Goal-directed Allocation of Gaze Reflects Situated Action Control in Dynamic Tasks

Nils Wendel Heinrich[1,3], Annika Österdiekhoff[2], Stefan Kopp[2], Nele Rußwinkel[1]
nils.heinrich@uni-luebeck.de
[1]Institute of Information Systems (IFIS), Department of Computer Science and Engineering, Universität zu Lübeck, Lübeck, Germany
[2]Social Cognitive Systems, Faculty of Technology, CITEC, Bielefeld University, Bielefeld, Germany
[3]Cognitive Psychology and Ergonomics, Technische Universität Berlin, Berlin, Germany

Abstract

How humans engage in goal-directed behavior within dynamic environments is still not completely understood. Pursuing goals in an environment that is characterized by constant unpredictable changes might be possible through the interaction of multiple layers of action control. A cognitive layer exerts situational control by selecting action intentions, while a motor control layer is responsible for execution. The motor layer informs the cognitive level, about disturbances during execution of these action intentions. We present an experimental dynamic environment, combining motor control manipulation and eye-tracking to investigate visuomotor grounding of cognitive processes. Our results indicate that inefficient motor control prompts strategic shifts in eye-movement behavior, with fixations closer to a reference point under moderate motor noise and further away under increased noise. We further find fixational and smooth pursuit eye movements that can be directly mapped to pursued action intentions. These findings shed light on the changes in action selection caused by noise in the motor system and can be used in a next step to investigate moment-to-moment changes in the pursuit of action intentions under inefficient motor control.

Keywords: situated action control; dynamic environments, control hierarchy; Sense of Control (SoC); sensorimotor integration; eye-movement control

Introduction

Situated action control is an intricate process that requires the interaction of various layers of control to enable goal-directed behavior within dynamic environments. Involved processes range from the selection of the action intention, over the decomposition of the intention into several motor commands, to their execution and thus the implementation of the action intention onto the environment (Grafton & Hamilton, 2007; Kahl et al., 2022). Each individual motor command is accompanied by a prediction, the sensorimotor consequence that is expected when the action is executed. The actual sensory feedback is then compared with the sensorimotor prediction (comparator model 3, Synofzik et al., 2008). If a mismatch between the two inputs is detected, i.e. if there is a prediction error, this negatively affects the feeling of being in control. This phenomenal experience is termed the Sense of Control (SoC) and helps the agent to identify self-produced changes in the environment.

Goal-directed behavior is particularly challenging within dynamic environments. These are characterized by continuously evolving in unpredictable ways. Nevertheless, humans excel at this through their ability to seamlessly adapt action intentions in response to evolving circumstances, at times even giving up on an intention to pursue another one. What enables us to do that? The given references postulate that the SoC might be the main contributor to that ability. It is a key experience informing us as agents about the current instance and our ability to bring about changes in it.

When experiencing changes in sensed control, the agent's behavior also changes. If the agent encounters difficulties in executing an action due to noise in motor control, leading to several prediction errors, it may adjust its original intentions to increase the chance of executing them successfully. Action intentions depend on the situation as it evolves, they are situated. Hence adapting the process of evaluating several possible options and finally selecting one is referred to as situational control (Pacherie, 2008). The interaction of these higher-level cognitive processes of action evaluation and selection with lower-level motor control processes is what is called situated action control. To our knowledge, however, there are no experiments in dynamic environments that investigate the effect of changes in action selection that stem from perturbing factors (predictable or uncertain) in motor control. We present here the newly developed Dodge Asteroids experimental environment (Heinrich et al., 2023) that features motor noise and situational complexity. Combining the Dodge Asteroids environment with eye tracking, we can assess action intentions that are allocated somewhere on the screen in relation to the agent as reference point. This experiment thus enables directly testing the changes in action intentions within dynamic environments and contributes to the evaluation and, if necessary, revision of models of action control already existent in literature.

