed below) randomly presented twice throughout the experiment. The EASY trials instructions were as follows:

1. Sort objects so that each row has the same shape or is empty
2. Sort objects so that each row has all unique shapes or is empty
3. Sort objects so that each row has the same color or is empty
4. Sort objects so that each row has all unique colors or is empty
5. Sort objects so that each column has the same shape or is empty
6. Sort objects so that each column has all unique shapes or is empty
7. Sort objects so that each column has the same color or is empty
8. Sort objects so that each column has all unique colors or is empty

The HARD trials instructions were as follows:

1. Sort objects so that each row has all the unique colors and all the unique shapes once
2. Sort objects so that each column has all the unique colors and all the unique shapes once
3. Sort objects so that each row and column has each of the four colors once.
4. Sort objects so that each row and column has each of the four shapes once.

2.4 Data pre-processing

2.4.1 Gaze Data

As a first step, using eye-in-head 3d gaze direction vector for the cyclopean eye we calculated the gaze angles 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 direction vector of each sample is represented in (x, y, z) coordinates as 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:

$$\theta_\text{h} = \frac{180}{\pi} * \arctan \frac{x}{z} \tag{1}$$

$$\theta_\text{v} = \frac{180}{\pi} * \arctan \frac{y}{z} \tag{2}$$

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:

$$\omega_\text{h} = \Delta\theta_\text{h} / \Delta t \tag{3}$$

$$\omega_{\mathrm{v}} = \Delta\theta_{\mathrm{v}}/\Delta t \tag{4}$$

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

$$\omega = \sqrt{\omega_h^2 + \omega_v^2} \tag{5}$$

To filter the samples where gaze was relatively stable, we used an adaptive threshold method for saccade detection described by Voloh et al. (2019). The schematic of the algorithm used for saccade detection is shown in figure 3. After this, we calculated the duration of the fixations and removed those fixation samples that had a duration less than 50ms.

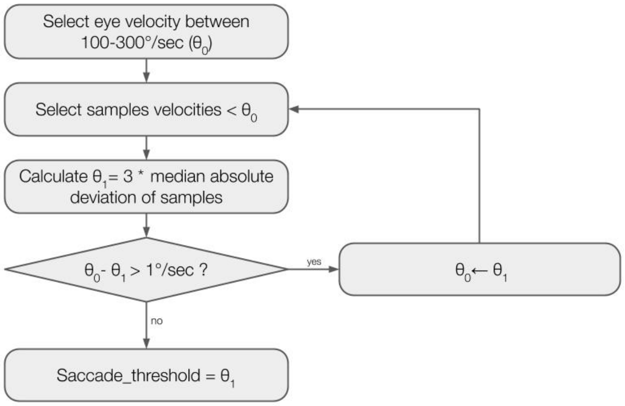

Select eye velocity between 100-300°/sec (θ_0)
Select samples velocities < θ_0
Calculate θ_1= 3 * median absolute deviation of samples
θ_0- θ_1 > 1°/sec ?
yes
$\theta_0 \leftarrow \theta_1$
no
Saccade_threshold = θ_1

Figure 3: Saccade detection algorithm. As the participants were completely mobile during the experiment, we used an adaptive method to determine the saccade velocity threshold. We selected an initial saccade velocity θ_0 of 200°/sec. All eye movement samples with angular velocity less than θ_0 were used to compute a new threshold θ_1 using 3 times the median absolute deviation of the selected samples. If the difference between θ_0 and θ_1 was less than or equal to 1°/sec θ_1 was selected as the saccade threshold, else algorithm was repeated with θ_1 as the new filtering threshold. The algorithm was repeated until the difference between the θ_0 and θ_1 was less than or equal to 1°/sec.

2.4.2 Hand controller data

Subjects used the trigger button of the HTC vive controller to virtually grasp the objects on the shelf and displace them to other locations. In the data, the trigger was recorded as a boolean which was set to TRUE when a grasp was initiated and was reset to FALSE when the grasp ended. Using the position of the controller in the world space, we determined the locations from the shelf where a grasp was initiated and ended. Next, we removed grasping periods where the beginning and final locations of the objects on the shelf were the same. We also removed trials where the controller data was showed implausible locations in the world space. These faulty data can be attributed to loss of tracking during the experiment.

2.5 Data Analysis

In order to study the function of eye movements for both action planning and execution, we divided each trial into 2 types of epochs. The action execution epoch spanned the time from start of object displacement to the end. The action planning epochs started from end of previous object displacement to start of current object displacement. The schematic of this epoch creation is illustrated in figure 4. This division of time within each trial into separate epochs allows us to parse the role of overt eye movements in planning and execution of object related actions separately.

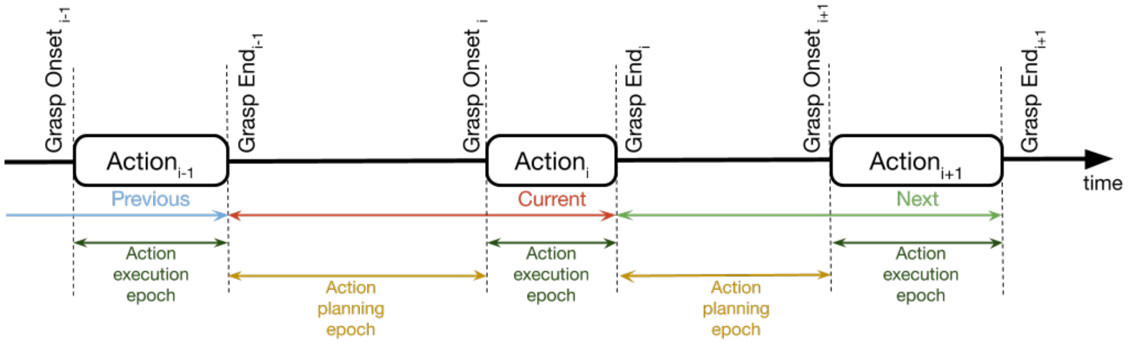


Figure 4: Action execution and planning epochs. In order to study the function of eye movements we divided the trials into action execution and action planning epochs. The action execution epochs start from grasp onset till grasp end for each object displacement, whereas the action planning epochs start from grasp end of previous object displacement and grasp onset of current object displacement.

