The role of reward in accelerating optimisation of information coding for cognitive control and decision-making

Marta Radzikowska

Date of submission: 16.09.2022

M.Sc. Brain and Cognitive Sciences, University of Amsterdam

Research Project Report

Supervisor: Dr. Sam Hall-McMaster**,** Max Planck Institute for Human Development

Second Examiner: Dr. Cyriel Pennartz, University of Amsterdam

Abstract

Encoding information in a way that reflects important features of the environment is crucial for supporting active behaviour. However, it is not clear what role rewards play in re-shaping representations into optimal formats. Here we draw on work demonstrating that the prospect of reward improves cognitive performance through transient enhancement in the encoding of task information. In this thesis, we extend this literature by investigating whether a continuous association between reward and task information changes information coding over longer timescales, to support decision-making. Using a novel task that required the re-formatting of task information for efficient behaviour, we tested whether rewards accelerate changes in internal representations used for decision-making, which would result in faster performance improvements over time. Across two behavioural experiments, we did not find converging evidence for the role of reward in guiding representational change into formats that optimise information processing for efficient goal-directed behaviour. Overall, this work provides a useful starting point for future investigations that explore the role of reward in shifting how information is represented over sustained timescales and how these changes affect cognitive performance.

Keywords: reward, cognitive control, decision-making, learning

Contents

1. Introduction

To navigate and understand the world, humans need cognitive maps of dependencies in their environments. How those cognitive maps are formed and represented in the brain over time remains an active area of interest (Behrens et al., 2018; Whittington et al., 2022). One notable theory suggests that cognitive maps for flexible behaviour are constructed from abstract representations that reflect the structure of the world (Behrens et al., 2018). According to this account, learning entails representational changes in how information is formatted (i.e. reducing dimensionality through constructing neural representations that reflect common structural elements). A strong implication of this theory is that shifting how information is represented into certain formats can optimise decision-making and allow knowledge to be generalised across different tasks with similar structure (Behrens et al., 2018).

The idea that information stored in neural circuits can be reformatted with time and experience is also captured in research examining representational drift (Rule et al., 2019,; Schoonover et al., 2021), which proposes that information coding is not a static process. Rather, it is subject to ongoing reorganisation. The reorganisation of neural activity identified in this research is associated with particular environments, stimuli and actions, and can sometimes occur without any apparent behavioural makers. Recent theoretical accounts suggest that drift might be an adaptive mechanism that reflects continuous learning, updating existing knowledge, and supports generalisation and flexible cognition by integrating experiences across time (Rule et al., 2019; Whittington et al., 2022).

These lines of research highlight the notion that neural representations change across time to optimise future behaviour. What constitutes optimal behaviour is usually defined on the basis of what is important or rewarded in the environment, such as actions and processes necessary to achieve task-defined goals (Karayanni & Nelken, 2022). If changes in neural coding serve a core

function in optimising cognition and behaviour in the pursuit of rewards, it is critical to understand how reward signals guide this optimisation process and change neural representations in a systematic manner. Although past research substantiates the theoretical importance of neural representations being reconfigured to support active behaviour (Behrens et al., 2018; Momennejad, 2020; Rule et al., 2019), the empirical role of rewards in this process has received more limited attention. As such, one central question remaining unclear is whether rewards accelerate representational changes, shifting information coding into formats most useful for decision-making, or whether such changes occur at the same rate, independent of motivational value. To gain more insight into this question, let us now turn to two prominent domains in which the effects of reward on behaviour and brain activity have been investigated: cognitive control and learning.

Rewards and Cognitive Control

Rewards in cognitive sciences, and in the world around us, are used to incentivise effort and improve performance. Vast evidence for the motivational effects of rewards comes from research on cognitive control (Botvinick & Braver, 2015). Broadly, cognitive control refers to a set of processes that enable goal-directed cognition and behaviour, such as forming goals and determining the actions necessary to achieve them (D'Mello et al., 2020). These processes can be modulated by rewards, as evidenced by an increasing number of studies showing that performance-contingent rewards enhance cognitive control operations, such as attentional control, task switching, working memory, information integration, and reduce the effects of interference (Ashby & O'Brien, 2007; Dixon & Christoff, 2012; Shen & Chun, 2011; Klink et al., 2017; Yamaguchi & Nishimura, 2019). A useful framework, that aims to account for the modulatory effects of rewards on cognitive control, is the Dual Mechanism of Control Theory (DMC) (Braver, 2012). DMC proposes that cognitive control has two main dimensions: proactive and reactive. Proactive control is engaged in preparation for upcoming cognitive demands, serving to strengthen the representation of

information about current goals during preparation, and bias subsequent sensorimotor processing to support goal-relevant action. Reactive control is engaged during conflict detection and resolution, serving to respond to unexpected changes in the environment. Consistent with DMC, empirical studies have indicated that the prospect of a high reward for performing well promotes preparation, preferentially enhancing the proactive dimension of cognitive control (Etzel et al., 2016; Frober & Dreisbach, 2016).

Whether the prospect of reward can be linked to changes in information coding has been extensively researched in the context of rule-based tasks (Etzel et al., 2016; Hall-McMaster et al., 2019; Padmala & Pessoa, 2011). On the basis of these studies, a central neural mechanism for the beneficial effect of rewards on cognitive control is that rewards increase the neural encoding of task-relevant information, within frontoparietal brain regions, particularly in lateral prefrontal cortex, which is also involved in attentional processes (Etzel et al., 2016). This strengthening of task-relevant information is especially crucial in the face of interference, where rewards lead to a reduction in conflict signals within medial prefrontal cortex (Padmala & Pessoa, 2011) and further alter information coding to reduce the similarity between neural coding patterns that could come into conflict (Hall-McMaster et al., 2019). When considered together, these studies demonstrate transient, cue-related enhancement in neural coding of task-relevant information at relatively short time-scales (i.e. trial-to-trial changes). However, it is not clear how rewards influence changes in neural representations of task-relevant information on longer timescales, when particular task information is continually associated with reward.

Rewards, Learning and Replay

In addition to its effect on cognitive control, reward also has an important role in learning. Notebaert & Braem (2016) posit that reward receipt is linked to reinforcement and associative learning processes that strengthen goal-relevant connections. These learning-driven changes can go

on to influence cognitive control performance. Behavioural evidence for this idea was demonstrated by Krebs et al., (2010), who observed that when reward on a Stroop task was associated with particular colours, the interference effect of distracting stimuli on those rewarded colours was reduced. Further, the authors also found greater interference of words that were semantically related to the rewarded colours in experimental blocks where the rules changed and attending to the rewarded-colour was no longer adaptive. These findings suggest that a consistent relationship between a specific stimulus and reward might trigger learning processes that result in enduring changes to internal task representations, in turn strengthening proactive control and improving task performance in contexts where enhanced activation of reward-related information is beneficial. What neural processes are driving the tuning of task representations into formats that emphasise reward-related information is an active area of research.

One candidate mechanism for tuning and adapting internal task representations into an efficient format is hippocampal replay (Momennnejad, 2020; Wittkuhn et al. 2021). Replay refers to sequential reactivation of learned information during periods of rest and sleep, as well as periods of active behaviour (Foster, 2017). Forward replay is proposed to support planning for goal-directed behaviour, while reverse replay is associated with learning acceleration and knowledge consolidation. Given replay's crucial role in learning, there has been a considerable interest in how it might interact with rewards. Current evidence suggests that rewards do not influence forward and backward replay uniformly. Findings from rodent studies show that reward increases the strength of the forward replay and frequency of backward replay during rest (Ambrose et al., 2016). This effect was observed for the current paths to reward as well as for other observed reward-related trajectories in the environment (Ambrose et al., 2016; Bahattari et al. 2020). In humans, one study explicitly studied the effect of reward motivation during learning on offline hippocampal dynamics during rest. They found that content associated with high-reward contexts is preferentially replayed during post learning rest. The extent of preferential replay is positively related to memory retention.

The authors suggest that their findings point towards modulatory role of replay in memory consolidation process (Gruber et al., 2016). In addition to these findings, recent theoretical work posits that replay could be a core mechanism through which information is restructured into formats that are optimised for decision-making (Wittkuhn et al., 2021). If replay supports the reconfiguration of information coding and increases as a function of reward, we would expect rewarded information to undergo more rapid representational changes.

The Current Study

In this project, we used behavioural measures to investigate whether rewards accelerate changes in information coding over time. Previous theoretical and empirical work within DMC (Braver, 2012; Yamaguchi & Nishimura, 2019), and studies showing that rewards dynamically modulate coding of task-relevant information (Etzel et al., 2016; Hall-McMaster et al., 2019) have established that reward has a transient impact on information coding, from trial-to-trial. Here we were interested in extending this work by examining whether continuous associations with rewards also led to representational changes over longer timescales (from block-to-block).

To test this, we developed a novel decision-making task to examine a) if changes in performance over the course of the experiment were associated with changes in how task information was encoded and b) whether the rate of representational change was influenced by the presence of reward. We examined these questions across two in-lab experiments (N = 18 and N = 23). As the experiments in this thesis were part of a larger neuroimaging project, our observations also aimed to assess the feasibility of this task for future behavioural and neuroimaging studies.