Theoretical Background

The multiple layers contributing to situated action control refer to the different types of regulatory control namely situational control and motor control, both associated with a distinct SoC (Pacherie, 2008). They live on different levels of the control hierarchy. Situational control is ascribed to proximal planning (proximal intentions; Pacherie, 2008) and decision-making processes at higher cognitive levels, whereas motor control is exerted by the sensorimotor system (Figure 1). Humans are usually aware of their adaptation in planning and decision-making processes, while regulatory control at the motor level is automated and happens without the agent being aware (Hacker, 1986; Kahl et al., 2022). Kahl et al. (2022) precisely defined the internal processes of each level. At the top of the hierarchy, in the cognitive control layer, an action field is compiled that contains actions that

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

can be executed at that very moment to potentially solve the task at hand. Out of these action possibilities, one is selected, the action goal. The current SoC associated with higher-level situational control directly affects the composition of the action field and the selection process of the final action goal. For example, the action field under a decreased SoC will only consist of action possibilities that are particularly easy to implement, even with inefficient control. Likewise, an action goal is selected that is associated with a general high chance of success. The action goal is then given to the lower level in the hierarchy, the sensorimotor control layer. Here, the action goal is split into several detailed motor programs that ultimately lead to the implementation of the action goal, getting one step closer with every command. The sensorimotor prediction of each individual motor program and the sensory feedback perceived after its execution are entered into a sensorimotor grounded comparator model. The output of the comparator is the difference between the two inputs. A sensitivity range indicates at which size the difference will still be accepted as a match (Synofzik et al., 2008). If the size of the difference falls outside this range, a prediction error has occurred. In this case, the SoC relating specifically to motor control is reduced. On the contrary, matches will lead to an increase of the low-level SoC. Thus, only large prediction errors or several consecutive prediction errors will lead to a significant drop in sensed motor control. The hierarchy has an awareness boundary in the form of the cognitive control layer threshold. The agent becomes aware of its insufficient motor control when the low-level SoC drops below a certain value in turn decreasing the high-level SoC. This can ultimately lead to abandoning the current action goal, building a new action field, selecting a new action goal, and its subsequent execution by the sensorimotor control layer.

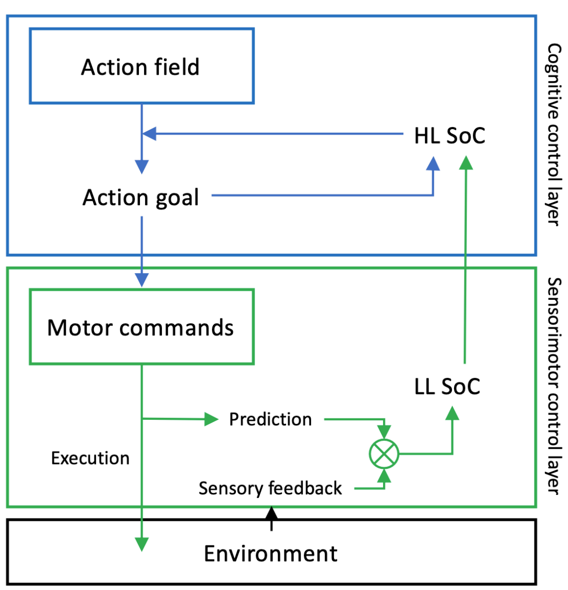

Figure 1: Depiction of two-layer architecture of the Kahl et al. computational model of situated action control (2022). Figure taken and adapted from Heinrich et al. (2023).

This description of the internal processes implies that the moment-to-moment changes of the SoC are therefore determined by overcoming the two threshold values. However, testing these dynamics is a challenge. To measure when exactly prediction errors occur, methods such as EEG or fMRI would probably have to be used. Further, when exactly regulatory control is applied at the cognitive level would have to be tested using sensitive behavioral measures. However, some things have to be addressed first: *how* is behavior regulated from the cognitive level? This is exactly what we aim to investigate. For this, we will draw from the Kahl et al. (2022) model and investigate changes in action selection by measuring the differences in the properties of action goals that are executed in environments imposing different levels of motor noise. Since the model implies that action selection processes are adjusted when the motor system constantly detects error signals, we should be able to determine exactly how situational control, regulatory control at the cognitive level, looks like by comparing the mean values of action goal properties.

