JUST-IN-TIME: GAZE GUIDANCE IN NATURAL BEHAVIOR

A PREPRINT

Ashima Keshava
University of Osnabrück
Germany
akeshava@uos.de

Farbod Nosrat Nezami
University of Osnabrück
Germany

Henri Neumann
Halocline GmbH & Co. KG
Germany

Krzysztof Izdebski
Halocline GmbH & Co. KG
Germany

Thomas Schüler
Halocline GmbH & Co. KG
Germany

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

January 30, 2024

ABSTRACT

Natural eye movements have primarily been studied for over-learned activities such as tea-making, sandwich-making, and hand-washing, which have a fixed sequence of associated actions. These studies indicate a sequential activation of low-level cognitive schemas facilitating task completion. However, it is unclear if these action schemas are activated in the same pattern when a task is novel and a sequence of actions must be planned in the moment. Here, we recorded gaze and body movements in a naturalistic task to study action-oriented gaze behavior. In a virtual environment, subjects moved objects on a life-size shelf to achieve a given order. In order to compel cognitive planning, we added complexity to the sorting tasks. Fixations aligned with the action onset showed gaze as tightly coupled with the action sequence, and task complexity moderately affected the proportion of fixations on the task-relevant regions. Our analysis further revealed that the gaze was allocated to action-relevant targets just in time. Planning behavior predominantly corresponded to a greater visual search for task-relevant objects before the action onset. The results support the idea that natural behavior relies very little on working memory, and humans refrain from encoding objects in the environment to plan long-term actions. Instead, they prefer just-in-time planning by searching for action-relevant items in the moment, directing their body and hand to it, monitoring the action until it is terminated, and moving on to the next action.

***K*eywords** action planning, gaze-guided planning, natural behavior, action schemas, virtual reality

Author Summary

Eye movements in the natural environment have primarily been studied for over-learned and habitual everyday activities (tea-making, sandwich-making, hand-washing) with a fixed sequence of associated actions. These studies show eye movements correspond to a specific order of actions learned over time. In this study, we were interested in how humans plan and execute actions for tasks that do not have an inherent action sequence. To that end, we asked subjects to sort objects based on object features on a life-size shelf in a virtual environment as we recorded their eye and body movements. We investigated the general characteristics of gaze behavior while acting under natural conditions. Our paper provides a comprehensive approach to preprocess naturalistic gaze data in virtual reality. Furthermore, we provide a data-driven method of analyzing the different action-oriented functions of gaze. The results show that bereft of a predefined action sequence, humans prefer to plan only their immediate actions where eye movements are used to search for the target object to immediately act on, then to guide the hand towards it, and to monitor the action until it is

terminated. Such a simplistic approach ensures that humans choose sub-optimal behavior over planning under sustained cognitive load.

1 Introduction

In a pragmatic turn in cognitive science, there is a significant push towards the study of cognition and cognitive processing during context-dependent interactions within the constraints of the local properties of an environment (Clark, 1998; Parada and Rossi, 2020). This pragmatic shift puts voluntary, proactive control at the heart of cognitive processing. Moreover, Engel et al. (2013) proposed that cognition encompasses the body, and bodily action can be used to infer cognition. Understanding the cognitive control of eye movements in natural, real-life situations requires a mobile setup that allows a human subject to be recorded while proactively interacting in a controlled yet unconstrained environment (König et al., 2018). Studying eye movements in mobile subjects might give us a richer, veridical view of cognitive processing in naturalistic settings.

Seminal studies have investigated eye movement behavior in the natural environment with fully mobile participants. 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 that involved manipulating objects one at a time to achieve a goal. Both studies showed that nearly all the fixations were primarily allocated in a task-oriented manner. Further studies investigated eye movements under a plethora of natural conditions while walking (Matthis et al., 2018), driving (Mars and Navarro, 2012; Sullivan et al., 2012; Navarro et al., 2020), hand-washing (Pelz and Canosa, 2001), hitting a ball (Land and McLeod, 2000), and free exploration (Schumann et al., 2008). These experiments in naturalistic settings have revealed several distinct functions of eye movements during habitual everyday tasks.

Investigation of habitual tasks has uncovered a systematic timing between visual fixations and object manipulation. Specifically, fixations are made to target objects about 600ms before manipulation. Notably, Ballard et al. (1995) proposed a "just-in-time" strategy that universally applies to this relative timing of fixations and actions. In other words, fixations that provide information for a particular action immediately precede that action and are crucial for the fast and economical execution of the task. Moreover, Land and Hayhoe (2001) have broadly categorized fixations while performing habitual tasks into four functional groups. "Locating" fixations retrieve visual information. "Directing" fixations acquire the position information of an object, accompany a manipulation action, and facilitate reaching movements. "Guiding" fixations alternate between manipulated objects, e.g., fixations on the knife, bread, and butter while buttering the bread. "Checking" fixations monitor where the task constraints in the scene have been met. These categories have been corroborated by Pelz and Canosa (2001) and Mennie et al. (2007). Hence, there is a just-in-time strategy of gaze allocation while performing common tasks.

Based on the above observations, Land and Hayhoe (2001) proposed a framework that outlines the flow of visual and motor control during task execution Figure 1A. 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. 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." (Land, 2006). In the action preparation period, the gaze control samples the information about the location and identity of the object and directs the hands to it. Subsequently, in the action execution period, the motor control system of the arms and hands implements the desired actions. Here, vision provides information about where to direct the body, which state to monitor, and determine when the action must be terminated. A 'script' of instructions is sequentially implemented where the eye movements earmark the task-relevant locations in the environment that demand attentional resources for that action.

An implicit theme in the studies investigating behavior in natural tasks (e.g., tea-making, sandwich-making, hand-washing) is that these tasks have an organized and well-known structure. They involve specific task-relevant objects, object-related actions such as picking up the teapot, pouring water, etc., and a predefined 'action script' for executing the tasks. Therefore, they study eye movements under strict control of a task sequence. Moreover, these tasks are over-learned as they are part of a habitual action repertoire for an adult human being. For these habitual tasks, cognitive schemas allocate gaze for specific information retrieval (locating, directing, guiding, monitoring) and are likely not executed under deliberate conscious control. Hence, it is unclear how cognitive schemas control gaze where an internal task script is unknown.

Furthermore, a core feature of Land and Hayhoe's gaze and motor control framework is its sparse use of visual working memory. In this framework, low-level cognitive schemas are activated successively as needed. For example, it states that a visual search is performed if the object's location is not in memory. The model does not distinguish whether or how additional cognitive load would affect the activation of each schema in its given sequence. As Hayhoe and Ballard (2005) states, fixations provide information at different levels, such as for behavior, arbitration, and the broader

context. At the level of individual behaviors, fixations provide the state of the environment and facilitate computations of whether a behavioral goal has been satisfied. A set of active behaviors might make perceptual and physical demands for attentional resources for which gaze can only be allocated sequentially. Hence, a mechanism is needed at the arbitration level to mediate conflicting resource demands. Finally, at the context level, gaze provides information to enable a set of context-specific behaviors, considering the state of the environment and the agent's goal. In well-learned tasks, action schemas are triggered automatically with little conscious control. However, when a task is relatively complex and requires top-down planning of subsequent actions, more information must be retained in memory as multiple schemas compete for attentional resources. Hence, a comprehensive action plan would need to maintain information about the object in working memory, how it should be manipulated, the conditions that signal the desired end state of the manipulation, and the subsequent actions specific to the context. As working memory is limited, it is pertinent to understand how attentional resources are arbitrated within the context of undefined action sequences and additional cognitive load.

To further pursue this line of research, performing such experiments under tightly controlled laboratory conditions is desirable. In recent years, virtual reality (VR) and mobile sensing have offered great opportunities to create controlled, naturalistic environments. Here, subjects' eye and body movements can be measured reliably along with their interactions with the environment (Keshava et al., 2023, 2020; Clay et al., 2019; Mann et al., 2019). Experiments in virtual environments have grown popular in recent years and have shown promise in studying cognition in naturalistic and controlled environments.

The present study aims to generalize the framework proposed by Land and Hayhoe (2001) to include the novelty of tasks and the effect of cognitive load. The tasks were novel as they did not have an associated predefined sequence of actions. Hence, the tasks required participants to plan actions for the defined context. Furthermore, we created two types of tasks that varied in cognitive load and required keeping multiple object features in memory. In virtual reality, we asked human volunteers to sort objects on a life-size shelf based on the object's features. The complexity of the task depended on sorting based on one object feature or two. We concurrently recorded the gaze positions and body movements while subjects performed the tasks. Thus, our study investigates the mechanism of gaze allocation in novel tasks with varying complexity in a naturalistic environment.

Our study explored the characteristics of action-related gaze behavior for different task demands. At the outset, we tested the hypothesis of whether additional cognitive load influenced behavioral outcomes. In this regard, we were interested in the duration of preparation and execution of actions and the optimality of the action choices. Hence, optimal planning would indicate an optimal use of working memory where an action script is devised in advance but requires additional time. Similarly, with optimal planning, the execution durations would not differ between task complexities. Conversely, if subjects do not formulate a high-level plan for the action sequences, the preparation and execution of these actions would differ considerably, and more complex tasks would show proportionally longer durations. We further explored the corresponding action-related gaze behavior and its differences with respect to cognitive load. With optimal planning, we expected initial fixations to encode the objects and corresponding action plans into memory. If an object's locations are kept in memory for a considerable time, the added task complexity would not markedly increase visual search at the level of preparing for and executing actions, especially at the beginning of the task. However, when planning is sub-optimal, data would show lower search behavior and greater just-in-time fixations during the action preparation period corresponding to choosing objects randomly and then figuring out how to move them, resulting in a greater visual search for appropriate locations during the execution period. Lastly, we tested the latency of fixations on the task-relevant regions and whether Land's framework adequately predicts the sequence of fixations that occur in a defined sequence relative to the action onset. Furthermore, we explored how the latency of fixations on task-relevant regions differed based on task complexity. In this way, we tested the generalization of the theoretical framework for attentional resource allocation in novel tasks.

2 Results

In the present study, 55 human volunteers performed object-sorting tasks in a virtual environment. They sorted objects on a 5x5 shelf that measured approximately 2m in width and height. The sorting tasks were based on the color and shape of the objects, where we experimentally modulated the task complexity into EASY and HARD trials. In the EASY trials, we gave participants instructions such as "Sort objects so that each row has the same color or is empty," and in the HARD tasks, an example trial instructed them to "sort objects so that each row has each unique color and unique shapes once." Participants wore an HTC Vive Pro Eye virtual reality head-mounted display (HMD) as they performed the task. Using the HMD, we measured their eye and head movements simultaneously, while the HTC hand controllers measured their hand movements and grasping behavior. Figure 1B illustrates the experimental setup. The eye and body tracking sensors were calibrated at regular intervals throughout the experiment session. Section 5.2 details the apparatus and the experiment setup.