Given the action planning and execution epochs, we examine the spatial and temporal characteristics of eye movements while performing the sorting tasks. We divided the object and shelf locations into 7 regions-of-interest (ROIs) comprising of previous, current and, next target object and target shelf. More specifically, the previous target object (prev_TO) refers to the object that was handled in the previous action epoch, and previous target shelf (prev_TS) as the shelf where the previous target object was placed. Similarly, the current target object (current_TO) refers to the object that is picked up and placed on the target shelf (current_TS) in the current epoch and the next target object (next_TO) and next target shelf (next_TS) in the immediately following epoch. All other regions which did not conform to the above 6 ROIs are categorized as 'other' and not relevant to the action sequence. As we need at least 3 object related actions within a trial to form the ROIs for the action planning and action execution epochs, we removed trials where subjects made fewer than three object displacements. In this format, we could parse the sequence of eye movements on the seven ROIs that are relevant for planning and execution of the object related actions.

To compute the scan paths within the action planning and execution epochs we created transition matrices that show the origin and destination locations of the fixations on the 7 ROIs. We used the steps described by Hooge and Camps (2013) to first create the scan paths and then the transition matrices. We computed the transition matrices summarizing gaze transitions from and to the 7 ROIs from the action planning and execution epochs within each trial. Using the transition matrices, we calculated the net and total transitions from and to each ROI. For every transition matrix 'A' per trial, net and total transition are defined as follows:

$$A_{net} = A - A^T \tag{6}$$

$$A_{total} = A + A^T \tag{7}$$

As discussed in Hooge and Camps (2013), if subjects make equal number of transitions between all ROIs, we can expect no transitions in the net transition matrix and can surmise that the gaze was allocated more randomly. Conversely, with strong gaze guidance we would expect more net transitions. Hence, using the net and total transitions per trial, we then calculated the relative net transitions denoted as F-value per trial as:

$$F = \frac{\sum A_{net}}{\sum A_{total}} \tag{8}$$

Further, we also calculated the time required to first fixation on the 7 ROIs using the median time to first fixation on each ROI in a given trial. This method was used by Montfoort et al. (2007) and further applied by Hooge and Camps (2013) to capture the gaze attraction power of ROIs. As defined by Montfoort et al. (2007) and Hooge and Camps (2013), we call this estimator T50 i.e. time to first fixation on the ROIs for 50% of the actions epochs.

2.5.1 Linear Mixed Effects Model

After cleaning the data set, we were left with data from 48 subjects and a total of 813 trials, with 17 trials per subject on average. Using these data, we modelled the linear relationship of the relative net transitions (F-value) dependent on trial type (EASY, HARD), epoch type (planning, execution) and number of object displacements and their interactions. All within-subject effects were modeled using random slopes grouped by subject and a random intercept for the subjects. The categorical variables trial_type and epoch_type were effect coded, so that the model coefficients could be interpreted as main effects. The object_displacement variable which pertained to the number of object displacements in the trial were coded as a continuous numeric variable. The full model is defined using the Wilkinson notation as follows:

$$F - value \sim 1 + trial\_type * epoch\_type * object\_displacements \tag{9}$$
$$+(1 + trial\_type * epoch\_type * object\_displacements|Subject) \tag{10}$$

We fit the model 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 bobyqa optimizer to find the best fit using 20000 interactions. As the maximal model did not converge successfully, we removed the random effects terms one by one and performed model comparison using the Bayesian Information Criterion (BIC). As we have a small number of trials per subject, we chose BIC which selects the model based on both the sample size and number of parameters used by the model. Given the minimum BIC, the following model was selected:

$$F - value \sim 1 + trial\_type * epoch\_type * object\_displacements \tag{11}$$
$$+(1|Subject) \tag{12}$$

3 Results

Our experiment measured the eye and body movements as subjects performed a sorting task in a virtual environment. The participants sorted objects based on the color and/or shape where we modulated the task complexity into EASY and HARD trials. We further divided the trials into planning and execution epochs where subjects planned the selection of the target objects to interact with and then executing the action of displacing it to target shelves, respectively. In this section, we report the behavioral differences of the subjects for the two task types (EASY, HARD), the scan-path differences in the planning and execution epochs and finally the timing of the first fixations on regions-of-interest based on the action sequences.

3.1 Task based Behavioral Differences

In the present study, the primary object related action was to repeatedly pickup objects and place them at a desired locations until they are sorted according to the sorting task. Figure 5a shows the differences in EASY and HARD trials based on the time taken to finish the sorting task. The EASY trials require sorting the objects based on a single feature, the trial duration is shorter (Mean=56.02 seconds, SD=23.308) as compared to HARD trials (Mean=132.16 seconds, SD=116.18) where subjects had to sort taking into account both features (color and shape) of the objects.

In order to compare the experimental object displacements for the two trial type, we also designed a depth-first search algorithm that computed the minimum number of displacements required to sort the objects for the 5000 random configurations of 16 objects in 25 shelf locations for both EASY and HARD trials. Figure 5b shows the comparisons in the object displacements made by the subjects and the optimal number of displacements as elicited by the search algorithm for both EASY and HARD trials. Subjects made lower number of object displacements in the EASY trials (Median=10, IQR=[8, 13]) compared to HARD trials (Median=15, IQR=[11, 20]). In the EASY trials subjects performed closely with the optimal search algorithm (Median=9, IQR=[9,10]), whereas, in the HARD trials, subject were more sub-optimal compared to the algorithm (Median=11, IQR=[9,13]).

Next, to check the propensity to pickup and drop-off objects from and to preferred locations on the shelf, we plotted the density of pickup locations and drop-off locations in Figure 6. Given the sorting tasks where subjects were presented with random initial configurations of the objects on the shelf locations, we did not expect any systematic spatial biases at play. Further, the expectation was that the subjects would move the objects randomly and not display a preference for object pickup and drop-off locations. As seen in Figure 6a at the start of the moving of the object, in the case of easy trials objects are preferentially picked up from the the rightmost column and bottom row of the shelf. Conversely, for the hard tasks the objects are picked up from all locations and there is no systematic preference for the pickup location on the shelf. Interestingly, for both EASY and HARD trials, subjects did not move the objects from the top and left-most locations of the shelf. Secondly, as seen in Figure 6b subjects show a propensity to drop the objects leaving out the right column and bottom row of the shelf for both EASY and HARD trials.