Drawing on the role of reward in learning (Notebaert & Braem, 2016) and a recently proposed role for replay in driving representational change (Momennejad, 2020; Wittkuhn et al., 2021), we predicted that reward would accelerate a shift from representing multiple features of the environment for decision-making into a more efficient, task-optimised format, wherein the sole

feature needed for decision-making was represented. Drawing on DMC (Braver, 2012), we predicted that these changes would in turn lead to more effective proactive control, resulting in improved performance under interference.

2. Materials & Methods

2.1. Study 1

Participants

A sample of 19 participants was recruited via an existing participant database ("Castellum") and completed the task in person. The experiment was part of the larger project and the sample size was selected to assess the feasibility of the task design for future neuroimaging work. During the experiment, one dataset was excluded due to a data saving error. The final sample consisted of 18 participants (11 female) between the age 21 and 35, with a mean age of 26 (*SD* = 3.61). Participants were fluent in English, reporting no history of neurological and psychiatric illnesses, normal or corrected-to-normal vision, and no colour-blindness. Participants received €10 for participation and could earn up to €15 in additional bonus based on their performance. The study was approved by the Max Planck Institute for Human Development ethics committee. All participants signed a document confirming their informed consent and compliance with the hygiene rules related to covid-19.

Materials

The task was run using Psychophysics Toolbox-3 on Matlab. Two computers were used for running the experiment, one with a screen resolution of 1800 x 1090 using Matlab version R2021b and one with a screen resolution of 1800 x 1200 using Matlab version R20219b. Animal videos and still images from the videos were used as cues in the task. Target stimuli were constructed from shape and pattern images obtained from copyright free online databases (see Appendix A for details). We

chose patterns that were relatively dissimilar to the shapes in their geometrical features. Feedback was displayed in white font while information about reward was displayed in green font. D, F, J, K, and L keys on a desktop QWERTZ keyboard were used to record participant responses. Analyses were performed in Matlab version R2021b and in IBM SPSS version 28.0.0.

Experimental Design

Main Task. Participants completed a virtual planning task. In the task, participants were asked to imagine they were a zookeeper distributing pellets to five animals in the zoo. Each animal was associated with a food pellet stimulus that had two dimensions (a shape and a pattern). Participants main task during the experiment was to plan trips around the zoo by ordering the sequence of pellets according to animals' preferences and zoo's structure (Fig. 1B).

Specifically, each trial had three main phases, a planning phase, a response phase and a feedback phase (Fig. 1A). During the planning phase, participants were shown their starting location in the zoo and had 9s to plan the sequence of four food items needed after leaving that location. The starting location was indicated with a short animal video lasting 0.75s, which froze on the last frame for the remainder of the planning phase. During the response phase, participants were shown partial information about the food pellets (shapes or patterns). These target stimuli were presented in a random order on screen to prevent participants from preparing a motor sequence prior to the response phase. Participants needed to select a sequence of 4 items, selecting the screen position of each item in turn. The response phase terminated as soon as 4 items had been selected, up to a maximum of 6.5s. During the feedback phase, participants were given feedback about their selected sequence. If the sequence was correct given the zoo structure, participants saw their selected sequence on screen with 'correct!' printed above it. Note that there were two correct paths from each starting animal, with the exception of the animal located in the middle of the zoo for which there were 4 correct paths. On rewarded trials (described in the next section), 'correct!' was

replaced with the amount of reward earned (Fig. 1C). If the sequence was incorrect given the zoo structure, participants saw a randomly selected example of a correct sequence, with 'incorrect! One possible sequence was:' printed above it. The feedback phase lasted 2s. The feedback phase was followed by an inter-trial interval (ITI), a blank screen with a mean duration of 2.5s. The specific ITI duration of a particular trial was randomly selected from a truncated exponential distribution with a mean of 2.5s, a lower bound of 1s and an upper bound of 12s.

Each block contained 20 trials. Across the 20 trials, the five starting locations in the zoo were presented four times. The order was randomised. At the end of the block, participants were informed about their accuracy and average reaction time (RT). Participants could also take a self-paced break. Participants completed 8 blocks in total. In four blocks, participants had to report paths based on the food pellets shapes. In four blocks, participants had to report paths based on the food pellets patterns. Shape and pattern blocks were interleaved. Whether the main task began with a shape or a pattern block was randomised.

The 5 animals in the zoo for each participant were selected at random from a set of 10 possible animals prior to the experiment. How the animals were mapped onto the graph structure, as well as the specific combination of shape and pattern that generated the pellet associated with each animal was also randomised and pre-computed prior to the experiment.

Reward Manipulation. During some blocks, participants could earn points for reporting correct sequences. For some participants, it was possible to earn points during shape planning blocks. For other participants, it was possible to earn points during pattern planning blocks. The rewarded block type was counterbalanced across participants. At the beginning of the experiment, participants were informed that they would be able to earn extra points on some trials, depending on the path taken. The best sequences were rewarded with €75, other sequences were rewarded with €25 and some sequences were not rewarded at all. The reward earned for reporting a sequence

depended on the end location of that sequence. High reward locations were placed on the one side of the zoo (e.g. the right nodes of the graph as shown in Fig. 1B), while low reward locations were placed on the other side (e.g. the left nodes). The rewarded side of the zoo was changed halfway through the experiment. Sequences terminating at the middle node in the zoo were not rewarded. Following Jimura, Locke & Braver (2010), this placement of rewards was designed so that rewarded blocks contained a mixture of rewarded and unrewarded trials. Specifically, we designed the placement of rewards so it was possible to reach the high reward locations in 4 steps when starting from the bottom two nodes and from the middle node. In contrast, it was not possible to reach the high reward location in 4 steps when starting from the top two nodes in the zoo graph. This meant that, within a rewarded block, 60% of trials ended in a rewarded location (high or low) and 40% of trials ended in a non-rewarded location. Participants were informed about possible sequence values (€75, €25, €0) and were instructed to try to find out which sequences were linked to the highest reward by trial and error. At the end of each rewarded block, participants were informed about the number of points they had earned and were shown a scale displaying how much of the maximum bonus they have earned so far. To achieve the maximum payout of €15, participants had to reach 75% of the maximum number of points available in the task.

Probe Trials. On 25% of the trials, target images were paired with novel combinations of the irrelevant feature during the response phase. For example, if shape was the relevant feature, an interference probe would pair shapes with random patterns that did not match any of the shape and pattern combinations determined in the graph structure (Fig. 1D). As shapes and patterns in these novel pairings related to different sequences around the zoo, this manipulation aimed to interfere with sequences participants thought of during the planning phase. This interference effect was expected be especially strong if participants are encoding both features during planning (shapes and patterns) and weaker if participants are encoding a single relevant feature (e.g. shapes alone).

Therefore, this manipulation aimed to inquire into the extent to which participants plan their sequence based on one or both features. Drawing on DMC (Braver, 2012), we predicted that greater prospective encoding of reward-related information would buffer against interference, resulting in diminishing effects of incongruent distractors in the response phase.

Binary Choice Trials. To gain more insight into participants' knowledge about the reward structure, we included a binary choice phase at the end of each task block. Within this phase, participants were shown all possible pairs of animals in the zoo and were asked select the animal they would like to visit next. The main idea behind these trials was that, if participants had gained knowledge about the high, low and no reward locations in the zoo, they would select the animal with the higher associated reward during binary choice. There were 10 binary choice trials per phase. In each trial, a pair of pseudorandomly selected animals appeared on the screen. One animal was shown on the left side of the screen, the other on the right. Whether a stimulus within each pair was presented on the left or right of the screen was determined at random. Participants indicated their selected animal stimulus using the F and J keys. Participants had 6.5s to enter their response and were instructed that the phase would not contribute to their final score on the task.

Training Procedure. Prior to the experiment, participants completed a virtual 'zookeeper' training. In the training stage, participants learned about the associations between animals and food pellets, as well as zoo's graph structure. The training consisted of several phases that increased in difficulty, building up to include all elements of the task.

In the first stage, participants learned about associations between animals and pellets. On each trial, participants were shown an animal video (0.75s), then a blank delay (0.5s). Following this, the 5 possible food pellets appeared on screen in a random order and participants needed to select the screen position of the associated food pellet. Feedback about the response accuracy was

presented for 2s. If the response was incorrect, or not entered with the allotted time, participants were shown the correct animal-food pellet pair during this time. The trial ended with a blank ITI (1s). Each training block lasted for 21 trials. This number allowed us to balance all pairwise trial transitions between the five animals. To begin with, participants had a 5s time limit to respond. Once participants scored 90% or higher, the response limit was reduced to 1.75s. Participants then needed to score 90% or higher one final time to complete the stage.

In the second stage, participants learned about possible transitions from one animal to the next. On each trial, a starting animal video was presented on the screen (0.75s). The final video frame then remained on screen and two different animal images were presented below it. Participants had to select which animal would be encountered next, after leaving the central animal enclosure (max. 4s). Following the response, participants were shown feedback about whether they had identified the correct transition (2s). If the response was incorrect, participants were shown what the correct transition would have been. The trial concluded with a 1s ITI. Each block consisted of 6 trials. This number was selected so that each of the five animals was shown as the central animals once within a block. The middle animal was displayed twice so that participants could learn about its two possible transitions (see Fig. 1B). After two blocks with less than 100% correct, participants were reminded that there was more than one possible pathway from some animals. After three blocks with less than 100% correct, participants were shown an image of the zoo's graph structure (displaying animals, without pellets) (Fig 1.B) to speed up the training. To complete this stage, participants needed to complete two successful blocks (100% correct) in a row.