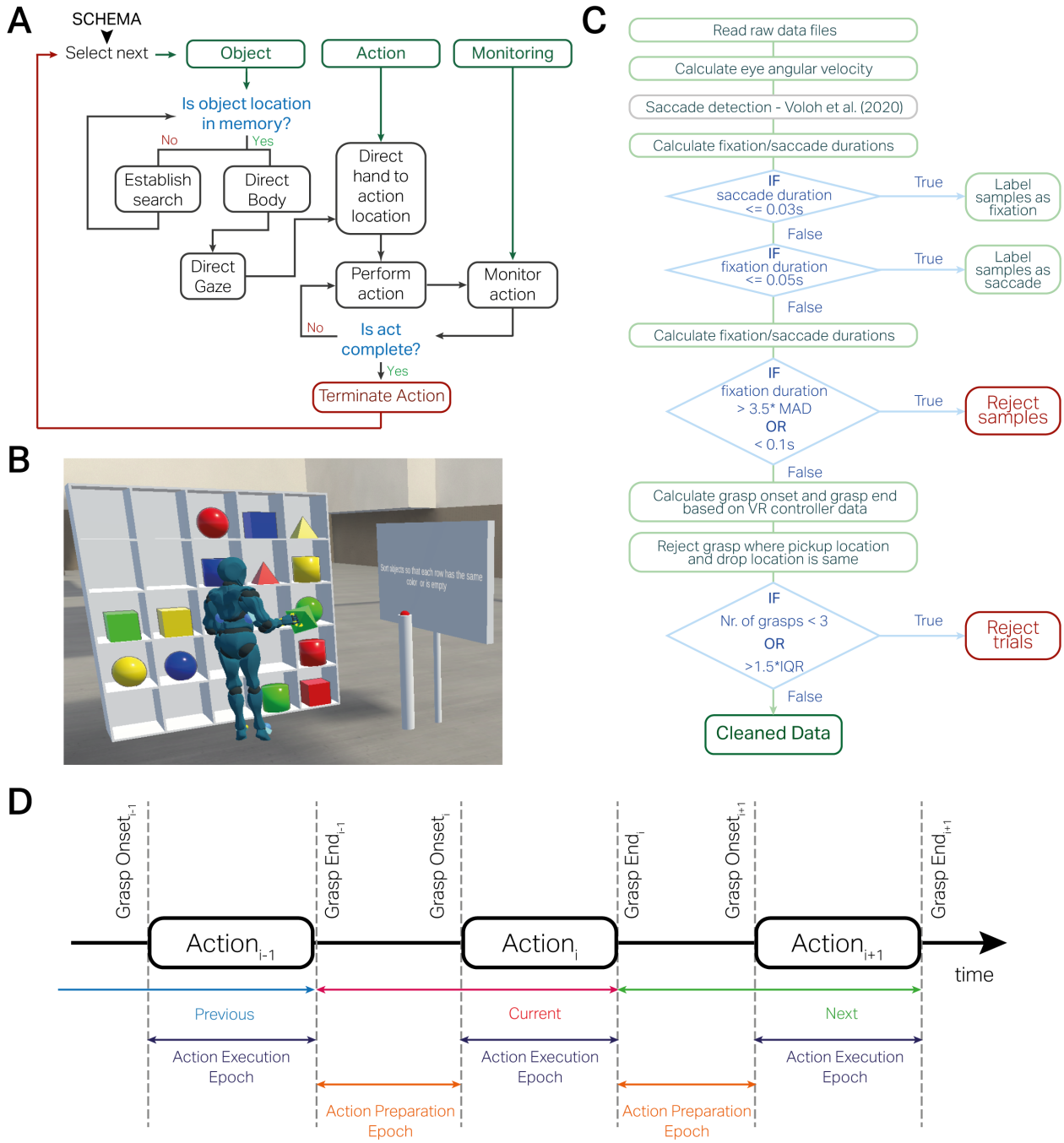

Figure 1: A. Schematic of motor and gaze control during performance of natural tasks, adapted from Land and Hayhoe (2001). A cognitive schema selects between the object of interest, the corresponding action and the monitoring of the action execution. If the object location is not in memory, a visual search is undertaken to locate the object, the body and hands are directed towards it in preparation for the action execution, and finally, the action is monitored until the behavioral goal has been achieved. **B.** Experimental Task. In a virtual environment, participants sorted 16 objects based on two 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 color. Trials, where objects were sorted based on just one object feature (color or shape), were categorized as EASY trials. Conversely, the trials where sorting was based on both features (color and shape) were categorized as HARD trials. All participants performed 24 trials (16 easy and 8 hard) without time constraints. **C.** Data preprocessing steps for fixation/saccade detection and data rejection. **D.** Labelling of the continuous data stream into action preparation and execution epochs. To study the function of eye movements, we divided each trial into action preparation and execution epochs. The action execution epochs start from the grasp onset till the grasp ends for each object displacement. In contrast, the action preparation epochs start from the end of grasping the previous object displacement and the grasp onset of the current object displacement.

We utilized three data streams for this study: the cyclopean (average of the left and right eye) gaze position, direction, and hand movement position. The pre-processing steps used to clean the data are outlined in 1C and described in detail

in section 5.4. At the outset, we rejected data from participants where the sensors malfunctioned mid-experiment. We derived the angular velocity of the cyclopean eye using the gaze position and direction data. We used a median absolute deviation-based saccade detection method to cluster gaze samples into either fixation or saccade. Following this, we classified fixations to gaze samples with a duration greater than 100 ms. We rejected the fixation samples with a duration greater than 3.5 median absolute deviation. This step rejected 2.4% of the outlying gaze data.

Next, we derived the onset and end of a grasp from the VR hand controller's trigger button, i.e., the time when a participant pressed the button with their right index finger to pick up an object and the time point when they released the button after moving it to a desired location. We did not consider grasps where the pickup and drop-off locations were the same. Finally, trials with fewer than three grasps were rejected. Also, we rejected the trials with more than 1.5 times the inter-quartile range (IQR) of the number of grasps by all participants. This step helped remove trials where subjects had outlying behavior. Hence, after pre-processing, we were left with 12403 grasps over 999 trials from 48 participants.

As a final step, we segmented the data stream into the action preparation and execution epochs. As illustrated in Figure 1D, these epochs marked the end of the previous grasp and the start of the immediate grasp and the time from the start of the current grasp to its end, respectively. Hence, every grasping action had a preparation and an execution phase with concurrent gaze data associated with it. Further, using the gaze data stream, we defined regions of interest (ROI) pertaining to the target object that was grasped and the target shelf where it was moved. These ROIs were further classified into the previous action, the current action, and the next action. The regions that did not correspond to targets in the action sequence were categorized as 'other objects' and 'other shelves'. This step prepared the data to demonstrate a "flow" of fixations from one action-relevant region to another during both the preparation and the execution phase of the grasping action.

2.1 Influence of Task Complexity on Behavior

In this study, the primary object-related action was to iteratively prepare and execute the pickup and placement of objects until they were sorted according to the task instructions. We used four measures to account for the behavioral differences between the task complexities. First is trial duration, i.e., the total time to finish the tasks. Second is the action preparation duration, which is the time between dropping the previous object and picking up the current target object. Third is execution duration, i.e., the time to move the target object to the desired location on the shelf. Fourth, the number of object displacements or "moves" made by the participants to complete the EASY and HARD tasks. These measures characterize the behavioral outcomes influenced by the task complexity.

Figure 2A shows the trial durations in EASY and HARD trials. A two-sample independent t-test showed that the trial duration for the two trial types was significantly different ($t = -10.92, p < 0.001$) where EASY trials were shorter (Mean = $51.15s$, SD = ± 10.21) as compared to HARD trials (Mean = $123.46 \pm 44.70s$). Hence, the increase in task complexity markedly increased the duration of the trials by a factor of 2.

At a more granular level, we were interested in how task complexity affected the average action preparation and execution duration. For EASY tasks, the mean preparation epoch duration was $2.10s$ (SD = ± 0.78), and the mean execution epoch duration was $1.76s$ (SD= ± 0.27). The mean preparation epoch duration for HARD trials was $4.05s$ (SD= ± 1.99), and the mean execution epoch duration was $2.15s$ (SD = ± 0.58). Figure 2B depicts the mean and variance of the preparation and execution duration for the EASY and the HARD tasks.

We tested the hypothesis that the complexity of HARD trials increased the action preparation and action execution duration. We used a linear mixed effects model (section 5.5.1) with epoch duration for a given grasp as the dependent variable and trial type (EASY, HARD), epoch type (PREPARATION, EXECUTION) as independent variables. We also added the grasp index i.e., an integer denoting the order of the object displacement in the trial (for example, 1st,2nd,3rd, ..) as a covariate. The independent variables are effect-coded, so the regression coefficients can be directly interpreted as main effects. Table 1 summarizes the model results. Epoch duration was significantly different for the two trial types, where epochs of HARD trials were $1.31s$ longer than EASY trials. Irrespective of the trial type, the action preparation epochs were significantly longer than the execution epochs by $1.08s$. An insignificant positive correlation of 0.002 with the grasp index indicates that overall, the epoch duration of initial moves did not differ much from the latter ones. The preparation duration was significantly longer in the HARD trials, as shown by the interaction between the trial and epoch types. Furthermore, the difference in the slope of epoch duration and grasp index of HARD and EASY trials was insignificant. Similarly, the difference in the slope of the duration with respect to the grasp index between preparation epochs and execution epochs was also not significant. However, the three-way interaction between trial type, epoch type, and the grasp index was 0.01 and significant, showing that the preparation duration of the HARD trials was significantly shorter at the beginning of the trial and increased linearly as the trial progressed. To summarize, the preparation and execution phases of the action were distinctly different, and task complexity specifically affected the duration of preparing actions and less so the time taken to execute them. Moreover, the significant interaction terms