This shows that subjects systematically, displace the objects leftward and upward employing an arbitrary heuristic to complete both task types. As the objects are instantiated on the shelf randomly, an optimal strategy would not show this behavior. We can conclude from the above that subjects offset their cognitive load of optimally completing the task by

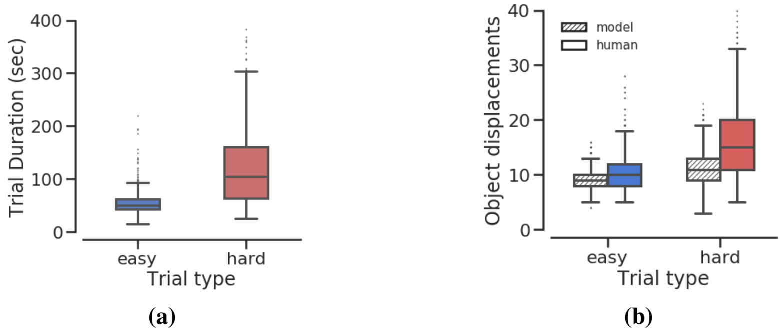


Figure 5: Behavioral differences based on the 2 tasks: (a) Trial duration of the EASY and HARD trials. The boxplots show the inter-quartile range (IQR) of the duration of the trials for the two different trial types for all trials and participants. The whiskers represent 1.5 times the IQR. The median duration for the EASY trials was 60 seconds while 110 seconds for HARD trials.(b) Distribution of number of object displacements for EASY (blue) and HARD (red) trials. The colored box plots show the inter-quartile range of the number of object displacements made by subjects per trial and per participant. The whiskers represent 1.5 times the IQR. The dashed box plots show optimal number of displacements required to sort the objects for a model computed with a depth-first search algorithm for 5000 random trial initialization for each trial type.

employing simple heuristics. Further, this behavior can also be explained as a way to reduce the physical effort and make ergonomic compensations by placing all objects at a higher ground level in order to bend the body fewer times during the task. In other words, in lieu of optimally performing the task and finishing it in a shorter time, subjects prefer to offload both cognitive and physical effort on the environment by adopting a more sub-optimal strategy.

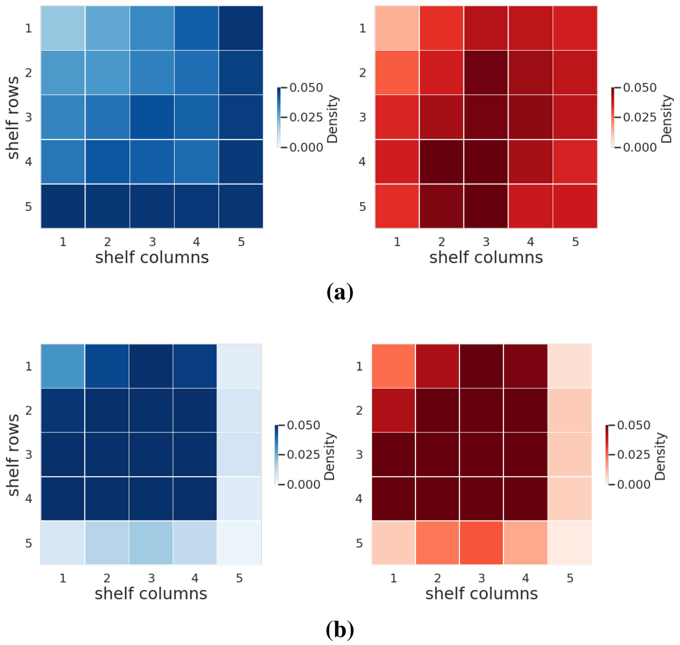


Figure 6: Object displacement behavior for 60 subjects and 16,500 displacements made over all shelf 5x5 locations with blue heatmaps showing probability density over EASY trials and red heatmaps showing density over HARD trials. (a) shows the probability density at the start of each object displacement. This indicates that subjects have a propensity to pickup objects from the rightmost column and bottom column for EASY trials (left) and conversely, in the HARD trials (right) subjects pickup objects from central locations. For both trial types, objects in the top-left corner of the shelf are not picked up. (b) shows the probability density at the end of the object displacement. This indicates that for both EASY (blue, left) and HARD (red, right) trials, subjects display a systematic propensity to place the objects every where other than the bottom row or rightmost column.

3.2 Role of Eye Movements in Action Execution & Planning

In this section, we investigate the role of eye movements specific to action planning and execution. The action of picking up an object and dropping it to a shelf location requires precise attention on the object to be grasped and then a shift of gaze to the location of the place where the object will be dropped off. The change of proportion of fixations on the task-relevant objects of interest over time would reveal the average shift of gaze before and during the action execution.

To do this, we employ seven regions-of-interest for each action of displacing an object. These regions-of-interest consist of the current target object (current_TO) that is being displaced, the current target shelf (current_TS) where the object is displaced finally, the previous target object (prev_TO) and previous target shelf (prev_TS) that were relevant in the previous action, the next target object (next_TO) and next target shelf (next_TS) that will be interacted with in the next action in the sequence, and finally all the other objects and shelves that were not relevant in the action sequence. We selected the 3s before the virtual hand touches the object (grasp onset) and 3s after and divided it into 0.25 second bins to calculate the proportion of fixations on all of the ROIs within each time bin. Hence, we can assess the dynamic changes in the proportion of fixations on the seven ROIs as the current action is planned and executed.