In the third stage, participants learned how to perform a planning task. The trial structure was similar to the main task (Fig. 1A), containing a planning phase, a response phase, a feedback phase and an ITI. At the beginning, participants were shown one of the animals during the planning phase and were asked to use their knowledge about the zoo to imagine the next animal that would be encountered and the food pellet it would need. Participants then selected this item during the

response phase (max 4s) and received feedback about whether their response was correct (2s). If the response was incorrect, or not entered within the time limit, participants were shown the correct response during this period. The trial ended with a 1s ITI. To complete a phase, participants needed to give at least one correct response for each starting animal within a block of 12 trials. As the training progressed the task gradually increased in difficulty. In each phase, the number of pellets participants had to report increased by one. That is, in second phase participants reported two consecutive pellets within 6 seconds, and in the third phase three consecutive pellets within 8 seconds. In the fourth phase, participants completed a full sequence plan, reporting four consecutive pellets within 8 seconds. In the fifth phase, blocks were separated into different types (shape blocks and pattern blocks). In shape blocks, participants were shown shapes (but not patterns) during the response phase. In pattern blocks, participants were shown patterns (but not shapes) during the response phase. As a final training stage, the time to enter the response was shortened to 6.5s, to further test participants' understanding of the task and to encourage planning prior to response period. This phase concluded once the participant had passed the block criterion above (one correct response for each starting animal entered within 6.5s) for both features.

At the end of the training, participants were familiarised with the distractor probe trials in preparation for the main task. Participants competed two blocks (one shape and one pattern block) of 5 probe trials. Note, the goal of this procedure was to familiarise participants with such trials, not to train them to ignore distractors. As such, there was no performance criterion for these blocks to move to the main task.

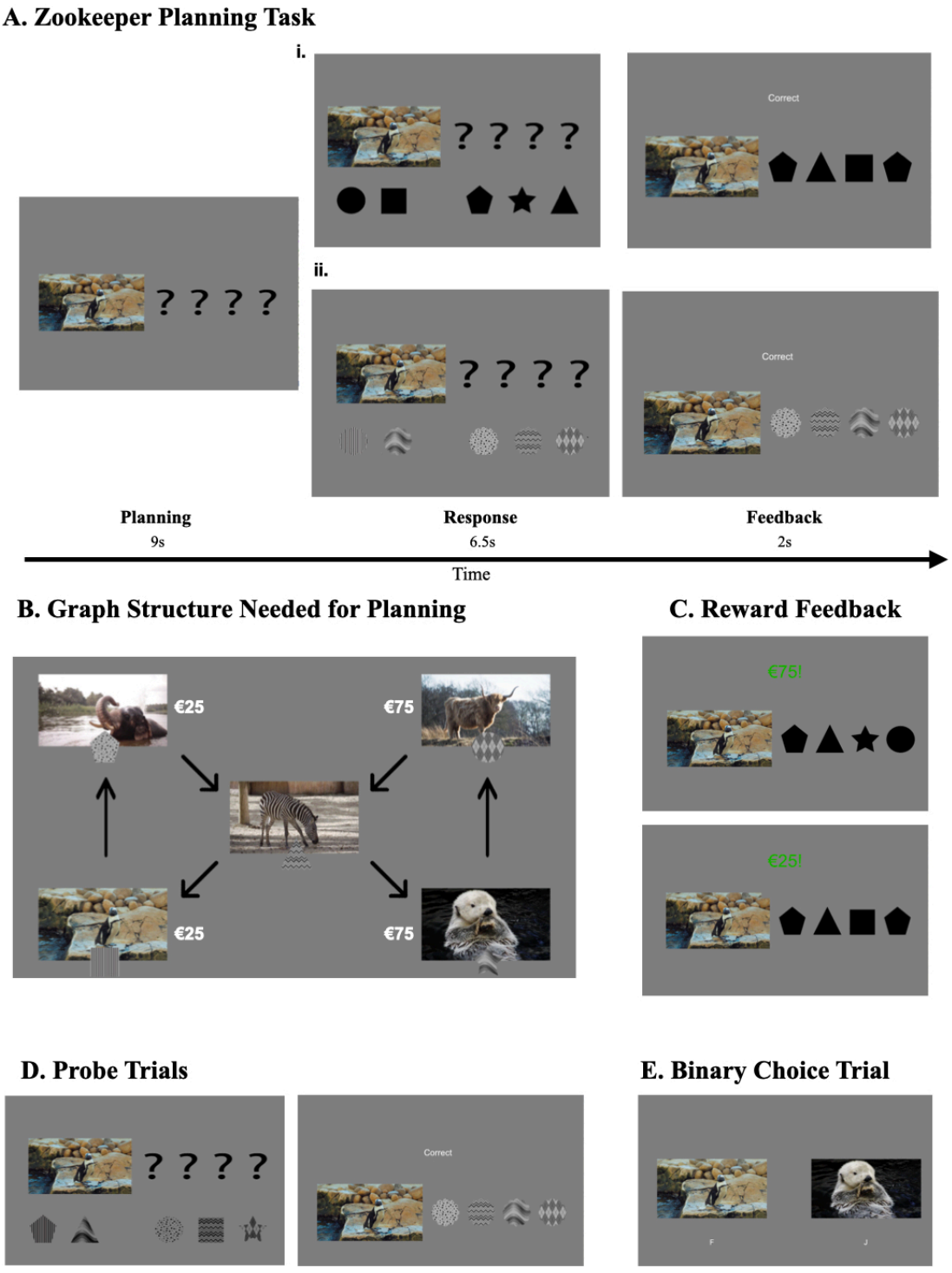

Figure 1. Experimental Design. Participants performed a virtual zookeeper task, which required them to plan trips around a virtual zoo to feed animals in specific sequences. Each animal had to be fed a unique food pellet stimulus (a textured shape), leading to unique plans from each starting position. **A.** *Trial Order.* In the planning phase, an animal indicating starting point of the sequence

appeared on the screen and participants had a total 9s to plan the upcoming sequence. During the response phase, participants were shown target items in random screen positions and had to select items in the correct order. Participants had up to 6.5s to enter their response in study 1 and 5.5s in study 2. i. In the shape blocks, participants were shown food pellet shapes as targets. ii. In the pattern block participants are shown food pellet patterns as targets. Participants were shown correct or incorrect feedback for 2s and the trial concluded with a jittered ITI (mean 2.5s). **B**. *Graph Structure Needed for Planning*. The possible transitions between animals were based on the underlying graph structure of the zoo. Participants learned this structure during training. **C.** *Reward feedback.* Reward was based on the end location (pellet) of the reported sequence. High reward, goal locations were placed on either right or left side of the zoo. If the reported sequence ended in a high reward location, participants received €75 as reward feedback. If the reported sequence ended in a low reward location (i.e. on the opposite side of the graph), participants received €25 as reward feedback. The placement of high and low reward locations was switched midway through the experiment to encourage ongoing engagement with the task. In the second study, participants could earn €50-€100 for ending at a high reward location and €5-€10 for ending at a low reward location, depending on their speed. **D**. *Pattern Probe Trials*. In pattern blocks, probe trials occurred on 25% of trials in study 1 (50% in study 2). In the response phase of a probe trial, patterns were presented with novel combinations of shapes, allowing us to assess whether participants were planning based on one or both target features. Shape blocks contained equivalent probe trials, which combined shapes with novel patterns during the response phase. **E.** *Binary Choice trial*. In binary choice trials, participants were asked to select which of the two animals they would prefer to visit. Participants made one binary choice for each possible pair of animals in study 1 and two binary choices for each pair in study 2, following each block of the planning task.

Statistical Analyses

The dependent measures in this experiment were accuracy (defined as the proportion of correct responses in the block), reaction time (RT) and participants' choices in binary selection trials. For the main behavioural analyses (accuracy and RT), data were analysed with repeated measures ANOVAs and t-tests. The Bonferroni correction was used to correct for multiple comparisons. For Binary Trials analyses we have explored how accurately participants select the animals associated with higher reward value in the previous block. We have analysed changes in preferences for locations associated with higher reward across blocks using ANOVAs and t-tests.

2.2. Study 2

Participants

For this experiment, we aimed for the sample of 25 participants. Two recruited participants were excluded as they could not complete the training in the time available. The final sample consisted of 23 individuals (15 female) between 18 and 35 years old, with a mean age was 26.9 (*SD* = 4.58). Recruitment and criteria for participation were identical to the previous study.

Experimental Design

Main task. The design of the main task was similar to the previous experiment. We made three key changes designed to increase power and boost the sensitivity of our design. First, the number of probe trials was increased from 25% to 50% in a block. This meant each block now had 10 trials where the irrelevant feature could interfere with sequence reporting. Increasing the number of probe trials allowed us to investigate the effect of reward on sequence encoding with more statistical power. Second, the time limit for entering the response after targets appeared on the screen was shortened from 6.5s to 5.5s. This change was designed to reduce possible ceiling effects, in which performance on the task was so high that a potential influence of reward on information

processing could not be detected. The change was also made in light of the short median RTs observed in the first study and has the additional benefit of incentivising planning prior to the response phase, since there was less time to plan during the response phase itself. Third, we made adjustments to the reward structure. In this experiment, reward was not only based on participants' accuracy, but also on their RT. For reporting a correct sequence that ended in a high reward location, participants could earn €50-€100 depending on their speed. For reporting a correct sequence that ended in a low reward location, participants could earn €5-€10 depending on their speed. The motivation for including this RT incentive was to allow reward-induced changes in information coding to be detected in the RT domain, even if participants showed high levels of accuracy in both rewarded and unrewarded blocks (as in study 1). To summarise the main changes, the design for study 2 increased the number of probe trials, reduced the response limit and incentivised RTs, all of which aimed to improve the power and sensitivity of our design.