Eye-Movement Control in Situated Action

The visual system is incredibly efficient at recognizing possibilities to act (Gibson, 1966). That means that visual perception is influenced by top-down processes that account for the motor repertoire of the agent (Gibson, 1977). At the same time, eye movements precede other bodily actions (Land, 2004, 2006). They guide goal-directed behavior directing attention towards crucial information needed by the motor system to prepare and execute the individual movements (Rothkopf et al., 2016; Wilmut et al., 2006). When grasping for an object for example, the eyes monitor how each individual motion reduces the distance between the hand and the object. Often peripheral vision will be enough to verify the accuracy of the grasping movement. Especially because humans can draw from other sensory information like proprioception as well. But the eyes will still be the most accurate sensors we can use for guiding manual movements. That is why we often glance shortly towards the object when reaching for it while facing something else. This is based on the neural coupling of eye and other bodily movements which has also been shown for steering in natural driving tasks (Marple-Horvat et al., 2005; Wilson et al., 2007), although in general eye-movement control is strongly adapted to the task and context at hand (Land & Hayhoe, 2001).

Following this line of reasoning, situational control at the cognitive level is grounded in eye-movement control. More specifically periods in which the eyes remain relatively still at a particular location within the environment may indicate action goals that are pursued. Additionally, contrary to the continuous stream of data corresponding to the position of an agent within an environment, eye movement data can be divided into individual events of fixations and smooth pursuits allowing us to assess individual action goals.

In this paper, we address the research question of whether key top-down processes involved in situated action control can be identified through visuomotor behavior. For this, we relate composing an action field to general visual exploration and pursuing an action goal to maintaining fixational and smooth pursuit eye movements of specific properties. Our assumption is therefore that if we combine a dynamic experimental environment with eye-tracking, it will allow us to assess action goals that are actively pursued.

Thus, we developed the Dodge Asteroids environment (Heinrich et al., 2023), a spaceship themed computer game which is inspired by the simulation environment of Kahl et al. (2022). Participants steer a spaceship and have to avoid crashing into oncoming obstacles that are randomly placed throughout the environment. The agent, the spaceship that is controlled by participants, stays at screen center at all times. All the while the participants gaze is being tracked. Within the environment, participants will encounter drift, a consistent motor influence from either side indicated by a red bar. Additionally, we manipulate motor control by introducing input noise, inaccurate control over the agent. The two environmental features of drift and input noise are opposites in how clearly they can be assigned to a definite environmental variable (with drift being predictable in its perturbing effect which can be assigned to the red bar, and input noise being unpredictable and occurring constantly). As eye movements and the task of steering are closely linked, we expect specific changes in oculomotor control and its underlying cognitive processes induced by input noise.

Hypotheses

Scanning the surrounding environment for action possibilities is done by visual exploration. The assumption is that under reduced motor control efficiency, individuals may adopt a more constrained or focused visual exploration, concentrating on action possibilities that are more likely to be realized. More precisely, hypothesis H1 states that higher input noise is associated with fixational eye movements being initiated closer to the agent.

Furthermore, we expect to be able to identify fixational or smooth pursuit eye movements that indicate action goals and feature the following traits: i) They are not foveating the spaceship or an obstacle, but rather empty space because they aim at a potential future location for the agent, and ii) over the period in which the fixation or smooth pursuit is maintained, the distance between the gaze location and the agent gradually decreases. This is because motor control is exercised explicitly with the goal of reducing the distance. Due to the agent being static at the screen center, a smooth pursuit foveating the goal position within the environment will approach the agent, i.e. the screen center. We refer to these as *foveated action goals*, as they combine fixational and smooth pursuit eye movements, and we expect that when facing input noise, they happen in a more confined space. First, we hypothesize that also in foveated action goals higher input noise is associated with smaller distances to the agent (H2a). Second, higher input noise is associated with greater distances to obstacles (H2b). This reflects the selection of action goals that are more likely to be successfully implemented without crashing. Both distances are assessed at the time of initiation of the foveated action goal. Lastly, we expect higher input noise to be associated with generally shorter durations in foveated action goals (H2c). This is due to the fact that while an action goal is being pursued, input noise causes to deviate too far from the path. The pursued action goal no longer appears to be realizable and is therefore abandoned. The fixation or smooth pursuit is then cancelled, and a new one is initiated somewhere else, which indicates the newly selected action goal. Additionally, this effect may also stem from the reduced distance to the agent which leads to the overall smooth pursuit taking less time to reach its goal. We assume this also to happen when drift appears on screen while an action goal is pursued. The sudden appearance alters the instance in a way that the current action goal is no longer effective and is consequently abandoned.