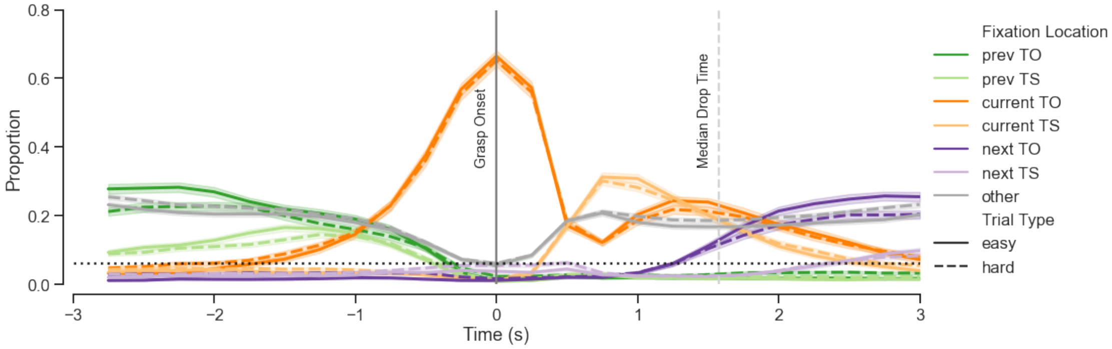
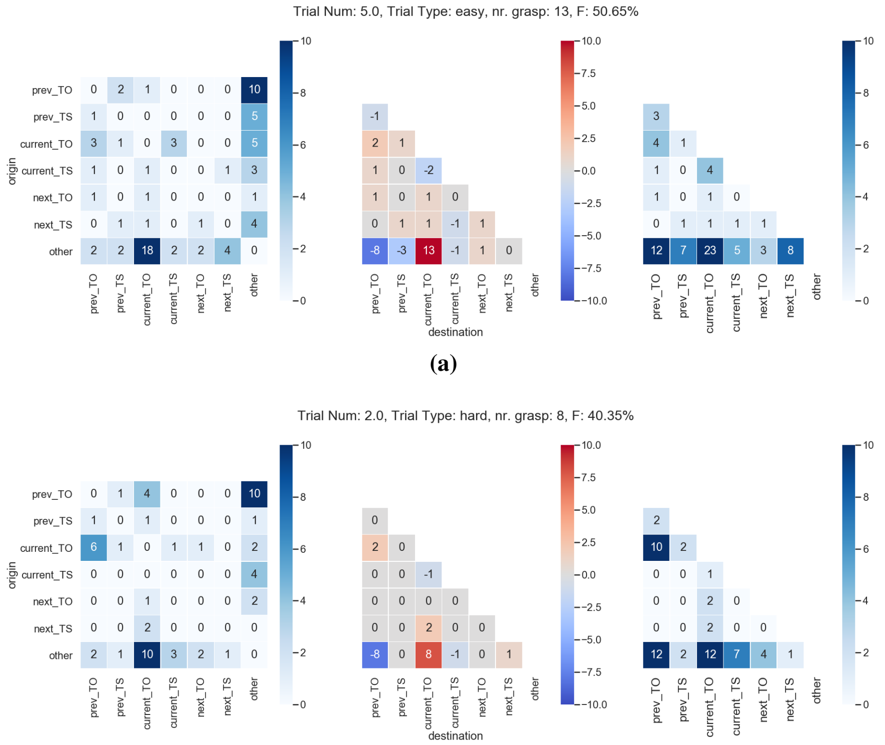


Figure 7: Time course of proportion of fixations centered on the object displacement initiation (grasp onset) at time=0 and 3 seconds before and after on the seven regions of interest and for the two trial types (EASY and HARD). The black dotted horizontal line indicate the chance level (1/16) of fixating on a target object. The shaded regions show 95% confidence interval of the mean proportion of gaze at each time-bin across all trials within a trial type. These regions-of-interest consist of the current target object (current_TO) that is being displaced, the current target shelf (current_TS) where the object is displaced finally, the previous target object (prev_TO) and previous target shelf (prev_TS) that were relevant in the previous action, the next target object (next_TO) and next target shelf (next_TS) that will be interacted with in the next action in the sequence, and finally all the other objects and shelves that were not relevant in the action sequence.

As a first step, we wanted to study the average oculomotor profiles across time over the seven ROIs and the trial types. The average profile will reveal any systematic gaze shifting strategy from previous, to current and then to next in the action sequence. Figure 7 shows the time course of proportion of fixations on the seven ROIs as described above for the two task types. We can see that proportion of fixations on the current target object are above chance level (1/16) at approximately 2 seconds before grasp onset with maximum proportion of fixations on the target object at the time of grasp onset. Before grasp onset, the average profile reveals equal proportion of fixations on the previous target object and other regions on the shelf. As the fixations on the previous target object and shelf decrease, the fixations on the current target object increase monotonically and are maximum at the grasp onset. Throughout the period before the grasp onset, gaze on the next target object and shelf remains below chance level. After the grasp is initiated, the gaze shifts from the current target object and goes to the current target shelf where the proportion of fixations on the target shelf and non-target other ROI increases monotonically and is maximum approximately 1 second after grasp onset. Moreover, the average proportion of fixations for the ’other’ ROI is consistently low during the action execution and only increase once the action is completed. Moreover, we see that before the median end of the action at 1.64 seconds after the grasp onset, the fixations are above chance level on the next target object. The presence of fixations on the next target object even before the end of the current action execution indicates, that the next action is queued in for execution shortly before ending the current action. Also, as there are no significant differences in the fixation probabilities on the ROIs across time for the two trial types, we can conclude that this oculomotor behavior is purely related to action execution and is not affected by the cognitive load of the different sorting tasks. Most importantly, we see that the fixations are made in a just-in-time manner with the gaze shifting systematically in sync with the previous, current and next action sequence one after the another. This indicates that the oculomotor system is engaged in a just-in-time manner for the sub-actions of picking up and dropping off one after the other.

The above results, illustrate the average spatial and temporal aspects of attention during action execution. However, the scanning behavior of subjects while they perform each action is "averaged out" with this technique. In order to study the scanning behavior while subjects plan and execute an action, we computed transition matrices to capture fixations to and from each of the seven ROI. The ROIs used are described in Section 2.5. With the transition matrices we wanted to capture the gaze guidance behavior of the subjects while they plan and execute the actions. The relative net-transitions within the planning and execution epochs of a trial tell us the gaze guidance behavior of the subjects during those epochs. With higher relative net transitions, we expect higher gaze guidance to the task-relevant ROIs, i.e, subjects do not necessarily search for the best action for the task. If subjects perform a search and fixate on multiple ROIs indiscriminately within an epoch, we would expect lower relative net transitions indicating a pattern of fixations related to planning the best (most optimal) actions.