Reward Manipulation. In the second study, participants could earn a reward based on whether they reached the goal locations as well as based on how quickly they entered their response on rewarded trials. By responding quickly, participants could earn between €10 and €50 when they reached high reward locations, and between €1 and €5 for low reward locations. The reward received was based on the RTs distribution in rewarded blocks. Responses that were shorter than 15%, 30%, 45%, 60% and 75% of previous RTs in the rewarded blocks earned additional €5/1, €20/2, €30/3, €40/4, €50/5 for high and low reward locations respectively. Initially the benchmarks for reward were set at 3.5s, 4s, 4.5s, 5s, 5.5s for each of the five reward levels. These initial benchmarks were based on the distribution of RTs observed during experiment 1. The bonus was added on top of the base reward for the reached location and the final sum was displayed to the participant during feedback phase. Therefore, the maximum amount participants could earn was €100 for the high reward, goal locations and €10 for the low reward locations. The implementation

of individualised RT reward followed the procedure outlined in Hall-McMaster et al. (2019). The bonus was based on participants RTs in the reward condition only, rather than on all responses in order to keep the bonus specific to the rewarded feature, and avoid a situation in which individuals intentionally performed slower in unrewarded blocks to lower the criteria for high RT bonuses in the rewarded blocks.

3. Results

3.1. Study 1: Main Analyses

Accuracy improved over time

First we examined how performance on the task changed across task blocks (Fig. 2). We hypothesised that participants would improve on the task over time, reflecting a shift in information coding from multiple stimulus features (shapes and pattern combinations) to individual stimulus features (shapes or patterns). Encoding individual stimulus features during planning would provide a more efficient representation of the information needed for behaviour within each block, optimising performance on the task.

Average accuracy, defined as the proportion of correct responses, across all eight experimental blocks was 0.89 (SD = 0.310). A response was considered to be correct when all four elements followed a path determined by the zoo graph structure and the response was entered within the time limit of 6.5s from the beginning of the response phase. Partial responses and partial accuracy were excluded from analyses. To assess whether accuracy changed with increasing time on task, we conducted a repeated measures ANOVA on accuracy scores, with a factor of task block (levels 1-8). The results showed a significant main effect of task blocks on task accuracy ($F(7,17)$ = 9.727, $p < 0.001$, $\eta_p^2 = 0.230$), confirming improved performance over time. Results of Bonferroni adjusted post-hoc comparisons indicate that the average accuracy improved significantly from block 1 to 2 (M = 0.79, SD = 0.411 and M = 0.86, SD = 0.349 respectively, p = 0.045) and 1 to 6 (M =

0.95, *SD* = 0.224, $p < 0.001$). Accuracy further improved from block 2 to 6 (*M* = 0.95, *SD* = 0.224, $p = 0.003$). Results of other pairwise comparisons were non-significant (all *ps* > 0.05). These results confirmed the prediction that participants would improve in task over time. Visual inspection of the data showed stable and high overall performance on the task, with average performance peaking in block 6, where accuracy reached 95%.

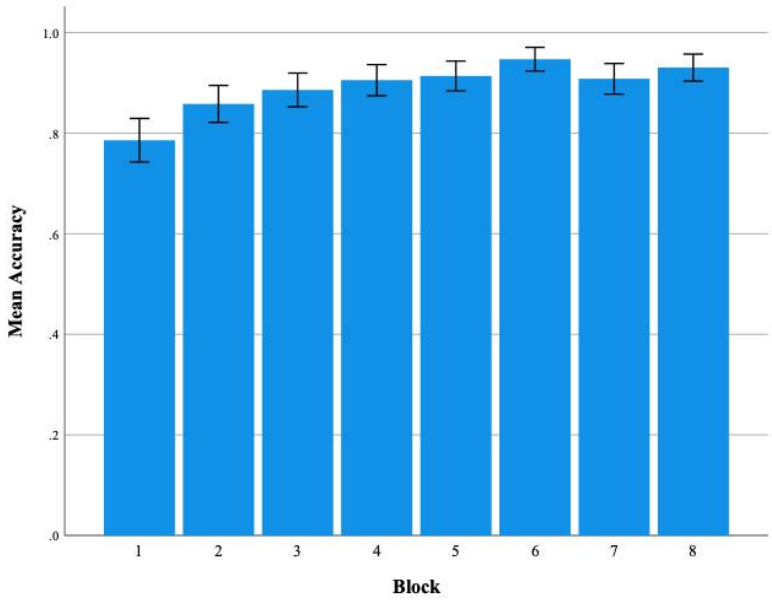

Figure 2. Mean accuracy across experimental blocks. Error bars indicate standard error of the mean (SEM).

Reward did not reliably improve accuracy over time

One of the main inquiries of the present study is whether rewards change how information is represented during planning, accelerating a shift from representing multiple features (sequences of shape-pattern combinations) to individual features (sequences of shapes or patterns). This shift was expected to improve performance because representing individual features within a given block was more efficient for the task at hand, providing all information needed for upcoming behaviour. We therefore expected a greater increase in block-to-block performance for rewarded blocks, compared

to non-rewarded blocks. For the analyses that follow, experimental blocks 1-8 were separated into the 1st, 2nd, 3rd and 4th rewarded block and the 1st, 2nd, 3rd and 4th unrewarded block (see Fig. 3).

To analyse the effect of reward on performance, we ran a 2x4 repeated measures ANOVA. The first factor was presence of reward for each feature (rewarded vs unrewarded). The second factor was the block number (1 through 4). Based on our central hypothesis, we predicted an interaction between reward condition and block number on accuracy, reflecting greater increases in accuracy over time in the reward condition. As the assumption of independence was violated, Greenhouse-Geisser corrected values are reported. The analysis revealed a main effect of reward ($F(1,17) = 4.997$, $p = 0.026$, $\eta_p^2 = 0.014$). Contrary to expectations, participants performed marginally better on non-rewarded blocks ($M = 0.80$, $SD = 0.294$) compared to rewarded blocks ($M = 0.88$, $SD = 0.163$). Overall, the results show a small, negative effect of reward on accuracy. The analysis also showed a significant main effect of block number ($F(3,38) = 15.633$, $p < 0.001$, $\eta_p^2 = 0.479$), confirming a general improvement in performance over time. Consistent with our core prediction, the interaction between reward and block number had a significant impact on accuracy ($F(42) = 3.468$, $p = 0.023$, $\eta_p^2 = 0.183$). Results of Bonferroni adjusted post-hoc comparisons indicate the interaction was driven by a more prominent increase in accuracy from the 2nd to 3rd rewarded block (mean difference (MD) = 0.081, $SD = 0.091$, $t(17) = -3.757$, $p < 0.001$), compared with the 2nd to 3rd non-rewarded block ($MD = 0.011$, $SD = 0.073$, $t(17) = 0.638$, $p = 0.333$; ps > 0.05 for all other comparisons). Paired samples statistics confirmed that the increase in accuracy between block 2 and 3 was significantly greater for the rewarded feature, compared to the unrewarded feature ($t(17) = 3.757$, $p < 0.001$)

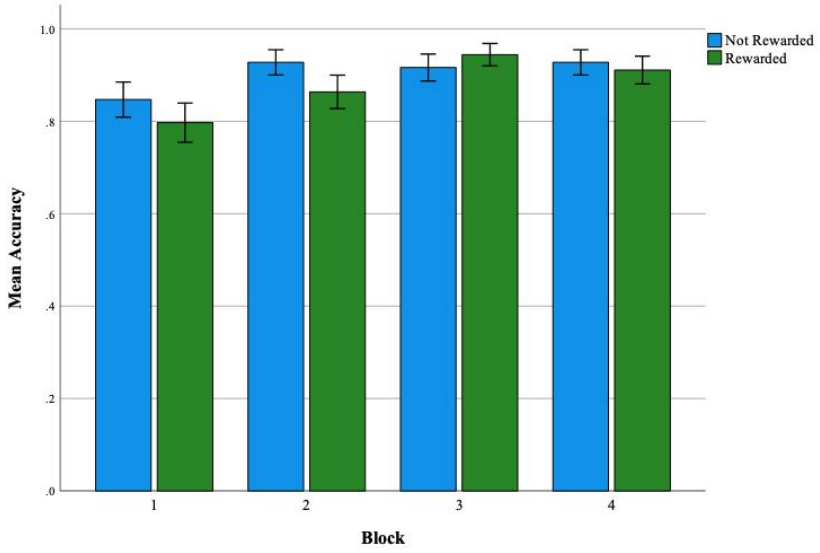

Figure 3. Accuracy for rewarded and unrewarded features for rewarded and unrewarded condition across blocks. Error bars represent SEM. For the analysis, experimental blocks 1-8 were separated into the 1st, 2nd, 3rd and 4th rewarded block and the 1st, 2nd, 3rd and 4th unrewarded block.