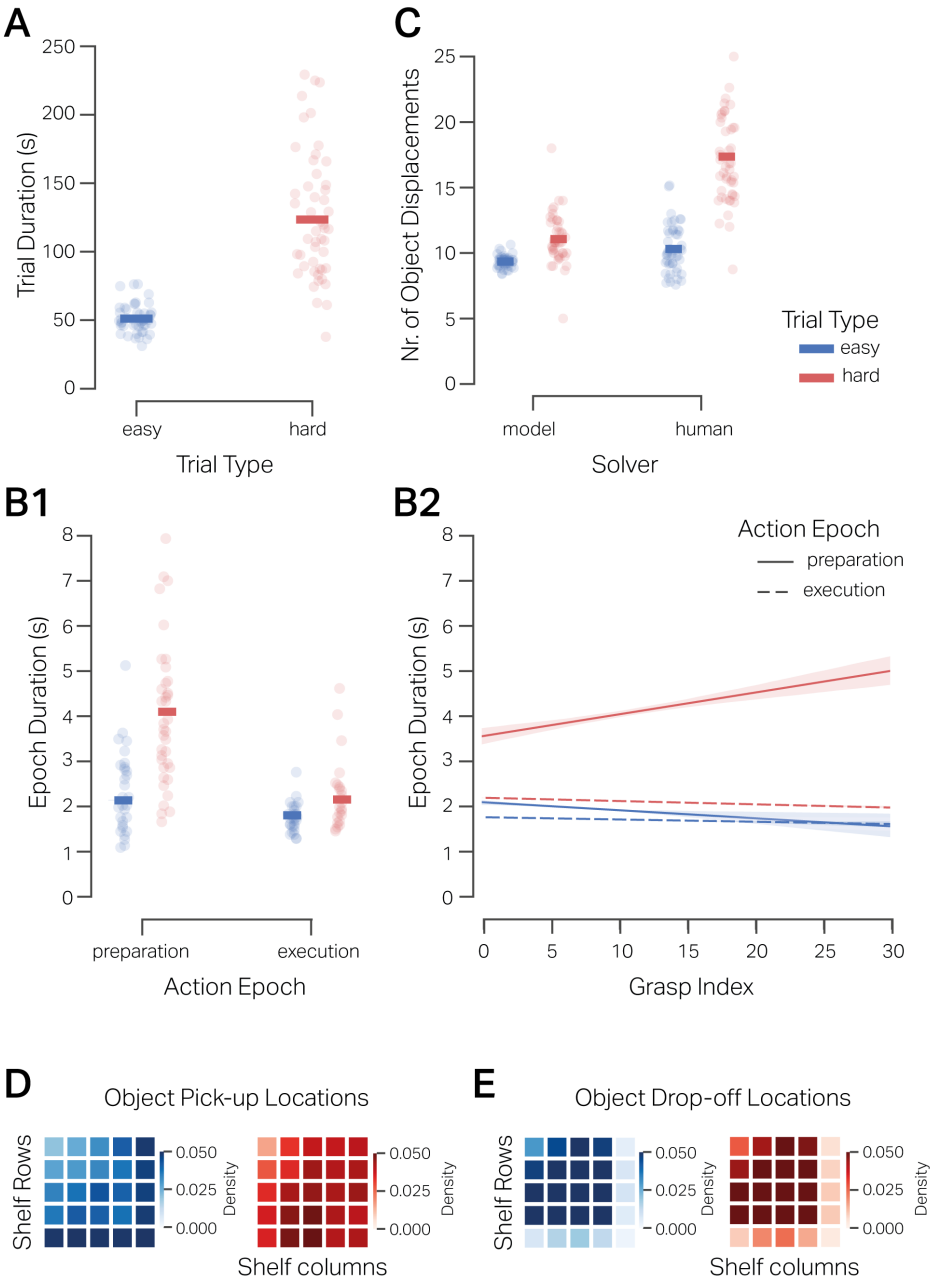

Figure 2: A. Trial duration of the EASY and HARD trials. Each dot represents the mean trial duration of each participant, and the solid bars indicate the mean duration over participants. **B1.** Duration of the action preparation and action execution epochs under the EASY and HARD trial conditions. Each dot represents each participant's mean action epoch duration, and the solid bars indicate the mean epoch duration over participants. **B2** Linear fit of epoch duration w.r.t. the index of the grasp/action in a trial (e.g., 0, 1, 2 indicates first, second, and third action of the trial respectively). The plot shows that the preparation duration of the early grasps in HARD tasks was shorter than in the later grasps. This was not the case for the preparation or execution duration of EASY trials. **C.** Number of object displacements made by participants vs. those calculated by a depth-first search model. The initial conditions of the task were inputted into a depth-first search algorithm, yielding the optimal number of moves required to complete the respective trial. Each dot represents the mean object displacement of a participant and the optimal counterpart over the EASY and HARD trials. The solid bars denote the mean over subjects. **D** Propensity of picking up objects from a preferred location on the 5x5 shelf locations with blue heatmaps showing the probability density over EASY trials and red heatmaps showing density over HARD trials. The probability density shows that subjects have the propensity to pick up objects from the rightmost column and bottom column for EASY trials (left), and conversely, in the HARD trials (right), subjects pick up objects from central locations. **E.** Propensity of dropping off objects to a preferred location on the 5x5 shelf locations. The probability density shows subjects' systematic preference to move objects to locations other than the bottom row or rightmost column.

of the model suggest that in HARD trials specifically, participants made fast moves at the beginning of the trials and spent more time preparing their actions in the later part.

Table 1: Fixed effects coefficients of the linear mixed model showing the estimated duration of epochs w.r.t task type, epoch type, and grasp index as a covariate. The coefficients and 95%CI have been back-transformed to the response scale and can be interpreted as main effects.

Model:

$$log(duration) \sim 1 + trial_type * epoch_type * grasp_index$$
$$+(1 + trial_type + epoch_type + grasp_index|Subject)$$

	Coefficient	95% CI	t	p-value
trial type (HARD - EASY)	1.31	[1.26, 1.37]	12.36	**< .001*****
epoch type (PREPARATION - EXECUTION)	1.08	[1.01, 1.16]	2.28	**0.02***
grasp index	0.002	[0.00, 0.00]	1.42	0.16
Interactions				
trial type : epoch type	1.36	[1.31 , 1.40]	17.13	**< .001*****
trial type : grasp index	-0.003	[-0.01, 0.0]	-1.44	0.15
epoch type : grasp index	0.002	[0.00, 0.01]	1.01	0.31
trial Type : epoch type : grasp index	0.01	[0.00, 0.02]	2.24	**0.006****

Table 2: Fixed effects coefficients of linear mixed model showing the estimated number of object displacements dependent on the trial type and the solver type

Model:

$$ObjectDisplacements \sim 1 + trial_type * solver_type$$
$$+(1 + trial_type + solver_type|Subject)$$

	Estimate	95% CI	t	p-value
trial type (HARD - EASY)	4.43	[3.81, 5.05]	14.01	**< .001*****
solver type (HUMAN - MODEL)	3.36	[2.82, 3.90]	12.16	**< .001*****
trial type : solver type	4.90	[4.25, 5.55]	14.78	**< .001*****

To further determine whether participants optimally planned and sorted the objects during the trials, we compared their number of object displacements with an optimal depth-first search algorithm. We fed the initial configuration of the shelf, as shown to the participants in each trial, into a depth-first search algorithm, which determined the optimal number of moves needed to reach the solution in that trial. Figure 2C shows the mean number of the object displacements made by the subjects and the optimal number of moves as elicited by the model for both EASY and HARD trials. In the EASY trials, subjects made an average of $10.30(SD = \pm 1.80)$ moves while the model optimal was $9.34(SD = \pm 0.45)$. In the HARD trials, participants made an average of $17.36(SD = \pm 4.04)$ moves when the optimal number was $11.06(SD = \pm 1.93)$. To test whether participants were optimal in their action choices, we fitted the data using a linear mixed model with the number of object displacements as the dependent variable and trial type (EASY, HARD), solver (HUMAN, MODEL) as the independent variables. Again, as the independent variables were effect-coded, the regression coefficients can be directly interpreted as main effects. Table 2 details that the trial type significantly affected the number of object displacements, irrespective of the solver, where HARD trials demand an estimated $4.43(95\%CI = [3.81, 5.05])$ more moves than the EASY trials. Similarly, irrespective of the trial type, participants made $3.32(95\%CI = [2.82, 3.90])$ more object displacements than the optimal. Finally, there was a significant interaction between the trial type and solver, i.e., the participants made many more moves than the model, specifically in the HARD tasks. In sum, with increased task complexity, participants were more sub-optimal in the HARD trials than in the EASY trials.

As participants deviated from the optimal in the EASY trials, especially in HARD trials, we checked if they employed any *salient* spatial heuristics to accomplish the tasks. To do this, we used the measure of propensity to pick up and drop off objects from and to preferred locations on the shelf. We calculated the counts of the object displacement for each pickup and drop-off location. We then normalized the counts with the total number of displacements in the EASY and HARD tasks. The normalized counts in the figures add up to 1. In Figure 2D shows that subjects preferred to pick up objects from the right-most column and bottom-most row of the shelf. Figure 2E shows that subjects had a propensity to drop the objects, leaving out the right column and bottom row of the shelf for both EASY and HARD trials. Given the sorting tasks where each trial began with a random initial configuration 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. These results show that subjects in our study systematically employed an arbitrary and non-task-specific heuristic to tackle the problem. As the objects are randomly instantiated on the shelf, an optimal strategy would not show this behavior, and all locations on the shelf would have an equal probability for object pick up and drop off. Hence, we can conclude from the above that subjects offset their cognitive load of optimally completing the task by employing simple spatial heuristics. In other words, in lieu of optimally performing the task, subjects preferred to offload cognitive effort on the environment by adopting a more sub-optimal strategy.

The above behavioral results show that the categories of action epoch (preparation, execution) and task complexity (EASY, HARD) have distinct differences and interactions. The increase in the trial duration in HARD tasks was primarily due to an increase in the time taken to plan each action and less so due to an increase in the time to execute them. With increased task complexity, participants were increasingly more sub-optimal in their actions. Lastly, although participants spent more time planning their immediate moves, especially within HARD tasks, they did not plan optimally and instead relied on a sub-optimal spatial heuristic.

2.2 Action-Locked Gaze Control

In the previous section, our analysis showed that task complexity significantly affects behavior in the preparation phase and less so in the execution phase. Next, we wanted to elucidate further the role of gaze control in the preparation and execution of these actions and how it is affected in complex task conditions. We also wanted to conduct analysis that would generalize over activities that utilize planning successive actions from one object to the next. We labeled eight regions of interest (ROIs) for each object displacement to do this. These ROIs denote the object and shelf locations handled in the previous, current, and next object displacements. Specifically, the current target object denoted the object currently being displaced, and the current target shelf as the location where the object would be placed. Similarly, the previous target object and previous target shelf represented the object that was handled in the previous object displacement and the shelf location where it was placed. It should be noted that the previous target object and the target shelf are spatially the same, as the action of displacing the object has already been accomplished. The next target object and target shelf represented the objects and locations used after the current object displacement ended. Lastly, the 'other objects' and 'other shelves' categories represent objects and shelf locations that were irrelevant to the sequence of previous, current, and next actions. Together, these ROIs helped us investigate the flow of gaze control from one action to the next during the different phases of each action.

First, our analysis focused on the absolute timing of fixations on the eight ROIs time-locked to the grasp onset, i.e., the time when the hand makes contact (and the VR hand controller trigger button is pressed) with the current target object. For the analysis, we chose the period from 3s before action onset to 2s after. The chosen time period approximately corresponds to the average duration of the preparation and execution epochs over the two trial types. We were also interested in the influence of task complexity on the average fixation behavior. Figure 3 shows the time course of the proportion of fixations on the eight ROIs as described in section 5.5.2 for the EASY and HARD tasks.

In the action preparation phase, the average fixation profile shows important aspects of oculomotor control 3s before the grasp onset. At the beginning of the preparation phase, there is an equal proportion of about 30% of fixations on the previous target object (dark blue trace), 'other objects' (dark grey trace), and 'other shelves' regions (light grey trace) on the shelf. The comparable proportion of fixations on previous targets and the 'other' categories indicates a search behavior for task-relevant objects/shelves at the start of the preparation period. Subsequently, as the fixations on the previous targets and 'other objects' decrease, there is a steep increase in the proportion of fixations on the current target object (dark red trace), reaching a maximum of 68% of fixations at the point of grasp onset (time= 0s). Throughout the preparation period, the proportion of fixations on the next target object (dark green trace) and shelf (light green trace) remain comparatively minuscule. The large proportion of fixations in the action preparation phase devoted to the current target object shortly before the hand makes contact with the object indicates just-in-time fixations used to direct and guide the hand toward the target object.

In the action execution phase, up to 2s after the grasp onset, the proportion of fixations on the different ROIs reveals further action-relevant oculomotor behaviors. Firstly, with time, the proportion decreases from the current target object and move towards the current target shelf (light red trace) where approximately 40% of the fixations are allocated around 700ms after grasp onset and well before the end of the action execution at 1.91s ($95\% CI[1.81, 2.01]$). Secondly, before the end of the action execution, there is an increase in the proportion of fixations on the next target object. The presence of approximately 10% of fixations on the next target object even before the end of the current action execution shows that the target for the next action is located shortly before ending the current action. Moreover, about 20% of the fixations are devoted to the 'other' ROIs after 1s. This is again indicative of a search process for task-relevant objects/shelves. In a nutshell, during the action execution phase, the majority of fixations suggest two primary functions