(a)

(b)

Figure 8: Exemplar transition matrices for gaze switching between end of previous grasp and start of current grasp for the two trial types. The ordinate defines the origin of the gaze i.e. where the gaze was before and the abscissa defines the destination of the gaze. The diagonal contains zero as the transition is defined as a saccade from one region of interest to another. In this example there are 7 regions of interest.They are defined here as the previous target object and shelf (prev_TO, prev_TS), the current target object and shelf (current_TO, current_TS) that are manipulated in the current epoch and the next target object and shelf (next_TO, next_TS) that will be used in the next grasping epoch. All other objects and shelf locations that are not specific to the above regions of interest are defined as 'other'. The left panel shows the transition matrix, A. The middle panel, shows the net transitions (A_{NET}) to the regions of interest. The positive number in the matrix denotes transitions from source ROI to destination ROI, whereas the negative numbers reverse the direction of transition and denote the transition away from destination to source. The right panel shows the total transitions (A_{TOTAL}) made between the 7 regions of interest. The F value in the in the figure title refers to the relative number of transitions in the net transition matrix.
Panel (a) shows the transition matrix of a given trial for the gaze switching in an EASY trial with 13 grasps/object displacements. As shown in the middle panel, in the trial, 8 transitions are made from prev_TO to other, whereas 13 transitions are made from 'other' to current_TO. The right panel shows the total number of transitions made between the ROIs. This tells us the that a majority of the transitions are from 'other' sources to the regions of interest. The F value for this trial shows that 50.65% net transitions explain the total transitions.
Panel (b) shows the transition matrix for the gaze switching of a given trial in a HARD trial with 8 grasps/object displacements. As shown in the middle panel, in the trial, 8 transitions are made from prev_TO to other, and 8 transitions are made from 'other' to current_TO. And other ROI have close to net zero transitions. The right panel shows the total number of transitions made between the the ROIs. This tells us the that a majority of the transitions are from 'other' sources to the regions of interest. The F-value for this trial shows that 40.35% net transitions explain the total transitions.

Figure 8 shows the exemplar transition matrices, net transitions and total transitions for an EASY and HARD trial. The matrices capture the gaze switches from and to the 7 ROIs taking together all the action planning epochs in a trial. As explained in Section 2.5, if the subjects make equal number of transitions between all ROIs we expect no transitions in the net transitions matrix. If gaze is biased towards one or more ROIs during action execution, we would see more net transitions to or from these ROIs. We denote the strength of gaze guidance behavior for each trial using the F-value which signifies the relative number of net transitions compared to total transitions for the 7 ROIs. Hence, higher F-values mean stronger gaze guidance towards one or more ROI during the action planning epochs.

Next, we wanted to relate gaze guidance behavior with the object displacement behavior within a trial. With higher gaze guidance, we can expect attention being guided to action relevant items without a search for an optimal action. MOreover, with higher F-values, we expected more number of object displacements which would signify a just-in-time strategy and not adequate planning of object selection. In order to understand the effects of trial complexity, the epochs of planning and execution of actions and the number of object displacements on F-values, we used a linear mixed effects model as described in section 2.5.1.

Figure 9 shows the regression fit over the fixed-effects the 2 trial types, 2 epoch types and the effect of object displacements on the relative net transitions. The fixed model coefficients are shown in table 1. The regression coefficients show that there is a significant main effect of trial type where the EASY trials show higher relative net

factor	Est	SE	LL	UL	Z	Pvalue
trial type	0.124	0.024	0.076	0.171	5.061	0.000 ***
epoch type	0.081	0.024	0.034	0.128	3.389	0.001 ***
object displacements	-0.003	0.001	-0.005	-0.001	-2.611	0.009 **
trial type:epoch type	0.145	0.048	0.051	0.239	3.019	0.003 **
trial type:object displacements	-0.015	0.002	-0.020	-0.011	-6.782	0.000 ***
epoch:object displacements	0.001	0.002	-0.003	0.006	0.635	0.526
trial type:epoch type:object displacements	-0.019	0.004	-0.028	-0.010	-4.321	0.000 ***

Table 1: Model coefficients of the linear fixed effects model.

transitions than HARD trials. We also see a significant main effect of epoch type where planning epochs had higher gaze guidance than execution trials. We further see a significant effect of object displacements where F-values decreased for increasing number of object displacement. The model coefficients further show significant interactions between epoch type and trial type where we have higher F-values in planning epochs of EASY trials. We also have significant interactions of trial type and object displacements where the F-value decreases at a higher rate for HARD trials than for EASY trials. Finally, we see a significant interaction between all of the factors showing planning epochs in HARD trials had a steeper decreasing slope for the F-values.

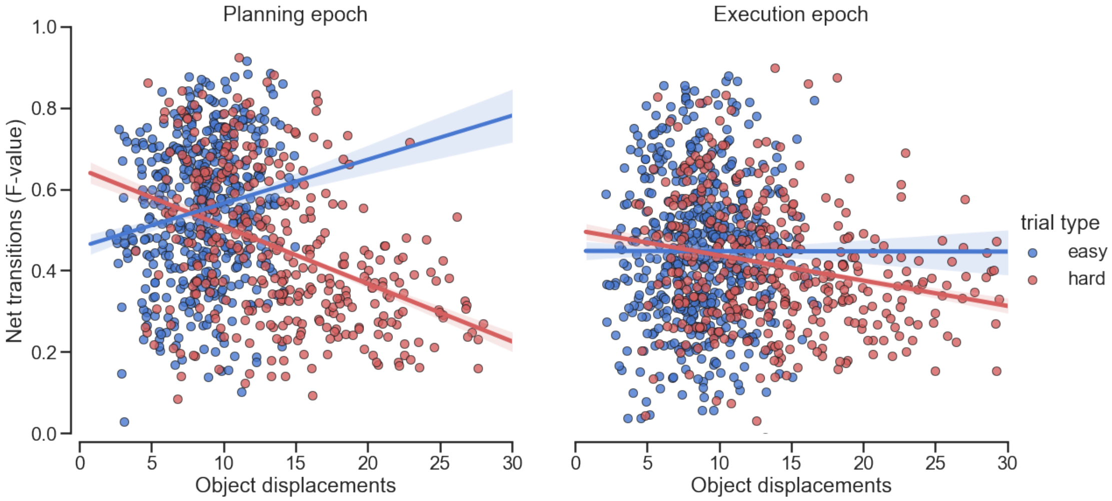


Figure 9: Regression fit over the fixed-effects of trial types, epoch types and object displacements on the relative net transitions. Figure shows a scatter plot of the relative net transition (F-value) vs. number of object displacements per trial and differentiated for the two trial types EASY(blue), HARD(red). Each point refers to one trial and it's F value vs. the number of object displacements in that trial. The lines denote the regression fit and the shaded region denotes 95% confidence interval.