Additionally, we relate every eye movement with the complexity of the situation in which it was executed. Here we assume that an increasing number of obstacles visible on the screen will elicit the same effects as input noise.

Methods

Dodge Asteroids Experimental Environment

We implemented the simulation environment of Kahl et al. (2022) as an experimental environment using Python (Van Rossum & Drake, 2009) and the PyGame package (Shinners, 2011). The now termed Dodge Asteroids environment runs with 60 FPS. An agent automatically traverses downwards through a funnel with walls on both sides. Obstacles are randomly distributed (uniform distribution with bounds equal to width and height of environment). Participants are tasked to make it to the bottom end of the environment without crashing into the walls or obstacles. They can steer the agent horizontally to avoid crashing using the Y and M keys on the keyboard (QWERTZ layout). A single trial consists of one attempt solving the environment regardless of crash or successful completion.

The environment has a width of 720 pixels and a height of 9000 or 18000 pixels. Automatic downwards traversal of the agent is 6 pixels each frame. The agent itself and the obstacles are 36 pixels in width and height. While solving the environment, participants encounter the two experimental manipulations, drift and input noise. Drift can be best described as wind coming from either side. In every frame within a drift section, the agent is pushed 3 pixels to either side. The sections as well as drift directions are indicated by red bars. Every aspect of drift in this experiment (when it applies and its effects) is meant to be highly predictable. Input noise directly affects the motor control of the participants. Contrary to drift, the effects of input noise are unpredictable. Without input noise horizontal movement of the agent will be 6 pixels each frame when participants steer the agent to either direction. In case of input noise however, the horizontal step between frames is sampled from a normal distribution centered above the normal step size of 6 pixels and with standard deviation being either 3 or 6 pixels, depending on the input noise level (*weak* and *strong* respectively).

The experiment was presented on a 28” ASUS PB277Q screen with a resolution of 1920x1080 pixels. The refresh rate was equal to the FPS of the Dodge Asteroids environment, 60Hz. Participants were seated with their heads rested in a chin rest 80cm away from the screen. We tracked their gazes using a ViewPixx TRACKPixx eye-tracker (VPixx Technologies, Saint-Bruno, QC, Canada) that recorded locations of both eyes with a sampling rate of 2,000Hz. As static reference point, the position of the agent on screen was fixed, with the agent remaining at position x = 954 and y = 270 pixels at all times (referring to the upper left corner of the sprite). Therefore, movement of the environment, either passive by automatic traversal through free fall and drift or active by steering, resulted in the environment moving around the agent. This way new objects appeared at the bottom of the screen and moved upwards until they again disappeared at the top (for reference, it takes roughly 2.45s for an object that appears on the bottom to disappear again at the top of the screen). At the bottom of the screen, a grey bar was drawn with 270 pixels height and spanning the whole width of the screen to prevent participants from moving their gaze beyond the screen. The table on which the setup is mounted is adjustable in height. This way, the chin rest and therefore the participants' faces always remain at the same height relative to the screen.

Procedure

Before we conducted the experiment, we random generated 6 different levels of the Dodge Asteroids environment. This was done in order to obtain a large number of different instances, which we can compare between participants. We generated 3 of the levels with a length of 9000 pixels and the other 3 with a length of 18000 pixels. Within the levels we placed a number of 12 to 168 obstacles with their x and y positions randomly sampled from a uniform distribution bounded by the level borders. 4 to 18 drift sections were randomly placed within each level, their starting y position also sampled uniformly from the length of the individual level. The length of drift sections was kept constant at 270 pixels. All of the 6 levels were played with all of the different input noise levels (no vs. weak vs. strong) and with drift either turned on or off. Therefore, each level was played in 6 different configurations.

While playing, only a small section of the entire environment is shown at any one time, the observation space (enlarged section on the left of Figure 3). An example level is shown on the right-hand side of Figure 3.