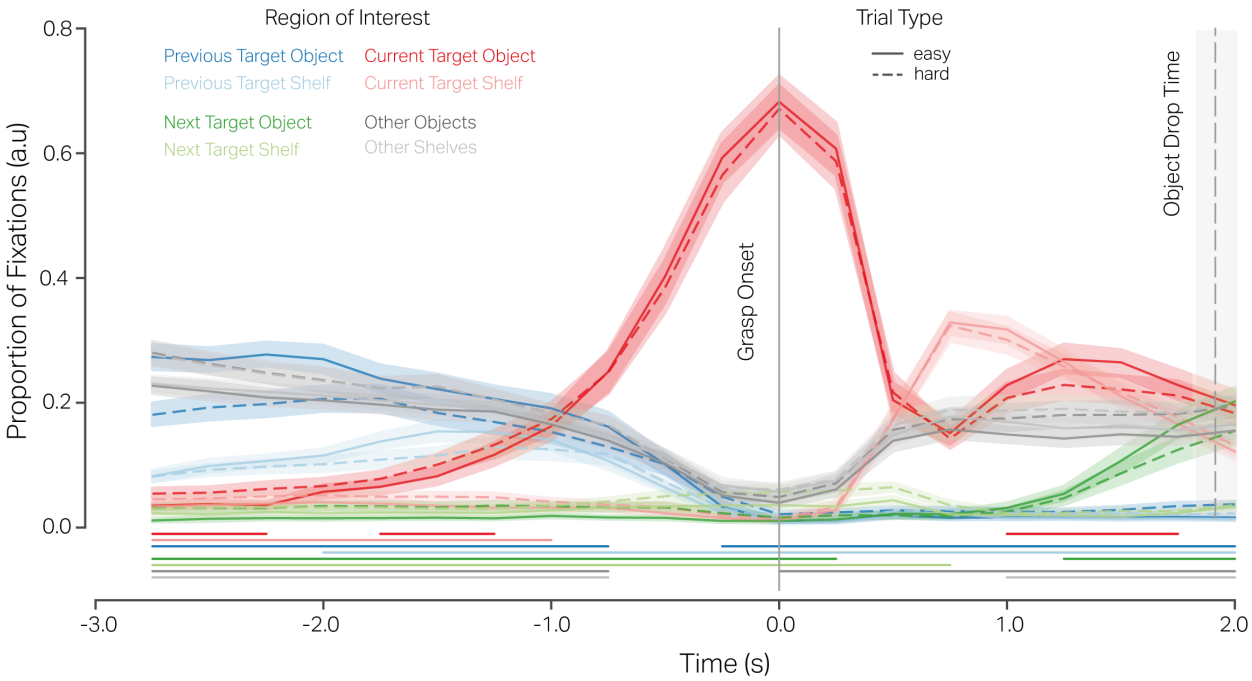

Figure 3: Proportion of fixations on the eight regions of interest time-locked to grasp onset (time=0s) from 3s before the onset to 2s after in the EASY tasks(solid trace) and HARD tasks(dashed trace). In each time bin of 0.25s, the proportion of fixations on all 8 ROIs for a trial type adds to 1. The shaded regions show the 95% CI of the mean proportion of fixations at each time-bin across all subjects. The dashed vertical line and shaded region denote the mean (time=1.99s)and 95%CI [1.81s, 2.01s] of the end of the action execution. The horizontal traces (corresponding to the same color of an ROI) at the bottom of the plot correspond to the time periods where the proportion of fixations on a given ROI was significantly different between the EASY and HARD trials as computed by a cluster permutation test.

Table 3: Significant differences in the proportion of fixations on the ROIs in the HARD and EASY trials as revealed by a cluster permutation test with a significance level of 0.006.

ROI	time start(s)	time end(s)	duration(s)	p-value
previous target object	-2.75	-0.75	2.00	<0.001
current target object	-2.75	-2.25	0.50	<0.001
current target shelf	-2.75	-1.00	1.75	<0.001
next target object	-2.75	0.25	3.00	<0.001
next target shelf	-2.75	0.75	3.50	<0.001
other objects	-2.75	-0.75	2.00	<0.001
other shelves	-2.75	-0.75	2.00	<0.001
previous target shelf	-2.00	2.00	4.00	<0.001
current target object	-1.75	-1.25	0.50	<0.001
previous target object	-0.25	2.00	2.25	<0.001
other objects	0.00	2.00	2.00	<0.001
current target object	1.00	1.75	0.75	<0.001
other shelves	1.00	2.00	1.00	<0.001
next target object	1.25	2.00	0.75	<0.001

of gaze: first, to aid the completion of the current action by fixating on the shelf location in anticipation of the object drop-off, and second, to search or locate the objects for the next action.

Importantly, the proportion of fixations on the eight ROIs in the EASY and HARD tasks does not deviate dramatically across time. Even though there are large behavioral differences in the EASY and HARD tasks, both in terms of time taken to prepare for the actions and the number of actions required, the fixation profiles time-locked to the grasp onset are comparable. Most importantly, we see that the fixations are made just in time, where gaze shifts occur systematically in sync with the previous, current, and next action sequences. Hence, we can conclude that the average oculomotor behavior described above is driven just-in-time by overt actions and less so by the cognitive load of the different sorting tasks.

There are, however, small differences in the proportion of fixations due to the task complexity. These differences point to aspects of the oculomotor behavior that are subject to the cognitive load imposed by task complexity. To quantify these differences, we used a cluster permutation analysis (section 5.5.3) on the average fixation time course over each ROI. The analysis revealed several time periods where the proportion of fixations significantly differed in the two task types on each of the eight ROIs. Table 3 details the time periods where the proportion of fixation on the ROIs significantly differed between the EASY and HARD trials. The data here numerically supplement the horizontal traces at the bottom of Figure 3, which depict the start and end of the time period when the differences in a given ROI were significant.

An interesting characteristic of the significant time periods is that they occur early in the action preparation epoch. There are fewer fixations on the previous target object (dashed dark blue trace) and previous target shelf (dashed light blue trace) and more fixations on the other target objects (dashed dark grey trace) and shelf (dashed light grey trace). Subsequently, in these tasks, there are more fixations on the next target object (dark green trace) and shelf (light green trace), which are of interest in the next action. This indicates a search for task-relevant target objects for the current action. Additionally, it is plausible that in some trials, participants 'located' the target objects for the next action concurrently with the current action. Also, in HARD trials, in the early preparation epoch, there are significantly more fixations on the current target object (dashed dark red trace) and target shelf (dashed light red trace). These early fixations on the target object indicate that they are 'locating' fixations. These results show that task complexity influences the average fixation profile early in the action preparation phase and primarily serves the function of searching and locating immediately relevant objects.

Concurrent with the action onset and execution of HARD tasks, there are more fixations on the previous target object and shelf, the next target shelf, and the other objects. More fixations on the previous target object and shelf are characteristic of 'checking' fixations that are used to monitor the current task execution w.r.t to the task condition. Similarly, more fixations on the next target object, which are present in concert with the grasp onset, show that these fixations play the role of both 'checking' fixations to monitor the task progression as well as 'locating' fixations to locations relevant for the next action in the sequence. Further, more fixations on the 'other' objects indicate a greater search for the next target object in the HARD tasks. With the increase in fixations on the other object ROIs in HARD trials, later in the execution phase, there are correspondingly fewer fixations on the current and next target object. As a proportional decrease is not seen in the fixations on the current target shelf, we can make the case here that this period is used to search for the next target object. Thus, specifically in the execution epochs of HARD tasks, the average fixations on the different ROIs show the functions of task monitoring, locating, and searching for the next targets occur in quick succession.

More importantly, there are no differences in the proportion of fixations on the current target object later in the preparation epoch or on the current target shelf during the execution epoch. Thus, task complexity does not play a role in 'directing' fixations that guide the hand to the action-relevant object. Hence, these fixations are tightly coupled with the overt action.

To summarize, the results in this section support two broad findings. First, despite large behavioral differences between the two task types, the eye movements show moderate deviations. This supports the idea that naturalistic eye movements are tightly coupled to the actions and less so with the cognitive load added by task complexity. Secondly, the observed significant differences between EASY and HARD trials are compatible with a slightly increased focus away from the previous action towards the current and on to the next step in the action sequence. This includes an earlier reduction of fixations on previous targets to facilitate search, the requirement for additional checking fixations after action onset, and a later attendance towards the next target object. Thus, the cognitive processes maintain a structural sequence corresponding to the searching and locating of a target object, guiding hands to it, and monitoring the task.

2.3 Gaze Transition Behavior

To further elucidate the concurrent cognitive processes, we focused our analysis on sequences of attended regions of interest in individual trials. Specifically, we computed transition matrices to capture fixations to and from each of the eight ROIs as described in section 5.5.3. Figure 4A shows the exemplar transition matrix for a preparation epoch in an EASY trial. The total transitions represent the total number of fixations to and from one ROI. On the other hand, the net transitions denote the asymmetry in gaze allocation from one ROI to another. For example, an equal number of fixations to and from pairs of ROIs would result in zero net transitions between those ROIs. Similarly, greater transitions from one ROI to another than vice versa would result in greater net transitions from that ROI to the other. We used the metric mean relative net transitions as the ratio of net transition by total transition for both action epochs of a trial. Hence, we computed the mean relative net transition for the preparation and execution epochs per trial. With maximum relative net transitions equal to 1, we can expect an equal number of the net and total transitions between ROIs, indicating all saccades between ROIs are made parsimoniously and show a strictly action-oriented gaze guidance to ROIs, e.g., from previous to current in the action preparation phase. Such a gaze behavior would show no intermediate search for target

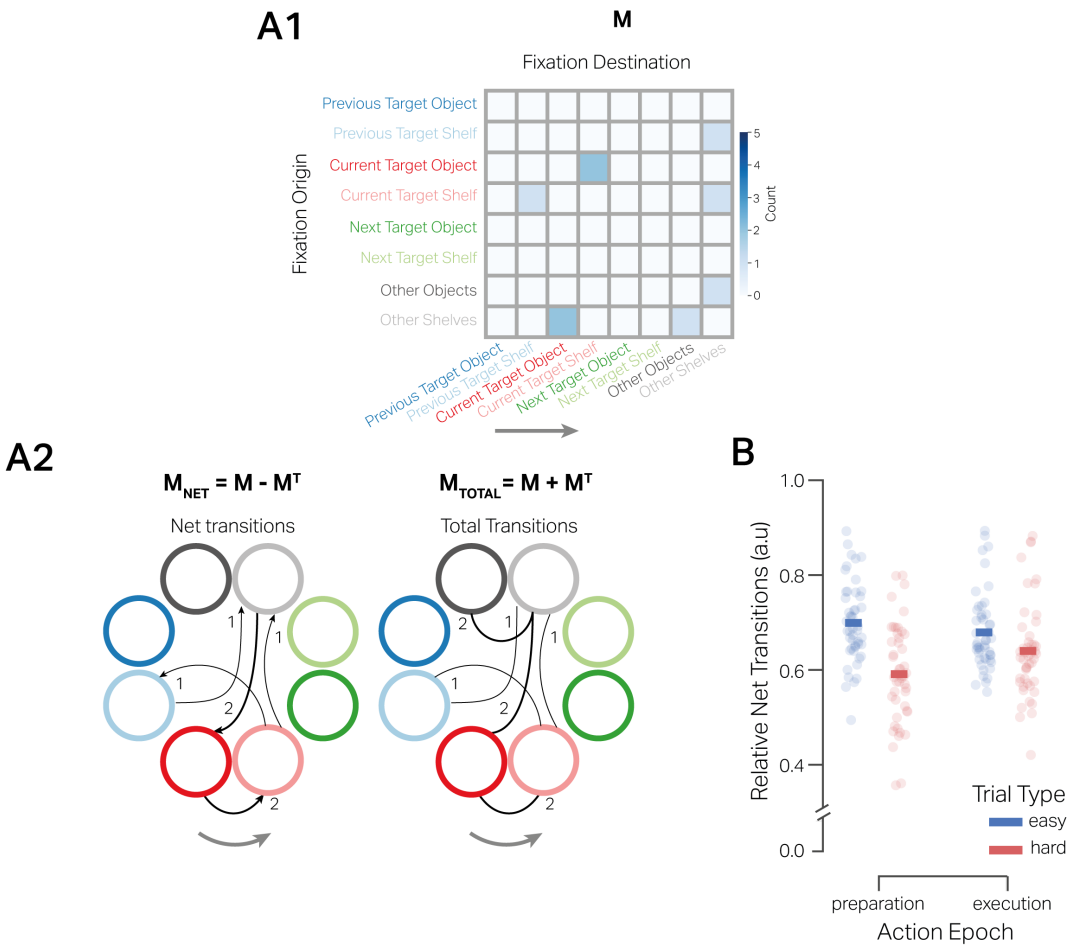

Figure 4: A. Exemplar transition matrix denoted as **M** for gaze switching between ROIs in a preparation epoch. The net (M_{net}) and total transitions (M_{total}) are calculated from matrix M. **B** Relative net transitions (ratio of net transitions and total transitions) for action preparation and execution (abscissa) in the two trial types (EASY and HARD). The bars denote the mean relative net transition across subjects, and the circles correspond to each participant.