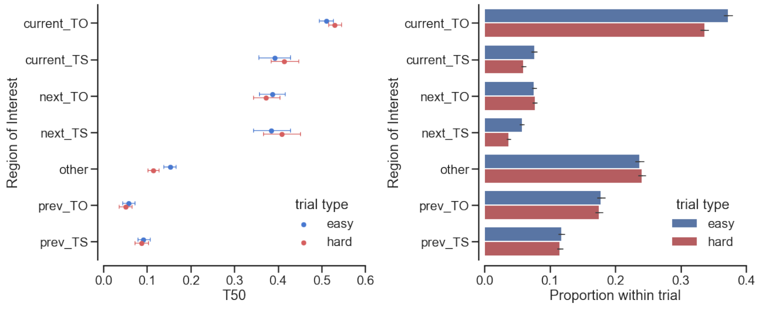
Finally, the analysis above tells us the gaze scan paths and attention guidance behavior to and from relevant ROIs. However, it does not illustrate the primacy of ROIs based on selection by gaze. Even though there are higher proportion of fixations on the current_TO and current_TS during action execution, we cannot determine which ROIs were first fixated on during the action the execution epochs. Hence, we were further interested in finding the attention attraction of power of the ROIs. In the case of planning epochs if the first fixation on current_TO is earlier than other ROIs, then the target location of the object is selected first and kept in memory while scanning other objects. If however, the first fixation on the current_TO is latest in the sequence of ROIs, it signifies a search process conducted at start of action planning epoch where the eyes search for the appropriate location to place the object, "finds" it and performs the action execution. This would signify a just-in-time strategy of fixating on ROIs as and when they are needed for an action. Similarly, in the action execution epochs, if the first fixation on current_TS is earlier than other ROIs, then the target location of the object is selected first and then other objects are scanned. If however, the first fixation on the current_TS is latest in the sequence of ROIs, it signifies a search process conducted at start of action execution epoch and once the target shelf is located the action is terminated.

As the action planning and execution epochs varied in duration, we normalized the time points by dividing them by the duration of the epoch. This way, time elapsed since start of epoch is comparable to all epochs across trials and subjects.

To estimate the time elapsed until first fixation on the ROIs, we used a measure known as T50 which estimates the median (50% quantile) time till first fixation on each of the 7 ROIs per trial. Hence, with multiple action execution epochs per trial, we estimate the T50 for each of the 7 ROI per trial. Figure 10a shows the distribution of T50 over the 7 ROIs in the action planning epochs. As the shown in the figure, there are early re-fixations on the the prev_TS and prev_TO as the gaze moves away from right after completing the previous action execution epoch followed by fixations on 'other' non-task relevant objects. In the later part of the epoch, T50 of first fixations on next_TO, next_TS and current_TS are close together. More importantly, T50 for the current_TO the shelf location where the target object is placed is last in the epoch after 50% of the epoch time has elapsed. The results above have to be accounted for given the proportion of occurrence of the ROIs within a trial. As shown in Figure 10b, current_TO and 'other' account for more than 50% of the first fixations in the trial. Less than 1/10th of the fixations in a trial are devoted to next_TO, next_TS and current_TS.

Figure 11a shows the distribution of T50 over the 7 ROIs. As the shown the figure, there are early re-fixations on the the current_TO right after the onset of the action execution epoch followed by fixations on 'other' non-task relevant objects. In the later part of the epoch, T50 of first fixations on prev_TO, prev_TS, next_TO and next_TS are close together. More importantly, T50 for the current_TS the shelf location where the target object is placed is last in the epoch after 50% of the epoch time has elapsed. The results above have to be accounted for given the proportion of occurrence of the ROIs within a trial. As shown in Figure 11b, current_TO and current_TS account for more than 50% of the first fixations in the trial, shortly followed by 'other' ROI. prev_TO, prev_TS and next_TO, next_TS comprise of less that 1/10th of the first fixations within a trial. This is an important caveat to note while assessing the T50 values for the ROIs where the T50 estimates for prev_T0, prev_TS, next_TO, and next_TS are based on small proportion of occurrences in a trial.

Taken together, this is further evidence of the a just-in-time strategy of gaze guidance during both action planning and execution epochs, where a "search" process is initiated for the current target object and shelf, with the search "ending" on the target object/shelf and immediately following an action of picking up the object or dropping it off. Most importantly, the latency of the fixation on the the current target object shows that object selection for action execution is not necessarily planned. In the action execution epochs, the latency of the T50 for the previous task related object and shelf show that a monitoring of the current action with respect to the previous action are at play, where fixations made to the prev_TO and prev_TS are made to confirm the choice of the current target shelf and might serve as look-back fixations. Further more, the latency of first fixations on the next task object close to the current target shelf indicates that in the 1/10th of the instances, gaze is used to pre-plan the next action before the end of the current action and might function as look-ahead fixations.

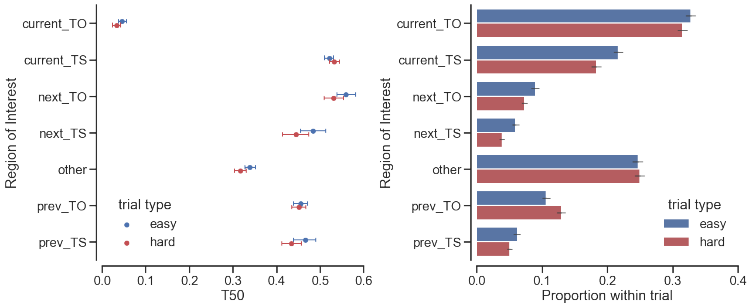


(a) (b)

Figure 10: T50 for the action planning epochs. In order to capture the latency T50 as an estimator for latency of first fixation on the 7 ROIs per trial. Panel (a) shows the mean and 95% confidence interval of T50 estimate for each trial. Blue points show the mean T50 for EASY trials and red points show the the T50 for the HARD trials. Panel (b) shows the proportion of first fixations on each of the 7 ROIs in the planning epochs within a trial. Blue bars show proportion of first fixation on given ROI for EASY trials, red bars for HARD trials. Error bars indicate 95% confidence interval.