Reward modulated improvements in accuracy on probe trials

In this experiment, probe trials were designed to inquire into the extent to which participants planned sequences based on one or both features. We expected that through the experiment, participants would improve their performance on probe trials, reflecting more feature-selective planning.

Following previous theoretical and experimental work within DMC (Braver et al., 2012), we predicted that the negative effect of probe trials on performance would be smaller for rewarded condition. To test this, we ran a 2x4 ANOVA using accuracy data on probe trials, with the same factors as the analysis of reward effects reported in the previous section. We did not find a significant main effect of reward ($F(1,17) = 1985$, $p = 0.177$, $\eta_p^2 = 0.105$). We did find significant effect of block number ($F(3,42) = 9.508$, $p < 0.001$, $\eta_p^2 = 0.359$). There was also a significant interaction effect between reward and block number ($F(3,49) = 4.498$, $p = 0.008$, $\eta_p^2 = 0.209$). Consistent with our predictions, Bonferroni corrected follow up analyses showed larger block-to-

block improvements on probe trials for the rewarded, compared to the unrewarded feature. Specifically, the block-to-block improvement from block 2 to block 3 was significantly larger in rewarded blocks compared to non-rewarded blocks (MD = 0.155, SD = 0.287, $t(17)$ = 2.296, p = 0.017). For rewarded probe trials, there were significant improvements in accuracy from block 1 (M = 0.58, SD = 0.342) to block 2 (MD = 0.178, SD = 0.246, M = 0.76 SD = 0.279, $t(17)$ =-3.063, p = 0.004), from block 1 to block 3 (MD = -0.333, SD = 0.336, $t(17)$ = -5.638, p = 0.001) and from block 2 to block 3 (MD = 0.156, SD = 0.055, M = 0.91, SD = 0.123, $t(17)$ = -4.208, $p < 0.001$). For unrewarded blocks, only the improvement from block 1 (M = 0.76, SD = 0.279) to block 2 (M = 0.88, SD = 0.170) was significant (MD = -0.122, SD = 0.207, $t(17)$ = -2.500, p = 0.11). All remaining pairwise comparisons were not significant (all $ps > 0.05$). To further understand the impact of reward on probe trials, we explored differences in accuracy between rewarded and unrewarded features in the first and last block. In block 1, we detected a significant difference in accuracy between rewarded and unrewarded feature blocks, with the direction of the effect indicating lower accuracy on rewarded probe trials (MD = -0.178, SD = 0.342, $t(17)$ = -2.204, p = 0.021). In block 4, the difference between the rewarded and unrewarded feature blocks was not significant (MD = -0.056, SD = 0.197, $t(17)$ = -1.230, p = 0.236). To summarise the main results in this section, performance on all probe trials improved across the experiment. However, this effect was more prominent for rewarded compared with unrewarded blocks.

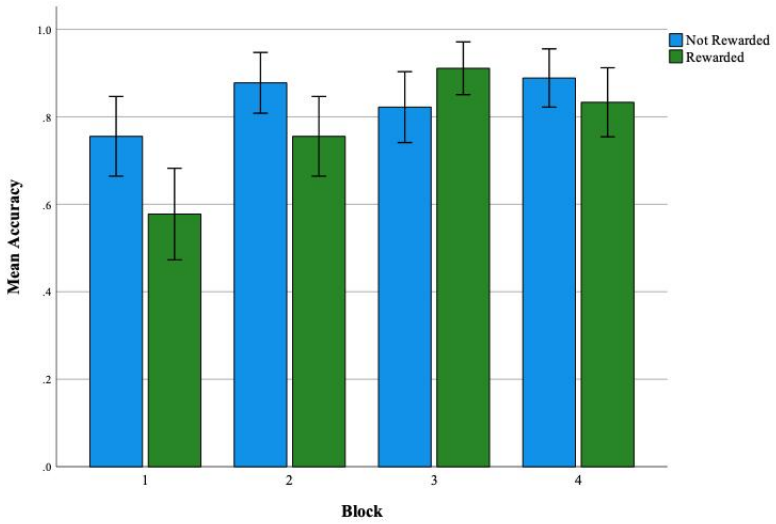

Figure 4. Accuracy on probe trials for rewarded and unrewarded conditions across blocks. Error bars show SEM. For the analysis, experimental blocks 1-8 were separated into the 1st, 2nd, 3rd and 4th rewarded block and the 1st, 2nd, 3rd and 4th unrewarded block.

3.2. Study 1: Exploratory Analyses

RT on probe trials improved over time

To get more insight into processes behind sequence planning, we performed additional exploratory analyses investigating variability in reaction times (RT) across conditions. The aim of this analysis was to inspect sensitivity of our design to RT differences. We expected that variability in RT across blocks and conditions would mirror the changes in accuracy, with RT decreasing across blocks. For this analysis, we computed median RTs for the last registered button press on probe trials, in which participants gave a correct response. We focused on probe trials, as these appeared to be more diagnostic in the accuracy analyses reported above.

The average RT for correct responses across all participants was 3.719s (*SD* = 0.942). For probe trials, it was 4.334s (*SD* = 0.989, 14.2% longer than the RT across all trials). A 2x4 repeated measures ANOVA on probe trial RT, with the same factors as the accuracy analyses presented above, revealed a significant effect of block number on RT ($F(3,37) = 5.509$, $p = 0.002$, $\eta_p^2 = 0.245$), indicating that RT decreased as participants gained more experience with the task (block 1:

M = 4.632s, *SD* = 0.920; block 2: M = 4.383s, *SD* = 0.779; block 3: *M* = 4.284s, *SD* = 0.746; block 4: *M* = 4.082s, *SD* = 0.766). Bonferroni corrected pairwise comparisons showed a significant decrease in RT from block 1 to 4 on rewarded trials (*MD* = 0.551, *SD* = 0.156, *p* = 0.013). All other comparisons were non-significant (p > 0.05). We did not detect a significant effect of reward on RT (*F*(1, 17) = 1.075, *p* = 0.314) or interaction between reward and block number on RT (*F*(1,40) = 0.545, *p* = 0.653). To summarise, exploratory analyses showed RT decreased across blocks. Visual inspection of the data suggests that the drop in RT was more pronounced on rewarded feature blocks (Fig. 5). However, our statistical tests indicate that reward did not have a significant influence on RTs across the task.

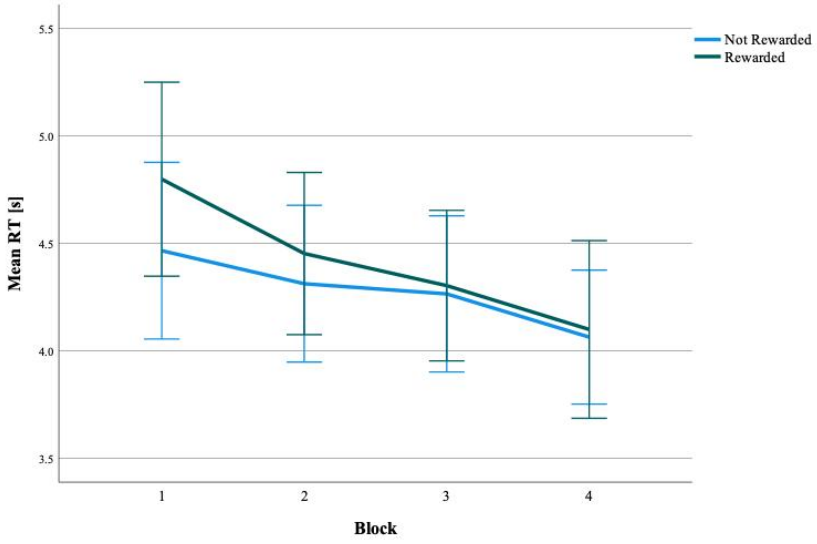

Figure 5. Mean RT on probe trials for rewarded and not rewarded condition across blocks. Error bars represent SEM. For the analysis, experimental blocks 1-8 were separated into the 1st, 2nd, 3rd and 4th rewarded block and the 1st, 2nd, 3rd and 4th unrewarded block.

Participants did not reliably improve in reaching the goal locations over time

In addition to examining cognitive control processes involved in sequential behaviour, our experiment also contained a strong decision-making component. Participants needed to decide

which path to take from the starting animal, to reach one of two high reward, goal locations and maximise their earnings. We therefore examined how often participants ended at one of the two goal locations, on trials where it was possible from the starting animal in four steps. The analysis considered correct trials only. Participants reached one of the goal locations on 63.0% (*SD* = 0.320) of correct trials, on which reaching the goal was possible. Numerical scores for reaching the goal location improved across the task (block 1: 57.5% *SD* = 0.344; block 2: 54.7%, *SD* = 0.355; block 3: 64.6%, *SD* = 0.279; block 4: 74.9%, *SD* = 0.282). However, this numerical improvement in reaching the goal location was not statistically reliable. A one-way ANOVA on the proportion of times the goal location was reached in rewarded blocks, with a factor of block number (levels 1 through 4), was non-significant ($F(3,17) = 1.442$, $p = 0.238$, $\eta_p^2 = 0.060$).