Table 4: Fixed effects coefficients of the linear mixed model showing estimated relative net transition w.r.t task type, epoch type

Model:

$$NetTransitions \sim 1 + trial_type * epoch_type$$
$$+(1 + trial_type * epoch_type | Subject)$$

	Estimate	95% CI	t	p-value
trial type (HARD - EASY)	-0.06	[-0.09, -0.06]	-8.85	**0.001*****
epoch type (PREPARATION - EXECUTION)	-0.02	[-0.04, 0.01]	-1.76	0.08
trial type : epoch type	-0.06	[-0.08, -0.03]	-4.58	**< .001*****

objects. Conversely, the minimum relative net transition of 0 would indicate an equal number of saccades from and to ROI pairs. Searching for target objects by fixating back and forth over the ROIs or switching gaze between ROI pairs to monitor the task progression would lead to 0 net transitions. Hence, relative net transitions would help distinguish the function of saccades between ROI pairs as locating or searching target objects or monitoring the task.

Here, we tested the hypothesis of whether task complexity influenced the net transitions of fixations during action preparation and execution. From the previous analysis we inferred search behavior from proportion of fixations during action preparation and execution phase. We hypothesized that in HARD tasks, subjects would show lower net transitions during the action preparation epochs due to increased search behavior and even lower net transitions during the execution

epoch due to increased monitoring and searching for the next action targets. The data shows subtle differences in the relative net transitions for the trial types and the action epochs Figure 4B1. In the EASY trials, the mean relative net transition for the preparation epochs was $0.69 SD = \pm 0.10$ and for the execution epochs 0.68 ± 0.08. In the HARD trials, the mean relative net transition for the preparation epochs was 0.59 ± 0.11, whereas it was 0.64 ± 0.10 in the execution epochs. We used a linear mixed effects model (section 5.5.3) to fit the mean relative net transitions in a trial for per action epoch and trial type. Table 4 details the model coefficients. A significant main effect of trial type showed that HARD trials had lower relative net transitions than EASY trials. No effect of epoch type showed that overall preparation epochs did not differ from execution epochs regarding relative net transitions. However, there was a significant interaction between trial type and epoch type, showing that preparation epochs in HARD trials had significantly lower relative net transitions. Thus, increased task complexity was associated with decreased relative net transitions. Notably, with increased task complexity, the reduction in net transitions was greater in the action preparation epochs than in the execution epochs.

The analysis above lends evidence to differential gaze guidance due to task complexity, specifically in the action preparation epochs. The higher relative net transitions in the EASY trials suggest saccades were made towards the immediate action-relevant objects with little requirement for search. The lower net transitions in the HARD tasks suggest significantly higher gaze switching to and from ROIs in search of task-relevant objects, especially during action preparation. The influence of task complexity was not similarly pronounced in the action execution epochs, indicating eye movements were engaged in a similar function of monitoring the task progression, resulting in back-and-forth saccades between the previous, current, and next target ROIs in the action sequence. Hence, with increased task complexity, there is an increase in visual search for action-relevant items during action preparation and less so at the level of task monitoring during action execution.

2.4 Latency of Fixations on Task-relevant ROIs

To further study the temporal aspects of gaze guidance behavior during action preparation and action execution, we were interested in the latency of the first fixations to the eight ROIs. As a measure of latency, we used the median time to first fixate on an ROI in each action epoch per trial. We hypothesized that if fixations are made to the task-relevant regions per a set cognitive scheme, we would see a strict temporal window where these fixations occur in the preparation and execution epochs. A cognitive schema that involves first searching for the targets, directing the hand towards the target, locating the next target, etc, would lead to first fixations on the 'other objects', then on the current target object, and then onto the next target object. On the other hand, if fixations do not occur based on a set cognitive schema behind it, the occurrence of fixations would not follow a temporal structure. Furthermore, to make the latency comparable across the different durations of preparation and execution epochs, we normalized it by the duration of the given epoch. Hence, the normalized latency of the first fixation to the eight ROI provides a temporal measure of when a schema is initiated during the preparation and execution of an action.

Here, we describe the latency of the first fixations on the eight ROIs normalized by the epoch duration. Figure 4A shows the distributions of the first fixations on the eight ROIs in the preparation phase. The median first fixation on the previous target shelf occurred at $0.09(SD = \pm 0.11)$ after epoch start in the EASY trials and at $0.07(SD = \pm 0.09)$ in the HARD trials. Similarly, the median first fixation on the previous target object occurred at $0.11(SD = \pm 0.13)$ in the EASY trials and at 0.06 $0.06(SD = \pm 0.09)$ in the HARD trials after the epoch start. The fixations then occurred at $0.18(SD = \pm 0.12)$ over the 'other shelves' ROI in EASY trials and at $0.13(SD = \pm 0.09)$ in the HARD trials. The fixations then occurred at $0.19(SD = \pm 0.09)$ over the 'other objects' ROI in EASY trials and at $0.13(SD = \pm 0.09)$ in the HARD trials. Subsequently, fixations occurred on the next target object at $0.39(SD = \pm 0.21)$ in the EASY trials and at $0.33(SD = \pm 0.19)$ in the HARD trials. Similarly, there were first fixations on the next target shelf at $0.39(SD = \pm 0.24)$ in the EASY trials and at $0.39(SD = \pm 0.23)$ in the HARD trials. First fixations on the current target shelf occurred at $0.40(SD = \pm 0.24)$ in EASY trials and at $0.40(SD = \pm 0.20)$ in the HARD trials. Finally, first fixations on the current target object were at $0.42(SD = \pm 0.13)$ in EASY trials and at $0.46(SD = \pm 0.11)$ in the HARD trials. As shown in Figure 4A, the distribution of the fixation latencies shows a structured sequence of attending to task-relevant ROIs. Notably, the first fixation on the current target object occurs only after 40% of the preparation epoch duration has elapsed.

We modeled the latency of first fixations dependent on ROI type and trial type using linear mixed models. Table 5 details the model coefficients of the linear regression model that compared the latency of the first fixation on the current target object with the rest of the ROIs and their interactions with respect to the task complexity as described in section 5.5.4. The fixed effects show that the latency of the first fixations on the ROIs significantly differed from the latency on the current target object. Notably, the first fixations go from the previous target object and shelf to the other shelves and objects, onto the next targets, and then finally to the current target object. Furthermore, the interactions show that task complexity significantly affected the latency of first fixations on the previous target object and shelf, as well as the other

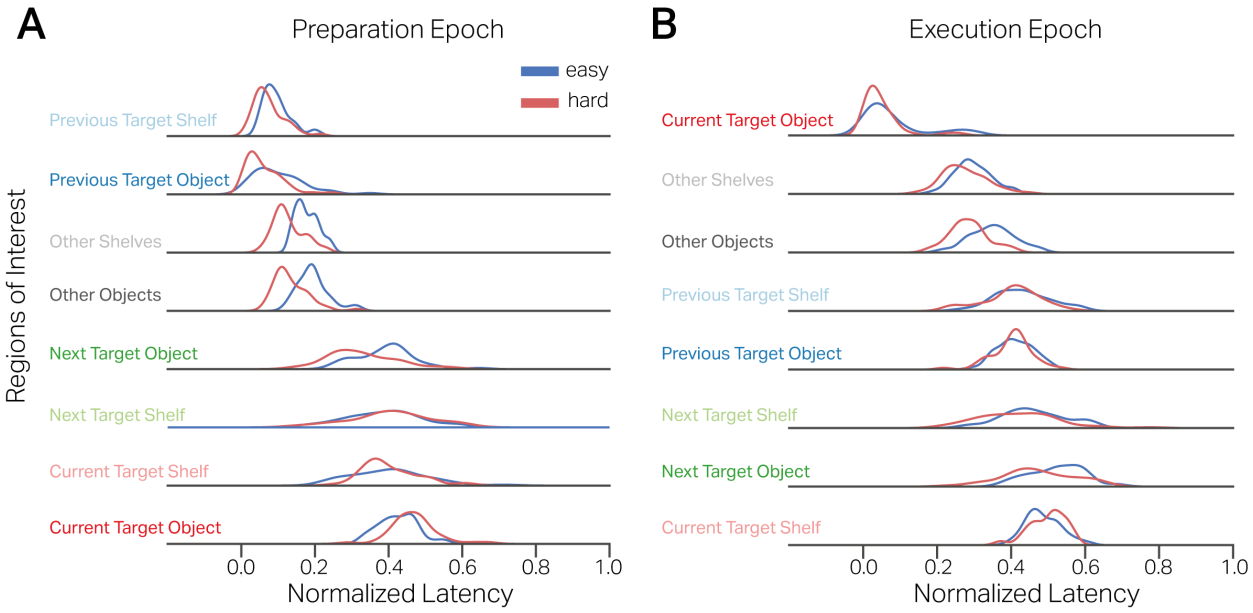

Figure 5: Distributions of median time to first fixation on the 8 ROIs for the action preparation and execution and trial types. The latency (abscissa) is normalized by the length of the preparation and execution epoch, respectively. The ordinate is sorted in ascending order of mean latency from epoch start.

and next target object, compared to the current target object. In HARD tasks, the first fixations on previous targets and the 'other' ROIs occurred significantly earlier than in the EASY tasks. Taken together, irrespective of task complexity, the action preparation epochs show a systematic progression of fixations from one ROI to another. This structured temporal sequence of fixations with a defined temporal window shows the fixations on the action-relevant ROIs are not incidental and are part of a set cognitive schema. Moreover, given the task complexity, the temporal profiles of these locating/searching fixations can change and occur earlier. Most importantly, the first fixations on the current target object occur last in sequence after more than 40% of the epoch time has elapsed, lending strong evidence to the just-in-time manner in which gaze is allocated to the immediate action requirements.

We similarly analyzed the latency of first fixations on the ROIs during the action execution epochs. Figure 4B summarizes the distribution of the latency of first fixations on the different ROIs during the action execution epoch. The first fixation on the current target object occurred at $0.1(SD = \pm 0.14)$ in the EASY trials and at $0.05(SD = \pm 0.08)$ in the HARD trials. Similarly, the first fixation on the 'other shelves' ROI occurred at $0.3(SD = \pm 0.12)$ in the EASY trials and at $0.27(SD = \pm 0.10)$ in the HARD trials after the epoch start. Shortly after, the fixations were on 'other objects' ROIs at $0.35(SD = \pm 0.16)$ in the EASY trials and at $0.28(SD = \pm 0.11$ in the HARD trials. The fixations occurred at $0.41(SD = \pm 0.17)$ on the previous target object in EASY trials and $0.40(SD = \pm 0.13)$ in the HARD trials. Subsequently, fixations occurred on the previous target shelf at $0.42(SD = \pm 0.21)$ in the EASY trials and at $0.40(SD = \pm 0.17)$ in the HARD trials. Similarly, the first fixations on the next target shelf occurred at $0.45(SD = \pm 0.24)$ in the EASY trials and at $0.41(SD = \pm 0.23)$ in the HARD trials. Fixations occurred on the current target shelf at $0.49(SD = \pm 0.11)$ in EASY trials and $0.50(SD = \pm 0.11)$ in the HARD trials. Finally, the first fixations on the next target object were at $0.52(SD = \pm 0.23)$ in EASY trials and at $0.47(SD = \pm 0.20)$ in the HARD trials. During action execution, we see a similar temporal structure of first fixations that go from 'other' objects and shelves to previous targets and have similar latencies on the next target object and the current target shelf. Here again, the fixations on the current target shelf occur after half of the epoch duration has elapsed.