4 Discussion and Conclusion

In the present study we investigated the allocation of gaze while action execution and action planning with varying task complexity. We report four main findings with this study. First, we found that subjects were not optimal in sorting the objects when compared to a greedy depth-first search algorithm. The data showed that subjects instead used a spatial heuristic to complete the task which resulted in more number of object displacements as compared to the model optimal especially for HARD tasks. These findings suggest that subjects did not optimally plan their actions and had to rely on ad-hoc heuristics to complete the task. Secondly, based on the scan path transitions, we observed greater net transitions



(a) (b)

Figure 11: T50 for the action execution epochs. In order to capture the latency T50 as an estimator for latency of first fixation on the 7 ROIs per trial. Panel (a) shows the mean and 95% confidence interval of T50 estimate for each trial. Blue points show the mean T50 for EASY trials and red points show the the T50 for the HARD trials. Panel (b) shows the proportion of first fixations on each of the 7 ROIs. Blue bars show proportion of first fixation on given ROI for EASY trials, red bars for HARD trials. Error bars indicate 95% confidence interval.

in the EASY trials compared to HARD trials. The lower proportion of net transitions to and from the ROIs in HARD trials is evidence for "instability" in attention and heightened distraction. Thirdly, the relative net transitions in the action planning phase of both EASY and HARD trials negatively correlated with the number of object displacements made by the subjects. Finally, the relative timing of the first fixations on the immediate task-relevant ROIs in the action planning and execution phase revealed that subjects predominantly made just-in-time fixations before executing the pick-up and drop-off actions.In sum, our findings reveal a just-in-time strategy for problem solving which is reflected in both behavior and attention allocation and is invariant to task complexity.

To assess the sorting behavior of the participants we compared their object displacement behavior with a greedy depth-first search algorithm which optimizes for the shortest path to the solution. Studies in human performance in reasoning tasks as well as combinatorial optimization problems (MacGregor and Chu, 2011) have revealed that humans solve these tasks in a self-paced manner rather than being dependent on the size of the problem. Pizlo and Li (2005) found that subjects do not perform an implicit search of the problem state space where they plan the moves without executing them, i.e. longer solution times would lead to shorter solution paths. Instead, they showed that humans break the problem down to component sub-tasks which gives rise to a low-complexity relationship between the problem size and time to solution. Further, Pizlo and Li (2005) show that instead of using implicit search, subjects use simpler heuristics to decide the next move. To this effect, while subjects in our study are sub-optimal compared to a depth-first search algorithm, humans in general are prone to use non-complex heuristics that favor limited allocation of resources in the working memory.

We further modelled the gaze guidance behavior of the participants while they performed the sorting task. Based on the regression model, we can conclude that the subjects had higher gaze guidance for the EASY trials compared to the HARD trials. These results suggest that subjects' gaze moved to the target locations smoothly without much search process in the EASY trials. This suggests that owing to the task complexity of the HARD trials, subjects made many fixations on different objects as a way to select the target objects and shelves. Moreover, we also see higher gaze guidance in the planning epochs as compared to execution epochs. The higher F-values in the planning phase suggests that subjects subjects picked up objects without doing much search, which also indicates why they showed sub-optimal behavior while sorting the task. The higher F-values in the execution epochs can be interpreted in two ways. Firstly, subjects maybe picked up the objects at whim and then searched for the best way to place the object, indicating the actual planning of the action happens after the action has already been initiated. Alternatively, the lower F-values would indicate that they moved their gaze to other objects as a way to monitor the outcome of the current task and produce the best outcome. We reject the second alternative, as subjects mostly show gaze transitions to other objects/shelf that are not relevant in the previous, current, next action sequence. If monitoring of the current action took place, subjects would have ideally made more fixations to the previous task relevant objects. Furthermore, our data also revealed that trials with greater number of object displacements showed lower F-values in the planning epochs in HARD trials, with F-values negatively correlated to the number of object displacements. Here, we argue that due to the complex nature of the HARD tasks, subjects fixated on many objects before selecting one for an action. As the subjects were more constrained in the HARD tasks they needed to search the scene for relevant objects more often. In sum, our findings show that the task complexity coupled with the situation of planning or executing the tasks had significant effects on gaze guidance behavior.

We also studied the latency of the first fixations on the task-relevant regions of interest for the trial types and action epochs. Our findings show that subjects made fixations to the task-relevant object or shelf towards the latter half of the epochs. These findings suggest that subjects predominantly used a just-in-time strategy for selecting objects, i.e. they acted on the target objects closely after fixating on them. Interestingly, in the action execution epochs, subjects make first fixations on the target objects that are relevant to the next action in the sequence. These fixations can be categorized as look-ahead fixations as have been found by Pelz et al. (ref), Mennie and Hayhoe (ref) Mars and Navarro (ref). These studies showed that look-ahead fixations have a low-probability of occurrence at about 20% of the reaching movements (ref). These fixations usually relate to anticipatory eye movements and are a task-dependent strategy to acquire information about objects for future manipulation.

Taken together, our findings are also consistent from the perspective of embodied cognition (Wilson, 2002; Ballard et al., 2013; Van der Stigchel, 2020), where: a) we off-load cognitive work on the environment; b) the environment is central to the cognitive system; b) cognition is for action; c) we off-load cognitive work on the environment; d) cognition is situated. The behavioral results show that subjects use spatial heuristics to complete the task indicating they exploit the external world to reduce the cognitive effort of selecting optimal actions. The oculomotor behavior further suggests that the just-in-time fixation strategy is primarily for action initiation. Further, Droll and Hayhoe (2007) have suggested that the just-in-time strategy to lower the cognitive cost of encoding objects in the world into the visual working memory. König et al. (2016) have further proposed eye movements reveal much about the cognitive state. In our study, the differences in the gaze guidance behavior for the different trial types and action epochs further reveal that the cognitive processes are situated in the the current context of the environment and inherently involves perception and action.

In conclusion, in the present study we investigated the oculomotor responses to novel task scenarios that varied in complexity. The main findings of the study was the just-in-time strategy of fixations employed by subjects to complete the task and that are tightly coupled with the action sequences in the task. We also show that subjects use this strategy to offset the cognitive effort of optimally planning their actions. This study extends the previous research showing the oculomotor behavior during natural tasks of tea-making, sandwich-making, hand-washing, etc., that are over-learned. Due to the abstract nature of the task, we believe our work offers a view of eye movement behavior in natural environments that generalizes across tasks.

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.