Participants did not reliably select locations with higher reward during binary choice

As a final exploratory analysis, we examined participants' binary choices at the end of each rewarded block. The main idea behind these trials was to test whether participants chose zoo locations associated with higher rewards, indicating knowledge about the reward structure of the task. For the analysis, we identified binary choice trials in which the two presented locations had different rewards during the preceding block. Trials in which a choice was made between two equally valuable locations were excluded. The resulting trials were scored based on whether participants chose the location associated with higher reward.

Participants selected the location linked to higher reward on 55.0% of eligible binary choice trials (*SD* = 0.234). A one sample t-test with a mean of 0.5 and unknown variance indicated that participants' binary choices were not significantly different from 0.5 ($t(71) = 1.762$, $p = 0.082$). Therefore, we were unable to confirm that participants selected locations linked to higher reward above chance. To test whether binary choice scores improved over time, we conducted a repeated measures ANOVA with rewarded block number as a factor (1 through 4). The analysis did not detect a significant effect of block on binary choice scores ($F(3, 36) = 1.336$, $p = 0.269$, $\eta_p^2 =$

0.074), indicating that participants did not show a robust increase in selecting more valuable zoo locations over time (block 1: *M* = 0.500, *SD* = 0.260; block 2: *M* = 0.502, *SD* = 0.272; block 3: *M* =0.590, *SD* = 0.200; block 4: *M* = 0.597, *SD* = 0.194). The analyses above pooled across three kinds of binary choice trials: high reward vs. no reward; high reward vs. low reward; and low reward vs. no reward. A full breakdown that examines the trial types individually is available in the appendix.

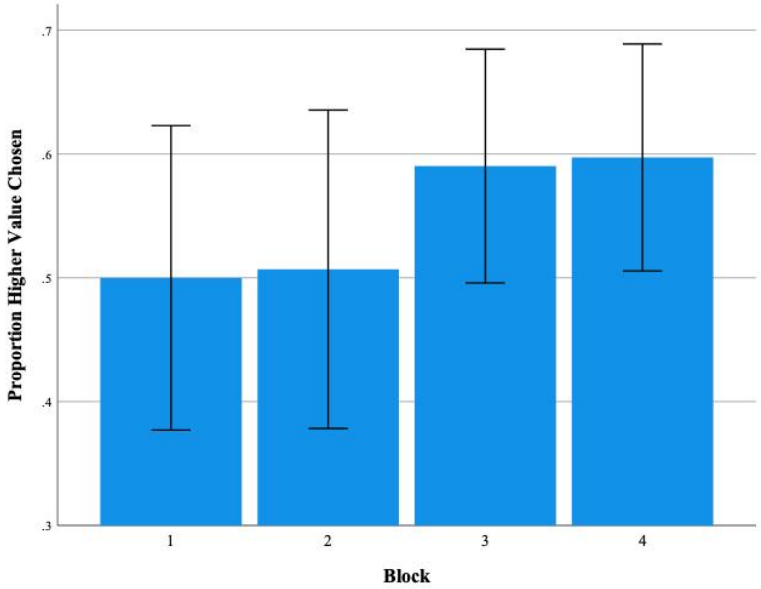

Figure 6. Binary choice results. Preference for locations associated with higher reward in the previous block. Error bas represent SEM. For the analysis, experimental blocks 1-8 were separated into the 1st, 2nd, 3rd and 4th rewarded block and the 1st, 2nd, 3rd and 4th unrewarded block.

3.3. Study 2: Main Analyses

The data from study 2 were analysed using the same structure and logic as study 1.

Accuracy improved over time

In study 2, the mean accuracy across all trials was 0.85 (*SD* = 0.354). We ran a repeated measures ANOVA to assess block-to-block changes in accuracy, with task block number (1 through 8) as a factor. This showed main effect of task block on accuracy ($F(3,44) = 24.734$, $p < 0.001$, $\eta_p^2 = 0.201$).

Bonferroni corrected comparisons revealed that the increase in accuracy was significant from block 1 (M = 0.77, SD = 0.171) to block 2 (M = 0.86, SD = 0.120, t(22) = -3.196, $p < 0.001$). Other comparisons were non significant (all $ps > 0.05$) This mirrors the results from study one, confirming improvements in accuracy early in the experiment, but not in the later blocks.

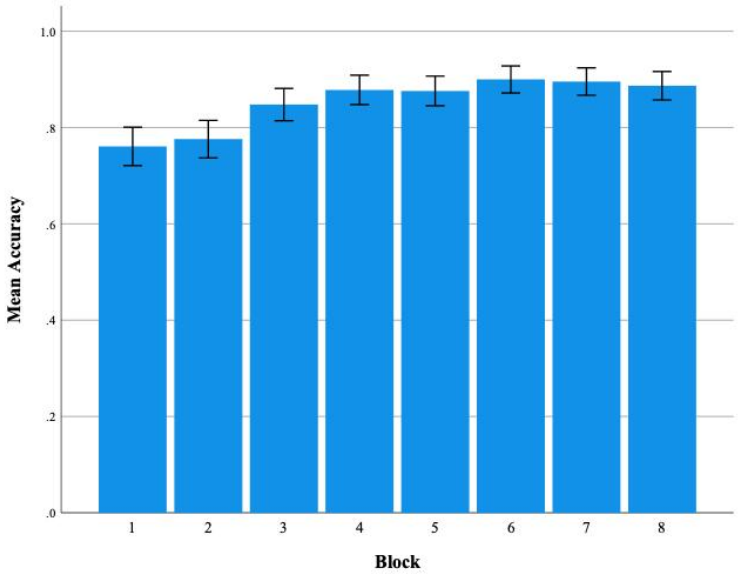

Figure 7. Mean accuracy on the task for each experimental block. Error bars represent SEM.

Reward did not modulate improvements in accuracy over time

To assess whether reward modulated accuracy improvements, we ran a 2x4 repeated measures ANOVA with factors of feature reward (rewarded vs unrewarded) and feature block number (1 though 4). We expected to replicate the primary finding from study 1, which showed a significant interaction between reward condition and block number. The ANOVA confirmed the main effect of block number ($F(3,44) = 16.792$, $p < 0.001$, $\eta_p^{22} = 0.433$). However, the analysis did not detect a main effect of reward on accuracy ($F(1,22) = 0.072$, $p = 0.791$, $\eta_p^2 = 0.003$) or significant interaction between the reward and block number ($F(3,54) = 1.655$, $p = 0.195$, $\eta_p^2 = 0.070$). Therefore, we were unable to replicate the effect of reward on accuracy across blocks in study 2. This stands in contrast to study 1, where we observed a small modulatory effect of reward on accuracy increases across the task.

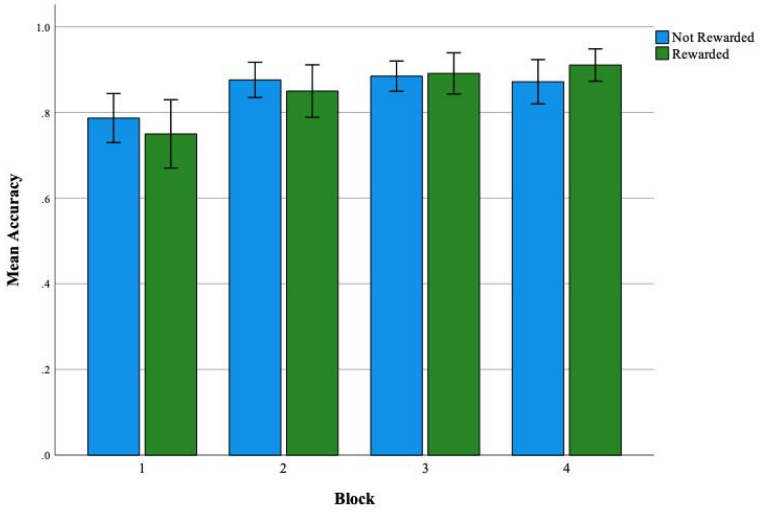

Figure 8. Accuracy blocks for rewarded and unrewarded condition. Error bars represent SEM. For the analysis, experimental blocks 1-8 were separated into the 1st, 2nd, 3rd and 4th rewarded block and the 1st, 2nd, 3rd and 4th unrewarded block.

Reward did not modulate probe trial accuracy

In the previous experiment, we found a weak modulatory effect of reward on probe trial performance. In study 2, we doubled the proportion of probe trials, to test the robustness of this effect with greater statistical power. Based on study 1, we expected to replicate the interaction between reward and block number on probe trial accuracy. We observed lower and more variable accuracy on probe trials ($M = 0.81$, $SD = 0.393$) than on standard trials ($M = 0.90$, $SD = 0.305$). A 2x4 ANOVA, with factors of feature reward (rewarded vs unrewarded) and feature block number (1-4), showed a significant main effect of block number on accuracy ($F(3,49) = 17.549$, $p < 0.001$, $\eta_p^2 = 0.444$). Bonferroni corrected pairwise comparisons show improved performance in the RT domain from block 1 to 2 ($MD = 0.117$, $SD = 0.023$, $p < 0.001$) to 1 to 3 ($MD = 0.172$, $SD = 0.034$, $p < 0.001$) and 1 to 4 ($MD = 0.157$, $SD = 0.029$, $p < 0.001$). Remaining pairwise comparisons were non-significant (all ps > 0.05). The analysis did not detect a significant main effect of reward ($F(1,22)= 0.642$, $p = 0.155$, $\eta_p^2 = 0.028$) or a significant interaction between reward and block

number ($F(3,62) = 1.564$, $p = 0.208$, $\eta_p^2 = 0.066$). In block 1 we found significant difference in accuracy on rewarded and unrewarded blocks ($MD = 0.083$, $SD = 0.197$, $t(22) = -2.012$, $p = 0.028$). In block 4 we did not find significant difference in accuracy between rewarded and unrewarded block ($MD = 0.013$, $SD = 0.158$, $t(22) = 0.397$, $p = 0.348$). Therefore, we did not find support for hypothesised modulatory effect of reward on probe trial accuracy.