As before, we used a linear mixed model to model the latency of the first fixations dependent on the ROI type and trial type. Table 5 shows the model coefficients of the linear model that compared the latency of the first fixation on the current target object with the rest of the ROIs and their interactions with respect to the task complexity in the action execution epoch. The fixed effects reveal that the difference in the latency of first fixations on all the ROIs was significantly different from the occurrence of the first fixation on the current target object. The fixations go from the current target object to the other objects and shelves, shortly followed by the previous targets and then in close succession from the next targets to the current target shelf. Furthermore, there is a small interaction of trial type wherein latency between the current target shelf and the current target object is significantly increased in the HARD trials. Specifically, task complexity did not affect the latency of fixations on the other seven ROIs. To summarize, irrespective of the task complexity, there is a systematic sequence of first fixations on the eight ROIs during action execution. The

Table 5: Fixed effects coefficients of the linear mixed model detailing estimated latency of first fixations w.r.t the current target object during the action preparation and execution

Model:

$$latency \sim 1 + trial_type * ROI_type$$
$$+(1 + trial_type + ROI_type|Subject)$$

	Estimate	95% CI	t	p
Action preparation Epochs				
Latency w.r.t Current Target Object				
Previous Target Shelf	-0.36	[-0.38, -0.34]	-35.49	**< 0.001*****
Previous Target Object	-0.36	[-0.39, -0.34]	-29.64	**< 0.001*****
Other Shelves	-0.29	[-0.31, -0.27]	-32.24	**< 0.001*****
Other Objects	-0.28	[-0.30, -0.26]	-31.05	**< 0.001*****
Next Target Object	-0.08	[-0.10, -0.06]	-7.85	**< 0.001*****
Next Target Shelf	-0.05	[-0.08, -0.03]	-3.97	**< 0.001*****
Current Target Shelf	-0.04	[-0.07, -0.02]	-3.25	**0.001*****
Interactions				
Trial Type : Previous Target Shelf	-0.06	[-0.09, -0.03]	-3.94	**<0.001*****
Trial Type : Previous Target Object	-0.08	[-0.11, -0.05]	-4.62	**<0.001*****
Trial Type : Other Shelves	-0.09	[-0.11, -0.06]	-5.72	**<0.001*****
Trial Type : Other Objects	-0.10	[-0.13, -0.07]	-6.55	**0.001****
Trial Type : Next Target Object	-0.10	[-0.13, -0.07]	-6.22	**0.001*****
Trial Type : Next Target Shelf	-0.03	[-0.06, -0.00]	-1.74	0.08
Trial Type : Current Target Shelf	-0.04	[-0.07, -0.01]	-2.42	**0.016***
Action Execution Epochs				
Latency w.r.t Current Target Object				
Other Shelves	0.22	[0.19, 0.26]	13.18	**< 0.001 *****
Other Objects	0.26	[0.22, 0.29]	14.97	**< 0.001 *****
Previous Target Object	0.34	[0.31, 0.38]	21.50	**< 0.001 *****
Previous Target Shelf	0.35	[0.32, 0.38]	23.35	**< 0.001 *****
Next Target Shelf	0.37	[0.34, 0.41]	23.30	**< 0.001 *****
Current Target Shelf	0.43	[0.40, 0.46]	27.77	**< 0.001 *****
Next Target Object	0.44	[0.40, 0.47]	25.01	**< 0.001 *****
Interactions				
Trial Type : Other Shelves	0.03	[-0.01, 0.07]	1.43	0.15
Trial Type : Other Objects	-0.01	[-0.06, 0.03]	-0.68	0.49
Trial Type : Previous Target Object	0.04	[0.00, 0.08]	1.94	0.053
Trial Type : Previous Target Shelf	0.03	[-0.02, 0.07]	1.20	0.22
Trial Type : Next Target Shelf	0.003	[-0.04, 0.05]	0.17	0.86
Trial Type : Current Target Shelf	0.06	[0.02, 0.10]	3.04	**0.002****
Trial Type : Next Target Object	0.004	[-0.04, 0.05]	0.20	0.84

early fixations on the previous target object and shelf indicate these fixations are related to monitoring/checking the current action execution with respect to the previously handled object. Similarly, later fixations on the next target object and shelf indicate fixations used for planning the next action even before the action has been completed. Importantly, the increased task complexity did not affect the latency of first fixations on the task-relevant ROIs during action execution. Here again, the first fixation on the current target shelf, i.e., where the target object is placed, is fixated towards the end after 50% of the time has elapsed, strongly indicating a just-in-time manner of the action-relevant fixations.

Taken together, the latency of first fixations on the ROIs revealed the activation of specific cognitive schemas and their relative time of occurrence in the action preparation and execution epochs. In the preparation epochs, early fixations on the 'other' objects and shelves indicate a visual search for target objects, closely followed by fixations on the next

target object and finally on the current target object. In the execution epochs, a similar temporal structure is found with early fixations on 'other' objects and shelves, indicating again a visual search, followed by fixations on the previous target, which suggests task monitoring, and then on the current target shelf. Task complexity affected the timing of first fixations, primarily during action preparation and not in the execution phase. Importantly, first fixations on the current targets occur after almost half the epoch time has elapsed, showing a just-in-time manner of gaze allocation. Supplementing these results with the average fixation profile from the previous section, we can conclude that the first half of the preparation and execution epochs is dedicated to locating the target and the later part to directing the hand and body to those targets.

The overall results suggest that action sequences are predominantly decided in a just-in-time manner for novel tasks, and task complexity influences distinct aspects of action-oriented fixations. The average fixation profile over the ROIs showed that fixations are driven just in time from previous to current and next actions. Task complexity affected the proportion of fixations early during the preparation period to facilitate the visual search for targets and during the execution period for increased action monitoring. The sequential gaze transitions between ROI pairs further demonstrated an increased visual search associated with task complexity in the action preparation period, while gaze guidance did not differ during the execution phase. Task complexity also affected the latency of the first fixations on the ROIs in the preparation epochs but not in the execution epochs. Importantly, the latency analysis emphasized the just-in-time nature of gaze guidance, where fixations are made to the immediate action targets after half the action epoch (both preparation and execution) has passed. Finally, the increased visual search process during the action preparation increased the time to finish complex tasks. However, it was not associated with optimal action choices. Instead, participants chose arbitrary spatial heuristics to possibly constrain the search space. Thus, gaze behavior is primarily controlled by low-level cognitive schemas of searching, locating, directing, monitoring, etc., in the context of novel tasks.

3 Discussion

In the present study, we investigated the spatiotemporal aspects of gaze control specific to the preparation of proactive actions and their execution. Our study shows crucial characteristics of natural cognition in choosing relevant actions and the corresponding gaze control under varying cognitive load. Increased cognitive load was associated with a disproportionate increase in the duration of planning the actions and less so with executing them. The average fixations aligned to the action onset revealed gaze guidance was driven just in time. Given the large behavioral differences induced by task complexity, the corresponding action-aligned average fixations differed moderately. These differences indicate an activation of visual search for targets early in the action preparation period and enhanced action monitoring concurrent with the action execution. This observation was validated further by studying the gaze transition behavior. The latency of first fixations showed gaze guidance to be just in time with overt actions. Task complexity also affected the latency of the first fixations on the task-relevant regions in the action preparation phase and not during the execution phase. Again, an increase in task complexity was associated with the activation of an early visual search for targets. Moreover, the substantial time dedicated to action planning and an increase in visual search behavior did not correspond with more optimal action choices. Instead, we found subjects used ad-hoc spatial heuristics to reduce the solution search space. Thus, our findings reveal an action-oriented structure of eye movements that primarily correspond to the activation of low-level cognitive schemas of locating target objects, directing the body and hands to those targets, and finally monitoring the execution of the manual action.

The central aim of our study is to generalize the cognitive mechanisms of gaze control to tasks that are not over-learned and routine. Building on the work of Land et al. (1999) where eye movements were studied while preparing tea, the tasks in our study are novel in the sense that they do not have an inherent action sequence associated with them. Our experimental setup provided a way to capture oculomotor behavior for tasks with varying complexity that did not have a strict action sequence. By studying the dispersion and timing of fixations on the previous, current, and next action-relevant regions, we show a structured sequence of fixations that may be classified based on previous work. Moreover, we add to the current body of research by providing a data-driven approach to studying locating, directing fixations, guiding fixations, and checking fixations. Our analysis and findings generalize the occurrence of these fixations explicitly to the action sequence by doing away with object identity completely.

Land et al. (1999) proposed the various functions of eye movements from locating, directing, guiding, and checking during a tea-making task. Our study also shows the general signature of these fixations time-locked to the action onset and the impact of cognitive load. Importantly, our study shows the occurrence of visual search early in the action preparation phase with increased search due to increased task complexity. The directing and guiding fixations occurred just-in-time of the action onset and were not affected by task complexity. The checking/monitoring fixations occurred concurrently with the action execution as fixations on the previous action targets. Considering the latency and the

proportional change in fixations on the task-relevant regions, our study further corroborates the cognitive schemas that sequentially support action preparation and execution.

Land and Furneaux (1997) additionally elaborated on the schemas that direct both eye movements and motor actions. They proposed a chaining mechanism by which one action leads to another, where the eyes supply constant information to the motor system via a buffer. In our analysis, the temporal structure of the first fixations on the ROIs lends further evidence of such a cognitive buffer that facilitates moving from one action to another where eye movements search and locate the regions necessary for current and (sometimes) future actions while also monitoring the task with previous action targets. Interestingly, the increased cognitive load did not affect this simple mechanism of cognitive arbitration where there is pronounced reliance on explicit visual search instead of encoding objects in memory. Thus, we show that cognitive schemas are activated in quick succession to inform the current action with scant use of the visual working memory.

Even though participants in our study spent an inordinate amount of time before an action onset, they were strikingly sub-optimal compared to the shortest path to the solutions. There is considerable psychological evidence that supports the idea that anticipated cognitive demand plays a significant role in behavioral decision-making. Further, all things being equal, humans show a propensity towards actions that reduce the cost of physical and cognitive effort (Kool et al., 2010). Studies in human performance in reasoning tasks and 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 implicitly search the problem space to plan the moves without executing them. Hence, longer task durations do not lead to shorter solution paths. Instead, they showed that humans break the problem down into component sub-tasks, giving rise to a low-complexity relationship between the problem size and time to solution. Further, Pizlo and Li (2005) showed that subjects use simpler heuristics to decide the next move. This is in line with Gigerenzer and Goldstein (1996) who posit that humans use simplifying strategies to gather and search information that results in sub-optimal behavior. In our study, participants used a spatial heuristic by arbitrarily organizing the objects from left to right, which might be an artifact of reading direction habit (Afsari et al., 2016). In more complex tasks, we observed higher planning durations and greater object displacements, but the action execution duration was similar to less complex tasks. Thus, participants in our study do not necessarily demonstrate a lack of planning. Instead, the results show they divided the overall task into sub-tasks and planned which object needed to be handled and where it should be moved.

From the perspective of embodied cognition (Wilson, 2002; Ballard et al., 2013; Van der Stigchel, 2020), humans off-load cognitive work on the environment, and the environment is central to the cognitive system. In this framework, cognition is situated and serves to produce actions. Draschkow et al. (2021) recently showed that humans do not rely much on working memory even when searching for objects is physically effortful. Our behavioral results show that subjects use spatial heuristics to reduce the search space and complete the tasks. This is indicative of exploiting the external world to reduce the cognitive effort of encoding object locations in memory and relying on repeated visual search to accomplish the task goals. Further, Droll and Hayhoe (2007) have suggested that a just-in-time strategy is also likely used to lower the cognitive cost of encoding objects in the world into the visual working memory. Thus, the prevalence of the just-in-time strategy observed in our study is further evidence of their use of the environment as an external memory.

We conducted this study in VR, which is a relatively new technology and comes with its own set of complications. The present state of VR has opened new directions of research in naturalistic human behavior. However, with the natural eye, head, and body movements, the experimental data can be plagued with noise due to the slippage of the head-mounted display (Niehorster et al., 2020) and other body sensors. To reduce such errors, we calibrated the eye tracker at regular intervals. We also calibrated the body-tracking sensors and the hand controller at the beginning of a trial and then halfway through the experiment. Moreover, participants in our study used VR hand controllers to grasp and move objects that could have seemed unnatural. To help them get accustomed to the VR environment, we gave them practice trials at the beginning of the experiment. Itaguchi (2021) showed that grasping movements in VR do not differ quantitatively from grasping kinematics in the real world. Consequently, even with its shortcomings, this work can find easy translation to real-world paradigms.