All 36 level configurations were presented in random order. Per level configuration participants were given 3 attempts. In case of a crash, the configuration was again presented at a later point during the experiment. If all 3 attempts resulted in a crash, the configuration was removed from the participant's experimental sequence. We did that to prevent participants from familiarizing with the specific instances (constellations of obstacles and drift sections) within the level. Before the start of the experiment, the eye-tracker was calibrated for every individual participant using a 9-point grid. After a training level of 36000 pixels length, the actual level configurations were presented. Before each level configuration the eye-tracker was recalibrated using a 5-point grid. This allowed for breaks between each trial if needed in which participants could disengage from the chin rest. On average, participants played for 31.12 minutes, ranging from 26.21 to 36.38 minutes.

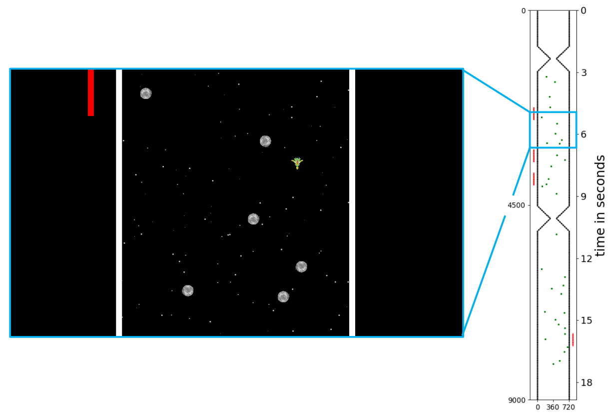

Figure 3: Visualization of an instance within a run in the *Dodge Asteroids* environment (figure taken from Heinrich et al., 2023). For visualization purposes, the agent (green spaceship) is displayed here at the actual position within the environment (not screen center). The complete environment for this run is shown on the right. Obstacles are scattered throughout the instance. The red bar signals drift that is directed to the right.

Variables & Data Analysis

To investigate the effects of input noise in eye-movement control, we used a velocity-based algorithm to detect fixations within the data samples of the eye-tracker. Rows were flagged as fixations when both eyes traveled less than or equal to 1.25 pixels between samples for at least 25 consecutive samples. Overall, we detected 67794 fixations that potentially included smooth pursuit eye movements due to the nature of the visualization. We verified our remaining data set by filtering saccades in a next step. For this, we used an established velocity-based saccade detection algorithm (Engbert & Kliegl, 2003; Engbert & Mergenthaler, 2006). Samples within the data were marked as a saccade when they exhibited a minimum amplitude of 0.5° in eye travel and exceeded a velocity threshold for at least four successive data samples (0.002ms). We computed the velocity threshold with a multiplier $\lambda = 6$. This step led to no actual filtering of the data; therefore all 67794 fixations were included in the analysis.

To determine the exact location of fixations, we calculated the center point of the x and y coordinates of both eyes. Distances to objects on the screen (agents or obstacles) were computed from the fixation location to the center of the object and are given in visual degrees. Second, we assessed fixation durations. To do this, we took the difference in time between the first and last frame in which the fixation was detected. Each fixation is related to the number of obstacles visible on screen at the beginning of the fixation and the input noise level applied in the trial in which the fixation happened. These two variables (N visible obstacles & input noise level) are used as fixed effects predicting the assessed variables. For each individual prediction we explore participant ID and the number of visible drift tiles at the time the fixation was initiated as random effects.

Analyses were done using (generalized) linear mixed models with the help of the MixedModels package (Bates, 2015) within the *Julia* programming language version 1.9.3 (Bezanson et al., 2017). Based on box-cox distributional analyses (Box & Cox, 1964), distance to the agent and the closest obstacle, and duration of fixations and smooth pursuits were transformed to the logarithmic scale. We explored random effects for each model individually, referring to the Bayesian information criterion (Chakrabarti & Ghosh, 2011) for model selection. First, models were fitted using the REML criterion. Then non-singular models were tested against each other, and the final selected model was fitted again using the maximum likelihood criterion. To ensure robustness of effects across samples, the selected model was subjected to a parametric bootstrap with 10000 replications. Hypothesis testing was based on the overlap of the 95% highest density confidence interval of the obtained bootstrap values with the 0-line (no overlap indicates a significant effect). The Random module's MersenneTwister(36) was used for replicability of the results.