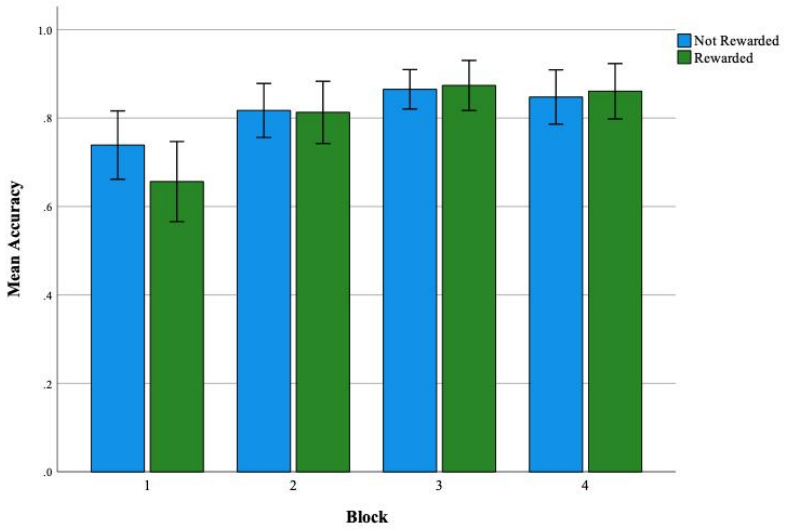

Figure 9. Accuracy on probe trials across blocks for rewarded and unrewarded condition. Error bars represent SEM. For the analysis, experimental blocks 1-8 were separated into the 1st, 2nd, 3rd and 4th rewarded block and the 1st, 2nd, 3rd and 4th unrewarded block.

RT on probe trials improved over time but was not modulated by reward

Two main differences between the current study and study 1 were additional incentives for fast performance and shortened response limits. In the current study, the bonus participants received when reporting a correct sequence that ended at a rewarded location, was based on value of the location (high or low reward), as well as the participant's RT. These changes were designed to make our task more sensitive to changes in reward-induced information coding, within the RT domain. Based on these changes, we expected that reward would modulate improvements in RT over time.

To test this prediction, we conducted a 2x4 repeated measures ANOVA on probe trial RT, using the same factors as in the previous results section (block number and reward). The dependent measure in these analyses was median RT, computed across correct probe trials in each block. The ANOVA detected a significant main effect of block number on RT ($F(3,58) = 16.810$, $p < 0.001$, $\eta_p^2 = 0.433$), suggesting that RT decreased across experimental blocks (block 1: $M = 4.079$, $SD = 0.655$; block 2: $M = 3.829$, $SD = 0.600$; block 3: $M = 3.683$, $SD = 0.549$; block 4: $M = 3.655$, SD = 0.582) (Fig. 10). Bonferroni corrected follow-up analyses confirmed significant reductions in RT from block: 1 to 2 ($MD = 0.250$, $SD = 0.068$, $p = 0.008$), 1 to 3 ($MD = 0.369$, $SD = 0.065$, $p < 0.001$), 1 to 4 ($MD = 0.423$, $SD = 0.76$, $p < 0.001$) and 2 to 3 ($SD = 0.146$, $MD = 0.047$, $p = 0.033$). The analysis did not detect a significant main effect of reward ($F(1, 22) = 0.268$, $p = 0.610$, $\eta_p^2 = 0.012$) or a significant interaction between reward and block number on RT ($F(3,59) = 0.547$, $p = 0.613$, $\eta_p^2 = 0.024$).

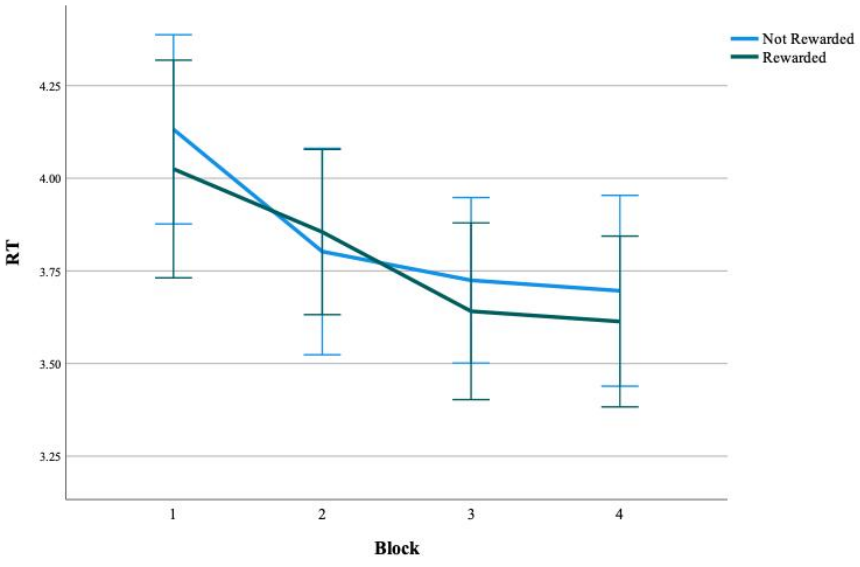

Figure 10. Mean RT on probe trials, in seconds, on rewarded and unrewarded condition across blocks. Error bars represent SEM.

3.4. Study 2: Exploratory Analyses

Participants did not reliably improve in reaching the goal locations over time

On average, participants ended a sequence high reward (goal) location on 58.1% of eligible trials (*SD* = 0.212), on which it was possible to reach a high reward location and participants entered a correct sequence. The percentage values showed that participants performance on reaching the high reward location improved numerically from block 1 to 2 and 3 to 4, but decreased from block 2 to 3 when the goal locations were changed (block 1: 59.4%, *SD* = 0.195; block 2: 64.2%, *SD* = 0.200; block 3: 50.5%, *SD* = 0.208; block 4: 58.1%, *SD* =0. 212). This numerical improvement in reaching the goal location was not statistically robust. A one-way ANOVA on the proportion of times the goal location was reached in rewarded blocks, with a factor of block number (levels 1 through 4), was non-significant ($F(3,37) = 1.974$, $p = 0.173$, $\eta_p^2 = 0.078$).

Participants did not reliably select locations with higher reward during binary choice

On average, participants selected zoo locations linked to higher reward on 51.5% of eligible binary choice trials (*SD* = 0.182), which occurred after rewarded blocks and involved stimuli linked to different reward magnitudes during the preceding task block. This percentage remained close to 0.5 across blocks (block 1: *M* = 0.511, *SD* = 0.198; block 2: *M* = 0.522, *SD* = 0.166; block 3: *M* = 0.533, *SD* = 0.199; block 4; *M* = 0.495, *SD* = 0.172). We did not detect evidence that participants selected higher reward locations above chance, or that their binary choice scores changed across blocks. A one sample t-test that examined whether the sample data differed from a population with a mean of 0.5 and an unknown variance was not significant ($t(91) = 0.787$, $p = 0.433$). A repeated-measures ANOVA with rewarded block number as a factor (1-4) did not find significant effect of block on location selection ($F(3,41) = 0.299$, $p = 0.728$, $\eta_p^2 = 0.013$) (Fig. 12). A full breakdown of choices for different types of binary choice trials is available in the appendix.

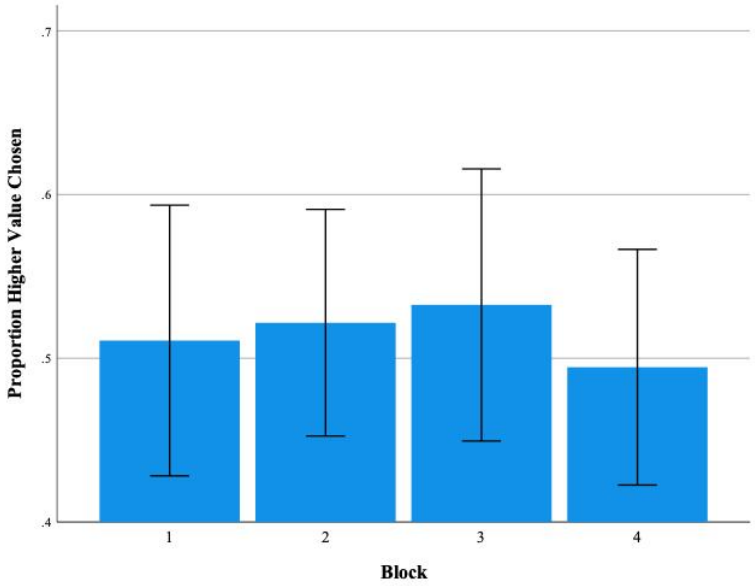

Figure 11. Preference for locations associated with higher reward in the previous block. Error bas represent SEM. For the analysis, experimental blocks 1-8 were separated into the 1st, 2nd, 3rd and 4th rewarded block and the 1st, 2nd, 3rd and 4th unrewarded block.