4 Conclusion

In conclusion, in the present study, we studied the nature of proactive action choices and the corresponding oculomotor control under novel and complex task conditions. Our results generalize past work on action-oriented eye movements to tasks that are not routine or over-learned. Our generalized approach explicitly investigates the flow of gaze associated with the action sequence. Our results show that action-oriented gaze primarily corresponds to low-level cognitive schemas such as searching and locating task-relevant objects of interest, directing and guiding the body and hands, and

monitoring the task progression. Furthermore, we show that these action schemas do not necessarily support optimal actions but rather only the production and execution of actions. More importantly, we show that fixations are tightly coupled with overt behavior and engaged in a just-in-time manner.

5 Methods

5.1 Participants

A total of 55 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 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 that could affect them in the virtual environment. After obtaining their informed consent, we briefed them on the experimental setup and task. The Ethics Committee of the University of Osnabrück approved the study (Ethik-37/2019).

5.2 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] with 120 Hz sampling frequency and $0.5° - 1.1°$ calibration error. With their right hand, participants used an HTC Vive controller to manipulate the objects during the experiment. The HTC Vive Lighthouse tracking system provided positional and rotational tracking and was calibrated for a 4m x 4m space. We used the 5-point calibration function provided by the manufacturer to calibrate the gaze parameters. To ensure the calibration error was less than $1°$, we performed a 5-point validation after each calibration. Due to the study design, which allowed many natural body movements, the eye tracker was calibrated repeatedly during the experiment after every three trials. Furthermore, subjects were fitted with HTC Vive trackers on both ankles and elbows and one on the midriff. The body trackers were calibrated subsequently to give a reliable pose estimation using inverse kinematics of the subject in the virtual environment. We designed the experiment using the Unity3D[2] version 2019.4.4f1 and SteamVR game engine and controlled the eye-tracking data recording using HTC VIVE Eye-Tracking SDK SRanipal[3] (v1.1.0.1).

The experimental setup consisted of 16 different objects placed on a shelf of a 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 a width of 20cm. The shelf was designed with a height and width of 2m with five 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 to indicate the end of the task.

5.3 Experimental Task

Subjects performed two practice trials where they familiarized themselves with handling the VR controller and the experimental setup. In these practice trials, they could explore the virtual environment and freely displace the objects. After the practice trials, subjects were asked to sort objects based on the object features. There were two types of trials: EASY and HARD. Subjects were not limited by time to complete the task. The participants were not explicitly told about the task category and were told to follow the instructions given to them at the beginning. Each subject performed 24 trials in total. Each trial type (as listed below) was randomly presented twice throughout the experiment. The experimental setup is illustrated in Figure 1B. 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

[1] https://enterprise.vive.com/us/product/vive-pro-eye-office/
[2] Unity, www.unity.com
[3] SRanipal, developer.vive.com/resources/vive-sense/sdk/vive-eye-tracking-sdk-sranipal/

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.

5.4 Data pre-processing

5.4.1 Gaze Data

The data preprocessing steps are illustrated in Figure 1C. As a first step, using the 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. Each sample's 3d gaze direction vector 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:

$$\theta_{\text{h}} = \frac{180}{\pi} * \arctan \frac{x}{z}$$
$$\theta_{\text{v}} = \frac{180}{\pi} * \arctan \frac{y}{z}$$

Next, we calculated the angular velocity of the eye in both the horizontal and vertical coordinates by taking the first difference of the gaze angles with respect to time using the formula below:

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

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

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

We used an adaptive threshold method for saccade detection described by Voloh et al. (2020) to filter the samples where gaze was relatively stable. We selected an initial saccade velocity threshold θ_0 of $200°/s$. All eye movement samples with an angular velocity less than θ_0 were used to compute a new threshold θ_1 as three times the median absolute deviation of the filtered samples. θ_1 was used as the new threshold to filter samples and the above process was repeated until the difference between θ_{n} and $\theta_{\text{n+1}}$ 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 saccades. To handle minuscule fixations and saccades, we labeled all samples with saccade duration less than 30 ms bracketed by fixations also as a fixation and merged the fixations before and after this interval. Similarly, we labeled all fixation samples with a duration of less than 50 ms, interspersed between two saccades, also as a saccade. Following this, we recalculated the fixation and saccade durations. This procedure suitably handles the misclassification of single samples or pairs of samples into fixations or saccades. Finally, we rejected all fixations with a duration greater than 3.5 times the median absolute deviation of the population fixation duration and fixations that were less than 100 ms long.

5.4.2 Grasp 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 with a default value of FALSE and set to TRUE when a grasp was initiated and then reset when the trigger was released to end the grasp. Using the position of the controller in the world space, we determined the locations from the shelf where a grasp was initiated and ended. We also removed trials where the controller data showed implausible locations in the world space. These faulty data

can be attributed to sensors losing tracking during the experiment. Next, we removed grasping periods where the beginning and final locations of the objects on the shelf were the same. We calculated the interquartile range (IQR) of the participants' number of object displacements for the two trial types (EASY and HARD). To remove the outlying object displacements in the trials, we removed the trials with 1.5 times the IQR of object displacements. We also removed those trials with fewer than three object displacements.

After pre-processing, we were left with data from 48 subjects with 999 trials in total and 20.8 ± 1.8 trials per subject. Furthermore, we had eye-tracking data corresponding to $12,403$ object displacements with 12.3 ± 5.11 object displacements per trial.

Using the grasp onset and end events, we divided the continuous data stream into action preparation and execution epochs. Specifically, the action preparation phase starts from the end of a previous grasp onset till the start of the immediate grasp onset. Similarly, the execution epoch started from the start of the current grasp onset to the end of that grasp. The division of the data into these epochs is depicted in Figure 1C.

We further organized the data to examine the spatial and temporal characteristics of eye movements during the action preparation and execution epochs. We labeled the fixation based on its location on the different objects and shelf locations into eight regions of interest (ROIs). These included fixations of targets in the previous, current, and next grasping actions. Specifically, the previous target object refers to the object that was handled in the previous action, and the previous target shelf is the shelf where the previous target object was placed. Similarly, the current target object refers to the object that is picked up and placed on the target shelf in the current action and the next target object and next target shelf in the immediately following action in the sequence. All other fixated regions that did not conform to the above six ROIs were categorized as 'other objects' and 'other shelves' that are irrelevant to the action sequence. In this format, we could parse the sequence of fixations on the eight ROIs that are relevant for preparing and executing the object-related actions.

5.5 Data Analysis

5.5.1 Influence of Task Complexity on Behavior

We used four measures to examine the grasping behavior during the experiment. First is trial duration, i.e., the total time to finish the tasks. Here, we tested the hypothesis that task complexity affected the total duration of finishing the tasks using a two-sample independent t-test. Second is the action preparation duration, which is the time between dropping the previous object and picking up the current target object. Third is execution duration, i.e., the time to move the target object to the desired location on the shelf. We tested the hypothesis that the action epochs durations differed based on the action epoch type and the task complexity. Finally, we were interested in the number of object displacements made by the participants to complete the tasks. We tested the hypothesis that participants were optimal in their action choices in the EASY and HARD conditions. Taken together, these measures capture the influence of task complexity and novelty on overt behavior.

We modeled the relationship between the action epoch duration dependent on the action epoch type (PREPARATION, EXECUTION) and the trial type (EASY, HARD). As the durations differed from the start of the trial to the end, we added the grasp index as a covariate in the model. All within-subject effects were modeled with random intercepts and slopes grouped by subject. The dependent variable was log-transformed so that the model residuals were normally distributed. The categorical variables trial type and epoch type were effect coded (Schad et al., 2020) so that the model coefficients could be interpreted as main effects. The model coefficients and the 95% confidence intervals were back-transformed to the response scale. The model fit was performed using restricted maximum likelihood (REML) estimation (Corbeil and Searle, 1976) using the lme4 package (v1.1-26) in R 3.6.1. We used the L-BFGS-B optimizer to find the best fit using 20000 iterations. Using the Satterthwaite method (Luke, 2017), we approximated degrees of freedom of the fixed effects. The full model in Wilkinson notation (Wilkinson and Rogers, 1973) is denoted as:

$$\begin{aligned} Log(duration) \sim 1 + trial_type * epoch_type * grasp_index \\ +(1 + trial_type + epoch_type + grasp_index | Subject) \end{aligned}$$

Similarly, we modeled the relationship between the number of moves made by subjects in each trial dependent on the trial type (EASY, HARD) and the solver type (MODEL, HUMAN). The model solver computed the shortest path to the object sorting solution for each trial instantiation using a depth-first search algorithm. As before, all within-subject effects were modeled with random intercepts and slopes grouped by subject. The independent variables were also

effect-codded. The full model in Wilkinson notation (Wilkinson and Rogers, 1973) is denoted as:

$$ObjectDisplacements \sim 1 + trial_type * solver_type$$
$$+(1 + trial_type + solver_type|Subject)$$

5.5.2 Action Locked Gaze Control

We analyzed the average fixation behavior aligned with the action onset. For each grasp onset in a trial, we chose the period from 3s before grasp onset to 2s after. We divided this 5s period into bins of 0.25 seconds and calculated the number of fixations on the eight ROIs described above. For each time bin, we calculated the proportion of fixations on the ROIs per trial type (EASY, HARD). We used the cluster permutation method to find the time points where there were significant differences between EASY and HARD trials for a given ROI. Here, we use the t-statistic as a test statistic for each time-bin, where t is defined as:

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

where x is the mean difference between the trial types, 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 at which the p-value is 0.05 in the t-distribution. We first found the time bins where the t-value was greater than the threshold. Then, we computed the sum of the t-values for these clustered time bins, which gave a single value 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. This gave us the null distribution over which we compared the cluster mass derived from the real data. To account for the multiple independent comparisons for the eight ROIs, we considered the significant clusters to have a Bonferroni corrected p-value less than 0.006 (0.05/8). In the results, we report the range of the significant time bins for the eight ROIs for the two trial types and the corresponding p-values.

5.5.3 Gaze Transition Behavior

To examine the gaze transitions within the action preparation and execution epochs we created transition matrices that show the origin and destination locations of the fixations on the eight ROIs. We first used the steps described by Hooge and Camps (2013) to create the scan paths and the transition matrices. A series of fixations on the ROIs in a given action epoch were transformed into a 2D matrix (M) with eight rows as the origin of the fixation and eight columns as the destination. The diagonal of the matrix was set to zeros as we were not interested in re-fixations on the same ROI. Using matrix M, we calculated the net and total transitions from and to each ROI. For every transition matrix **M** per action preparation and execution epoch, the net (M_{net}), and total (M_{total}) transition are defined as follows:

$$M_{net} = M - M^T$$

$$M_{total} = M + M^T$$

If subjects make equal transitions between the ROI pairs in both directions, we can expect zero transitions in M_{net}. Conversely, with strong gaze guidance towards a particular ROI, we would expect more net transitions. Hence, using the net and total transitions per action epoch, we calculated the relative net transitions (R) as:

$$R = \frac{\sum M_{net}}{\sum M_{total}}$$

We then took the mean of the relative net transitions per trial to measure asymmetric gaze guidance behavior in that trial.