References

D. H. Ballard, M. M. Hayhoe, and J. B. Pelz. Memory representations in natural tasks. *J. Cogn. Neurosci.*, 7(1):66–80, 1995. URL http://dx.doi.org/10.1162/jocn.1995.7.1.66.

D. H. Ballard, D. Kit, C. A. Rothkopf, and B. Sullivan. A hierarchical modular architecture for embodied cognition. *Multisens Res*, 26(1-2):177–204, 2013. URL http://dx.doi.org/10.1163/22134808-00002414.

G. T. Buswell. How people look at pictures: a study of the psychology and perception in art. *Chicago University Press*, 198, 1935. URL https://psycnet.apa.org/fulltext/1935-05800-000.pdf.

V. Clay, P. König, and S. König. Eye tracking in virtual reality, 2019. URL http://dx.doi.org/10.16910/JEMR.12.1.3.

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

J. A. Droll and M. M. Hayhoe. Trade-offs between gaze and working memory use. *J. Exp. Psychol. Hum. Percept. Perform.*, 33(6):1352–1365, Dec. 2007. URL http://dx.doi.org/10.1037/0096-1523.33.6.1352.

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

M. M. Hayhoe, A. Shrivastava, R. Mruczek, and J. B. Pelz. Visual memory and motor planning in a natural task. *J. Vis.*, 3(1):49–63, 2003. URL http://dx.doi.org/10.1167/3.1.6.

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

I. Hooge and G. Camps. Scan path entropy and arrow plots: capturing scanning behavior of multiple observers. *Front. Psychol.*, 4:996, Dec. 2013. URL http://dx.doi.org/10.3389/fpsyg.2013.00996.

W. James. *The principles of psychology*, volume 1. Cosimo, Inc., 2007.

A. Keshava, A. Aumeistere, K. Izdebski, and P. König. 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.

S. Kollmorgen, N. Nortmann, S. Schröder, and P. König. Influence of low-level stimulus features, task dependent factors, and spatial biases on overt visual attention. *PLoS Comput. Biol.*, 6(5):e1000791, May 2010. URL http://dx.doi.org/10.1371/journal.pcbi.1000791.

P. König, N. Wilming, T. C. Kietzmann, J. P. Ossandón, S. Onat, B. V. Ehinger, R. R. Gameiro, and K. Kaspar. Eye movements as a window to cognitive processes. *researchgate.net*, 9(5), Dec. 2016.

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

M. F. Land. Eye movements and the control of actions in everyday life. *Prog. Retin. Eye Res.*, 25(3):296–324, May 2006. URL http://dx.doi.org/10.1016/j.preteyeres.2006.01.002.

M. F. Land and M. Hayhoe. In what ways do eye movements contribute to everyday activities? *Vision Res.*, 41(25-26): 3559–3565, 2001. URL https://www.ncbi.nlm.nih.gov/pubmed/11718795.

M. F. Land and P. McLeod. From eye movements to actions: how batsmen hit the ball. *Nat. Neurosci.*, 3(12):1340–1345, Dec. 2000. URL http://dx.doi.org/10.1038/81887.

J. N. MacGregor and Y. Chu. Human performance on the traveling salesman and related problems: A review. *The Journal of Problem Solving*, 3(2):2, 2011. URL https://docs.lib.purdue.edu/jps/vol3/iss2/2/.

D. L. Mann, H. Nakamoto, N. Logt, L. Sikkink, and E. Brenner. Predictive eye movements when hitting a bouncing ball. *J. Vis.*, 19(14):28, Dec. 2019. URL http://dx.doi.org/10.1167/19.14.28.

F. Mars and J. Navarro. Where we look when we drive with or without active steering wheel control. *PLoS One*, 7(8): e43858, Aug. 2012. URL http://dx.doi.org/10.1371/journal.pone.0043858.

J. S. Matthis, J. L. Yates, and M. M. Hayhoe. Gaze and the control of foot placement when walking in natural terrain. *Curr. Biol.*, 28(8):1224–1233.e5, Apr. 2018. URL http://dx.doi.org/10.1016/j.cub.2018.03.008.

N. Mennie, M. M. Hayhoe, and B. Sullivan. Look-ahead fixations: anticipatory eye movements in natural tasks. *Exp. Brain Res.*, 179(3):427–442, May 2007. URL http://dx.doi.org/10.1007/s00221-006-0804-0.

I. Montfoort, M. A. Frens, I. T. C. Hooge, G. C. L.-v. Haselen, and J. N. van der Geest. Visual search deficits in Williams-Beuren syndrome. *Neuropsychologia*, 45(5):931–938, Mar. 2007. URL http://dx.doi.org/10.1016/j.neuropsychologia.2006.08.022.

D. A. Norman and T. Shallice. Attention to action. In R. J. Davidson, G. E. Schwartz, and D. Shapiro, editors, *Consciousness and Self-Regulation: Advances in Research and Theory Volume 4*, pages 1–18. Springer US, Boston, MA, 1986. URL https://doi.org/10.1007/978-1-4757-0629-1_1.

F. J. Parada and A. Rossi. Perfect timing: Mobile Brain/Body imaging scaffolds the 4e-cognition research program. Mar. 2020. URL psyarxiv.com/wru4j.

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

Z. Pizlo and Z. Li. Solving combinatorial problems: the 15-puzzle. *Mem. Cognit.*, 33(6):1069–1084, Sept. 2005. URL http://dx.doi.org/10.3758/bf03193214.

B. T. Sullivan, L. Johnson, C. A. Rothkopf, D. Ballard, and M. Hayhoe. The role of uncertainty and reward on eye movements in a virtual driving task. *J. Vis.*, 12(13):19, Dec. 2012. URL http://dx.doi.org/10.1167/12.13.19.

S. Van der Stigchel. An embodied account of visual working memory. *Vis. cogn.*, 28(5-8):414–419, Sept. 2020. URL https://doi.org/10.1080/13506285.2020.1742827.

B. Voloh, M. Watson, S. König, and T. Womelsdorf. MAD saccade: statistically robust saccade threshold estimation. Dec. 2019. URL https://osf.io/rzd6v.

M. Wilson. Six views of embodied cognition. *Psychon. Bull. Rev.*, 9(4):625–636, Dec. 2002. URL http://dx.doi.org/10.3758/bf03196322.

A. L. Yarbus. *Eye Movements and Vision*. Springer, 1967.