We specified contrasts for input noise so that the first comparison is always made between weak input noise and no input noise, and the second comparison is made between strong input noise and weak input noise.

Results

To investigate visual exploration under various degrees of noise in motor control (H1), we build a model predicting the distance to the agent with the fixed effects being the number of obstacles on screen and the level of input noise. The final selected model included random intercept effects for the number of visible drift tiles and participant ID. It further included a random slope effect for the number of visible obstacles on both random intercepts individually. No correlation between random effects was allowed. The confidence intervals of the parametric bootstrap revealed significant effects for both levels of input noise. Compared to no input noise, *weak input noise* significantly *decreased* the distance to the agent [-0.036, -0.008]. However, compared to weak input noise, *strong input noise* led to an *increase* in the distance to the agent [0.001, 0.030]. The *number of visible obstacles* did not affect the predicted variable [-0.028, 0.026].

As we analyzed the fixations in more detail, we noticed that many of the fixations seemed to have a very specific purpose. Participants moved their gaze further down in areas that lay ahead of the agent, only keeping the agent in peripheral vision. They then initiated fixations that lasted on average 0.39 s and remained at the relative position within the game environment foveating empty space and moving along with the environment. These fixations always approached the agent over time, effectively falling into the categorization of smooth pursuits. Therefore, exerted motor control via key presses served to bring the gaze location closer to the agent. Before the smooth pursuit could finally reach the agent, it was canceled, and the gaze was again allocated to the lower half of the screen. We defined a set of filter conditions from the above eye-movement behavior (smooth pursuit initiated on lower half of the screen in empty space, at least 5° away from the agent, and the distance between smooth pursuit location and agent being less when the smooth pursuit is terminated than at its initiation) and filtered the data set accordingly. The resulting data set, called foveated action goals, comprised a total of 42315 smooth pursuits. The following analyses are based only on these foveated action goals.

We again build a model predicting the distance to the agent (H2a). We entered the number of visible obstacles and input noise level as fixed effects. The final selected model included the number of visible drift tiles and participant ID as random intercept effects. No random slope effects were included. *Number of visible obstacles* significantly *decreased* the distance to the agent [-0.012, -0.01], as did *weak input noise* compared to no input noise [-0.034, -0.016]. Compared to weak input noise, *strong input noise increased* the distance to the agent [0.029, 0.047].

Further investigating how action goals are selected from a more confined space in the visual environment, we built a model that predicts the distance to the closest obstacle (H2b). Here we entered input noise level as single fixed effect. We needed to control for variance caused by the number of visible obstacles because more obstacles on screen will automatically lead to a smaller distance between foveated action goal location and the closest obstacle. We therefore explored a random intercept effect for the number of visible obstacles in addition to the number of visible drift tiles and participant ID. All of these were included in the final selected model. No random slope effects were included. The obtained highest density confidence intervals revealed no significant effect for the comparison between *weak input noise* and no input noise [-0.008, 0.011], nor for the comparison between *strong input noise* and weak input noise [-0.012, 0.007].

Finally, we tested what leads to foveated action goals being pursued for shorter periods of time (H2c). We entered the number of visible obstacles, input noise level, and additionally drift tile onset during the foveated action goal as fixed effects into the model predicting foveated action goal duration. The final selected model included a random intercept effect for participant ID and a non-correlating random slope effect for the number of visible obstacles on participant ID. Here the β-values obtained by the parametric bootstrap revealed that the *number of visible obstacles increases* foveated action goal duration [0.005, 0.021]. No effect for the comparison between weak and no input noise was found [-0.006, 0.024], however *strong input noise* led to shorter foveated action goal durations compared to weak input noise [-0.035, -0.004]. Lastly, when a drift tile appeared on screen while the smooth pursuit was executed, it resulted in longer foveated action goal durations [0.426, 0.515].