We modeled the linear relationship of the relative net transitions dependent on the trial type (EASY, HARD), epoch type (preparation, execution), and their interactions. All within-subject effects were modeled with random intercept and slopes grouped by subject. The categorical variables trial type and epoch type were effect-coded. The full model in Wilkinson notation (Wilkinson and Rogers, 1973) is defined as:

$$Relative_{Transitions} \sim 1 + trial_type * epoch_type$$
$$+(1 + trial_type * epoch_type|Subject)$$

5.5.4 Latency of First Fixationson Task-relevant ROIs

Finally, we also calculated the latency of the first fixation on the eight ROIs from the start of the preparation and execution epochs. We computed the median time to first fixate on an ROI per trial. As the action preparation and execution epochs varied in duration, we normalized the time points by dividing them by the duration of the epoch. Thus, the normalized elapsed time from the start of an epoch is comparable to all epochs across trials and subjects.

We modeled the linear relationship of the normalized median time to the first fixation dependent on the trial type (EASY, HARD) and the eight ROIs and their interactions. We computed two models for the action preparation and execution epochs. All within-subject effects were modeled with random intercepts and slopes grouped by subject. The categorical variables trial_type and ROI were effect coded as before, and the model coefficients could be interpreted as main effects. For both models, we chose the latency of the first fixation on the current target object as the reference factor so that the latency of the fixation of all other ROIs could be compared to it. The full model in Wilkinson notation (Wilkinson and Rogers, 1973) is defined as:

$$Fixation_{time} \sim 1 + trial_type * ROI$$
$$+(1 + trial_type * ROI|Subject)$$

Author Contributions

AK, PK, TS: conceived and designed the study. TS, PK: Procurement of funding. HN, KI: programmed the experiment in virtual reality. programmed the LeapMotion study. AK: data collection. AK, FNN: data preprocessing. AK: data analysis. AK: initial draft of the manuscript. AK, FNN, KI, TS, PK: revision and finalizing the manuscript. All authors contributed to the article and approved the submitted version.

Data availability statement

The experimental data can be found at `https://osf.io/gq6aj/`. The analysis code can be found at `https://github.com/ashimakeshava/ergovr_gaze_guidance`

Acknowledgement

We are grateful for the financial support from 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. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Z. Afsari, J. P. Ossandón, and P. König. The dynamic effect of reading direction habit on spatial asymmetry of image perception. *Journal of vision*, 16(11):8, Sept. 2016. ISSN 1534-7362. doi: 10.1167/16.11.8. URL `http://dx.doi.org/10.1167/16.11.8`.

D. H. Ballard, M. M. Hayhoe, and J. B. Pelz. Memory representations in natural tasks. *Journal of cognitive neuroscience*, 7(1):66–80, 1995. ISSN 0898-929X. doi: 10.1162/jocn.1995.7.1.66. 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. *Multisensory research*, 26(1-2):177–204, 2013. ISSN 2213-4794. doi: 10.1163/22134808-00002414. URL `http://dx.doi.org/10.1163/22134808-00002414`.

A. Clark. *Being There: Putting Brain, Body, and World Together Again*. MIT Press, Jan. 1998. ISBN 9780262260527. URL https://play.google.com/store/books/details?id=i03NKy0ml1gC.

V. Clay, P. König, and S. U. 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: a journal of statistics for the physical, chemical, and engineering sciences*, 18(1): 31–38, Feb. 1976. ISSN 0040-1706. doi: 10.1080/00401706.1976.10489397. URL https://www.tandfonline.com/doi/abs/10.1080/00401706.1976.10489397.

D. Draschkow, M. Kallmayer, and A. C. Nobre. When natural behavior engages working memory. *Current biology: CB*, 31(4):869–874.e5, Feb. 2021. ISSN 0960-9822, 1879-0445. doi: 10.1016/j.cub.2020.11.013. URL http://dx.doi.org/10.1016/j.cub.2020.11.013.

J. A. Droll and M. M. Hayhoe. Trade-offs between gaze and working memory use. *Journal of experimental psychology. Human perception and performance*, 33(6):1352–1365, Dec. 2007. ISSN 0096-1523. doi: 10.1037/0096-1523.33.6.1352. URL http://dx.doi.org/10.1037/0096-1523.33.6.1352.

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. ISSN 1364-6613, 1879-307X. doi: 10.1016/j.tics.2013.03.006. URL http://dx.doi.org/10.1016/j.tics.2013.03.006.

G. Gigerenzer and D. G. Goldstein. Reasoning the fast and frugal way: Models of bounded rationality. *Psychological review*, 103(4):650–669, Oct. 1996. ISSN 0033-295X, 1939-1471. doi: 10.1037/0033-295X.103.4.650. URL https://psycnet.apa.org/fulltext/1996-06397-002.pdf.

M. Hayhoe and D. Ballard. Eye movements in natural behavior. *Trends in cognitive sciences*, 9(4):188–194, Apr. 2005. ISSN 1364-6613. doi: 10.1016/j.tics.2005.02.009. URL http://dx.doi.org/10.1016/j.tics.2005.02.009.

M. M. Hayhoe, A. Shrivastava, R. Mruczek, and J. B. Pelz. Visual memory and motor planning in a natural task. *Journal of vision*, 3(1):49–63, 2003. ISSN 1534-7362. doi: 10.1167/3.1.6. URL http://dx.doi.org/10.1167/3.1.6.

I. Hooge and G. Camps. Scan path entropy and arrow plots: capturing scanning behavior of multiple observers. *Frontiers in psychology*, 4:996, Dec. 2013. ISSN 1664-1078. doi: 10.3389/fpsyg.2013.00996. URL http://dx.doi.org/10.3389/fpsyg.2013.00996.

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. doi: 10.3389/frvir.2021.712378. URL https://www.frontiersin.org/article/10.3389/frvir.2021.712378.

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. ISBN 9781450371346. doi: 10.1145/3379156.3391338. URL https://doi.org/10.1145/3379156.3391338.

A. Keshava, N. Gottschewsky, S. Balle, F. N. Nezami, T. Schüler, and P. König. Action affordance affects proximal and distal goal-oriented planning. *The European journal of neuroscience*, 57(9):1546–1560, May 2023. ISSN 0953-816X, 1460-9568. doi: 10.1111/ejn.15963. URL http://dx.doi.org/10.1111/ejn.15963.

P. König, A. Melnik, C. Goeke, A. L. Gert, S. U. König, and T. C. Kietzmann. Embodied cognition. In *2018 6th International Conference on Brain-Computer Interface (BCI)*, pages 1–4. ieeexplore.ieee.org, Jan. 2018. doi: 10.1109/IWW-BCI.2018.8311486. URL http://dx.doi.org/10.1109/IWW-BCI.2018.8311486.

W. Kool, J. T. McGuire, Z. B. Rosen, and M. M. Botvinick. Decision making and the avoidance of cognitive demand. *Journal of experimental psychology. General*, 139(4):665–682, Nov. 2010. ISSN 0096-3445. doi: 10.1037/a0020198. URL http://dx.doi.org/10.1037/a0020198.

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. ISSN 0301-0066. doi: 10.1068/p2935. URL http://dx.doi.org/10.1068/p2935.

M. F. Land. Eye movements and the control of actions in everyday life. *Progress in retinal and eye research*, 25(3): 296–324, May 2006. ISSN 1350-9462. doi: 10.1016/j.preteyeres.2006.01.002. URL http://dx.doi.org/10.1016/j.preteyeres.2006.01.002.

M. F. Land and S. Furneaux. The knowledge base of the oculomotor system. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 352(1358):1231–1239, Aug. 1997. ISSN 0962-8436. doi: 10.1098/rstb.1997.0105. URL http://dx.doi.org/10.1098/rstb.1997.0105.

M. F. Land and M. Hayhoe. In what ways do eye movements contribute to everyday activities? *Vision research*, 41 (25-26):3559–3565, 2001. ISSN 0042-6989. doi: 10.1016/s0042-6989(01)00102-x. URL http://dx.doi.org/10.1016/s0042-6989(01)00102-x.

M. F. Land and P. McLeod. From eye movements to actions: how batsmen hit the ball. *Nature neuroscience*, 3(12): 1340–1345, Dec. 2000. ISSN 1097-6256. doi: 10.1038/81887. URL http://dx.doi.org/10.1038/81887.

S. G. Luke. Evaluating significance in linear mixed-effects models in R. *Behavior research methods*, 49(4):1494–1502, Aug. 2017. ISSN 1554-351X, 1554-3528. doi: 10.3758/s13428-016-0809-y. URL http://dx.doi.org/10.3758/s13428-016-0809-y.

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. ISSN 1932-6246. doi: 10.7771/1932-6246.1090. 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. *Journal of vision*, 19(14):28, Dec. 2019. ISSN 1534-7362. doi: 10.1167/19.14.28. 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. ISSN 1932-6203. doi: 10.1371/journal.pone.0043858. 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. *Current biology: CB*, 28(8):1224–1233.e5, Apr. 2018. ISSN 0960-9822, 1879-0445. doi: 10.1016/j.cub.2018.03.008. URL http://dx.doi.org/10.1016/j.cub.2018.03.008.

N. Mennie, M. Hayhoe, and B. Sullivan. Look-ahead fixations: anticipatory eye movements in natural tasks. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, 179(3):427–442, May 2007. ISSN 0014-4819. doi: 10.1007/s00221-006-0804-0. URL http://dx.doi.org/10.1007/s00221-006-0804-0.

J. Navarro, E. Hernout, F. Osiurak, and E. Reynaud. On the nature of eye-hand coordination in natural steering behavior. *PloS one*, 15(11):e0242818, Nov. 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0242818. URL http://dx.doi.org/10.1371/journal.pone.0242818.

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. ISSN 1554-351X, 1554-3528. doi: 10.3758/s13428-019-01307-0. URL http://dx.doi.org/10.3758/s13428-019-01307-0.

F. J. Parada and A. Rossi. Perfect timing: Mobile brain/body imaging scaffolds the 4e-cognition research program. *The European journal of neuroscience*, 54(12):8081–8091, May 2020. ISSN 0953-816X, 1460-9568. doi: 10.1111/ejn.14783. URL http://dx.doi.org/10.1111/ejn.14783.

J. B. Pelz and R. Canosa. Oculomotor behavior and perceptual strategies in complex tasks. *Vision research*, 41(25-26): 3587–3596, 2001. ISSN 0042-6989. doi: 10.1016/s0042-6989(01)00245-0. URL http://dx.doi.org/10.1016/s0042-6989(01)00245-0.

Z. Pizlo and Z. Li. Solving combinatorial problems: the 15-puzzle. *Memory & cognition*, 33(6):1069–1084, Sept. 2005. ISSN 0090-502X. doi: 10.3758/bf03193214. URL http://dx.doi.org/10.3758/bf03193214.

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, Feb. 2020. ISSN 0749-596X. doi: 10.1016/j.jml.2019.104038. URL https://www.sciencedirect.com/science/article/pii/S0749596X19300695.

F. Schumann, W. Einhäuser-Treyer, J. Vockeroth, K. Bartl, E. Schneider, and P. König. Salient features in gaze-aligned recordings of human visual input during free exploration of natural environments. *Journal of vision*, 8(14):12.1–17, Oct. 2008. ISSN 1534-7362. doi: 10.1167/8.14.12. URL http://dx.doi.org/10.1167/8.14.12.

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. *Journal of vision*, 12(13):19, Dec. 2012. ISSN 1534-7362. doi: 10.1167/12.13.19. URL http://dx.doi.org/10.1167/12.13.19.

S. Van der Stigchel. An embodied account of visual working memory. *Visual cognition*, 28(5-8):414–419, Sept. 2020. ISSN 1350-6285. doi: 10.1080/13506285.2020.1742827. URL https://doi.org/10.1080/13506285.2020.1742827.

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. ISSN 1995-8692. doi: 10.16910/jemr.12.8.3. 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):392–399, 1973. ISSN 0035-9254, 1467-9876. doi: 10.2307/2346786. URL http://www.jstor.org/stable/2346786.

M. Wilson. Six views of embodied cognition. *Psychonomic bulletin & review*, 9(4):625–636, Dec. 2002. ISSN 1069-9384. doi: 10.3758/bf03196322. URL http://dx.doi.org/10.3758/bf03196322.