All the above results relating to changes in motor control efficiency can also be found in Table 1. It contains the name of the contrast and the corresponding boundaries of the 95% confidence interval of the β-values obtained from the parametric bootstrap. Significant effects are indicated.

Table 1: Main results for input noise.

Hypothesis	Contrast	95% CI bounds
H1 (distance to agent)	weak vs. no input noise	[-0.036, -0.008] *
	strong vs. weak input noise	[0.001, 0.03] *
H2a (distance to agent)	weak vs. no input noise	[-0.034, -0.016] *
	strong vs. weak input noise	[0.029, 0.047] *
H2b (distance to obstacle)	weak vs. no input noise	[-0.008, 0.011]
	strong vs. weak input noise	[-0.012, 0.007]
H2c (fixation duration)	weak vs. no input noise	[-0.006, 0.024]
	strong vs. weak input noise	[-0.035, -0.004] *
	drift tile onset (yes vs. no)	[0.426, 0.515] *

Discussion

In this article we have introduced a new experimental environment. It allows us to investigate the interaction of different levels of control in a dynamic context. Furthermore, it can be combined with eye-tracking to implicitly measure key cognitive processes that are involved and how exactly top-down regulatory control looks like. The results of our study shed light on the changes in goal-directed gaze allocation within dynamic tasks, particularly in the context of situated action control. Drawing upon theoretical foundations from Grafton and Hamilton (2007), Kahl et al. (2022), and Synofzik et al. (2008), our investigation explores the interplay between noisy motor control and statistics of fixational and smooth pursuit eye movements.

Our first hypothesis (H1) proposed that higher input noise is associated with fixations initiated closer to the agent. The findings align with this hypothesis, revealing that increased input noise correlates with shorter distances to the agent. This result suggests that under conditions of reduced motor control efficiency, participants tend to concentrate their visual exploration closer to the agent. This focused gaze allocation might reflect a strategy to monitor small but unpredictable motor perturbations. Though, this only held true for weak input noise. Strong input noise, which is associated with large motor perturbations, triggered even longer distances. In foveated action goals, we observed the same nuanced responses to weak and strong input noise. Here again weak input noise elicited an effect that agrees with our hypothesis (H2a). In contrast, strong input noise led to foveated action goals being initiated further away. This surprising result raises questions about top-down regulatory control strategies that are temporarily being adopted here.

We found no effect of input noise on the distance to obstacles (H2b). Participants did not initiate foveated action goals further away from obstacles when they had less accurate control over the spaceship. They did not allocate their gaze in a way that would reflect selecting action goals that might be safer to pursue and it needs to be investigated whether this effect is inherent to our experimental environment or whether it can be generalized. Nevertheless, we could show that input noise leads to shorter duration in foveated action goals (H2c). This may be because input noise lead to straying too far of the path of realizing the action goal and thus abandoning it. Contrary to our hypotheses, the sudden appearance of a drift tile increases foveated action goal duration. This indicates that changes to the continuously evolving instance could easily be factored into motor control giving no need to abandon the action goal. Drift simply led to participants taking longer to reach the intended location, resulting in longer durations. But how much of this can be attributed to the fact that the perturbing effect of drift is predictable? With respect to top-down regulatory control processes, it might be worth to analyze general strategies of how drift is countered by steering the spaceship in specific ways in more detail (horizontal positioning in preparation to drift onset). This is an aspect that we would like to investigate in future experiments in which we introduce different types of uncertainty and manipulate the predictability of environmental features.

The complexity of the visual environment emerged as a crucial factor influencing visual exploration in general but also foveated action goals. The number of obstacles significantly impacted location and duration of fixations and smooth pursuits, emphasizing the role of the configuration of the visual scene in shaping participants' gaze behavior. This is not surprising but can lead to interesting interactions between motor perturbations of different uncertainty and the environmental complexity.

We expanded on the concepts of the Kahl et al. computational model by exploring the implications of sensorimotor grounding of the cognitive control functions featured in the model. More specifically, we investigated how elicited fixational and smooth pursuit eye movements reflect the composition of the action field and the action goal that is actively pursued. Finally, we showed the effects of changes in motor control on eye movement behavior, which allows inferences about the effects on top-down regulatory control.

Our findings could be incorporated into a computational model, that is based on the theoretical assumptions of Kahl et al. (2022) and features the two distinct thresholds of the sensitivity range for sensorimotor error signals and the awareness boundary. Model simulations could predict when exactly top-down regulatory control is initiated within a trial. Model output of eye-movement control might be fitted to human data. Ultimately, this enables a time-dependent investigation of the individual components of situated action control.

Acknowledgements

This research was funded by the German Research Foundation (DFG) Priority Program 2134 "The Active Self".

References

Bates, D. (2015). *MixedModels: A Julia Package for Fitting (Statistical Mixed-Effects Model* (Julia package version 0.3-22) [Software]. https://github.com/dmbates/MixedModels.jl

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, *59*(1), 65–98. https://doi.org/10.1137/141000671

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *26*(2), 211–243.

Chakrabarti, A., & Ghosh, J. K. (2011). AIC, BIC and Recent Advances in Model Selection. In P. S. Bandyopadhyay & M. R. Forster (Hrsg.), *Philosophy of Statistics* (Bd. 7, S. 583–605). North-Holland. https://doi.org/10.1016/B978-0-444-51862-0.50018-6

Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research*, *43*(9), 1035–1045. https://doi.org/10.1016/S0042-6989(03)00084-1

Engbert, R., & Mergenthaler, K. (2006). Microsaccades are triggered by low retinal image slip. *Proceedings of the National Academy of Sciences*, *103*(18), 7192–7197. https://doi.org/10.1073/pnas.0509557103

Gibson, J. J. (1966). *The senses considered as perceptual systems.*

Gibson, J. J. (1977). The theory of affordances. *Hilldale, USA*, *1*(2), 67–82.

Grafton, S. T., & Hamilton, A. F. de C. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human movement science*, *26*(4), 590–616.

Hacker, P. M. S. (1986). *Insight and illusion: Bd. Themes in the philosophy of Wittgenstein (2nd wydanie)*. St Augustine's Press; New ed of 2 Revised ed Edition.

Heinrich, N. W., Russwinkel, N., Österdiekhoff, A., & Kopp, S. (2023, Juli). A Straightforward Implementation of Sensorimotor Abstraction in a Two-Layer Architecture for Dynamic Decision-Making. *MathPsych/ICCM/EMPG 2023*.

Kahl, S., Wiese, S., Russwinkel, N., & Kopp, S. (2022). Towards autonomous artificial agents with an active self: Modeling sense of control in situated action. *Cognitive Systems Research*, *72*, 50–62. https://doi.org/10.1016/j.cogsys.2021.11.005

Land, M. F. (2004). The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. *Experimental Brain Research*, *159*(2), 151–160. https://doi.org/10.1007/s00221-004-1951-9

Land, M. F. (2006). Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research*, *25*(3), 296–324. https://doi.org/10.1016/j.preteyeres.2006.01.002

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision research*, *41*(25–26), 3559–3565.

Marple-Horvat, D. E., Chattington, M., Anglesea, M., Ashford, D. G., Wilson, M., & Keil, D. (2005). Prevention of coordinated eye movements and steering impairs driving performance. *Experimental Brain Research*, *163*(4), 411–420. https://doi.org/10.1007/s00221-004-2192-7

Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, *107*(1), 179–217. https://doi.org/10.1016/j.cognition.2007.09.003

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2016). Task and context determine where you look. *Journal of Vision*, *7*(14), 16. https://doi.org/10.1167/7.14.16

Shinners, P. (2011). *PyGame* (2.1.2) [Python]. http://pygame.org

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition*, *17*(1), 219–239. https://doi.org/10.1016/j.concog.2007.03.010

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

Wilmut, K., Wann, J. P., & Brown, J. H. (2006). How active gaze informs the hand in sequential pointing movements. *Experimental Brain Research*, *175*(4), 654–666. https://doi.org/10.1007/s00221-006-0580-x

Wilson, M., Stephenson, S., Chattington, M., & Marple-Horvat, D. E. (2007). Eye movements coordinated with steering benefit performance even when vision is denied. *Experimental Brain Research*, *176*(3), 397–412. https://doi.org/10.1007/s00221-006-0623-3