4.Discussion

In this project, we investigated the role of reward in shifting task representation into formats optimised for cognitive control and decision-making. To address this, we developed a complex behavioural task with interleaved rewarded and unrewarded blocks. In each block, some trials included distracting stimuli, which allowed us to probe the internal content participants focused on while making a decision. Based on the proposal that task representations change over time to optimise goal-directed behaviour (Whittington et al., 2022, Momennejad, 2020), we expected that information coding would change throughout the experiment, reflecting a shift from multiple stimulus features to the single relevant feature needed for decision-making and regulation of cognitive control. Drawing on the literature on modulatory effects of rewards in cognitive control within DMC (Braver, 2012) and the proposed role of hippocampal replay in optimising task representations (Momennejad, 2020; Wittkuhn et al., 2021), we hypothesised that reward would accelerate the shift in information coding outlined above, resulting in greater prospective encoding of decision-relevant information over time. In turn, this was expected to lead to more rapid performance improvements and buffer against interference across rewarded blocks, resulting in

diminishing effects of incongruent distractors. We tested these hypotheses in two studies. We found improvement on the task in both accuracy and RT domain, consistent with the proposal that the coding of task information changes over time, into formats readied to support active behaviour (Momennejad, 2020; Wittkuhn et al., 2021). However, the role of reward in accelerating this process was not consistent across studies.

In study 1, we found that increasing accuracy on the task was influenced by an interaction between reward and block number. Larger increases in accuracy were observed from the second to third rewarded block than the equivalent unrewarded blocks. This interaction was observed when considering all trials and when considering probe trials. This interaction result was a core prediction of this project, consistent with the proposal rewards accelerate changes in information coding. One interpretation for the significant interaction on probe trials is that the anticipation of reward reduced the negative impact of distracting stimuli more quickly over time. From a DMC perspective (Braver, 2012), this is consistent with the proposal that reward prospect leads to enhancements in proactive cognitive control, shielding performance on the task from the negative effects of interference. Under this interpretation, the present results suggest that this process could occur over longer (block-to-block) timescales than the trial-to-trial changes often studied under DMC (e.g. Hall-McMaster et al., 2019). Based on other results, however, it is not clear that rewards truly altered information coding to optimise task performance. For one, average accuracy across all trials was found to be significantly lower on rewarded blocks. In addition, performance on probe trials was not better during rewarded blocks than unrewarded blocks on average, with no significant difference between block types. When considering specific blocks, probe trial performance was significantly lower on the first rewarded block than the first unrewarded block, with rewarded performance becoming numerically higher in the third block set, and no difference in the last block of the experiment. Speculatively, the initial difference in accuracy could be partially attributed to the surprise caused by the introduction of rewards, or reflect the trial-and-error learning behaviour when participants attempted to find paths that led to the highest rewards. With these considerations in mind, study 1 showed a modulatory effect of reward that could be viewed as consistent with the proposal that rewards facilitate changes in task representation to optimise cognitive control. However, we are cautious not to over interpret the interaction effect observed in this study. It is an encouraging starting point for future work but it is not conclusive, since the performance on the rewarded blocks did not reliably surpass the performance on the unrewarded blocks, as would be

expected if rewards were truly optimising neural coding patterns for the purpose of improving behaviour.

We further explored whether the faster improvement on rewarded probe trials was evident in the RT domain. We observed an interference effect of distracting stimuli, as evidenced by longer RTs on probe trials. However, we did not find the beneficial effect of reward on RTs. This result is not entirely surprising, since the design for study 1 was not optimised to detect changes in RT. Since there was no additional incentive for fast performance in study 1, it is likely that participants reasonably prioritised accuracy over speed.

In study 2 participants could earn additional points for fast performance. In this experiment, we also increased the number of trials with distracting stimuli to investigate the modulatory effect of reward observed in study 1 with greater power. This revealed significant improvements in RT and accuracy over time, as well as reduced interference from distracting stimuli. These improvements are consistent with the notion that internal task representations were being optimised to reflect important features over time (Wittkuhn et al., 2021), which in turn affects cognitive control performance that relies on these representations (Notebaert & Braem, 2016). Despite these general performance improvements across the experiment, we did not detect significant behavioural effects of reward on accuracy or RT. Therefore, in study 2 we observed changes in accuracy and RT consistent with the idea that information coding was fine-tuned over time to support active behaviour. However, in contrast to the significant interaction effects observed in study 1, we did not see evidence indicating that reward had a possible role in modulating this process.

One rationale for why our results do not converge with the previous work demonstrating beneficial effects of reward on cognitive control (Braver, 2012; Hall-McMaster et al., 2019;Klink et al., 2017; Yamaguchi & Nishimura, 2019) is that our task differed from traditional experimental paradigms, in that it combined cognitive control and decision-making elements. The decision-making element was, arguably, stronger on the rewarded blocks relative to the unrewarded blocks. In rewarded blocks, participants not only had to report the sequences and deal with the interference but also weigh the value of different paths through the task's graph structure. In non-rewarded blocks, participants could rely more heavily on processes practiced during training. This difference in the level of difficulty between the two reward conditions could explain the worse performance on first rewarded block compared to the first non-rewarded block, observed in both studies. One further way to assess the influence of the decision-making component on task accuracy would be to

test the difference in performance on trials in the rewarded blocks where obtaining a reward was possible and where it was not possible, regardless of the path taken. If the deciding between paths introduced substantially greater difficulty, one would predict worse performance on rewarded trials than on non-rewarded trials, within the rewarded blocks. Due to the scope of the project, this analysis was not performed as part of the presented thesis. Future studies could reduce an asymmetry in decision difficulty between rewarded and non-rewarded blocks by introducing a decision-making component in all conditions. For example, the non-rewarded condition could be adapted to a low reward condition, so that participants need to weigh the value of different paths in all blocks, not just in the high reward blocks.

Despite the potential difficulty difference between conditions, a core strength of the presented work is the introduction of an experimental manipulation that can be used to probe potential shifts in internal task coding using behavioural measurements. Across analyses, we observed a diminishing effect of distracting stimuli over time, consistent with the notion that, as participants gained experience with the task, their encoding of task-information shifted into a more feature selective format that was less susceptible to interference from goal-irrelevant information. This finding substantiates the ability of our probes to tap into the internal content used for decision-making. Future work could validate our approach by introducing incongruence between the feature participants were preparing to report and the one they are asked to report. For example, participants could be asked to prepare a shape sequence but unexpectedly need to report a pattern sequence. We would expect performance on these incongruent trials to mirror the results of the interference probe trials, such that a shift in internal representations from multiple features to single feature representations is reflected in worse behavioural performance. One potential limitation of the probes used in our experiments and the new ones outlined above is that these manipulations are not necessarily a purely passive measurement of internal content, and their presence could also promote in information coding over time. While this limitation is important to be aware of, these manipulations still offer a useful readout of internal content that can used in behavioural studies, especially when more direct neural recordings are not possible.

We assessed the decision-making component in the present experiments by exploring participants' ability to report paths that lead to higher reward, and their preferences for stimuli associated with a higher reward on binary choice trials. Overall, participants had low levels of accuracy in making decisions that led to the highest reward (i.e. selecting paths that culminated in

the location associated with the highest reward) (63% in study 1 and 58% in study 2). There was no reliable improvement in making these decisions over time. Moreover, no significant preference for the stimuli associated with higher reward was detected during binary choice trials (55% in study 1 and 51.5% in study 2). One possible reason for the low scores on both decision-making metrics and the absence of reward-driven improvements is that the present experiments did not explicitly cue rewarded blocks or trials. Rather, reward was associated with one of the two features of the target stimuli (shape or pattern), with the most rewarding path on a given trial depending on the final sequence element reported. Participants were expected to learn this association while performing the task. It is possible that some participants simply did not manage to make the link between stimuli and reward to the degree sufficient for inducing changes in task representations, cognitive control benefits predicted by the DMC (Braver, 2012), and optimising decision-making overall.

Addressing this aspect in more detail, it is possible that the reward structure in this study was not optimal to induce learning and accelerate changes in information coding. Following Jimura, Locke & Braver's (2010) work on reward-related enhancement in cognitive control, we used static, performance-contingent, monetary rewards. This reward structure is empirically linked to improvements in cognitive control (Notebaert & Braem, 2016). However, recent studies on reward-learning and hippocampal replay suggest that facilitating learning to improve future decision-making might be best achieved via stochastic rewards with fluctuating probability, as these generate greater reward prediction errors (Diederen & Schultz, 2015, Liu et a., 2021). Reward prediction errors are signals that account for the difference in predicted and obtained rewards, communicated through midbrain dopamine release, that are important for guiding learning and adaptive behavior (Schultz, 2017). Recent theoretical and computational work has proposed that reward prediction errors and corresponding dopaminergic activity, are not only implicated in associative learning but might also facilitate adjustment of how information is encoded (Alexander & Gershman, 2021). This account posits that prediction errors provide a signal that regulates information coding in a way that preferentially represents important aspects of the task space, which in turn could help to guide top-down cognitive processes, such as proactive control and decision-making. If representational changes are linked to prediction errors and prediction errors are more effectively induced by fluctuating rewards, the static reward structure in our task could have been insufficient in generating prediction errors and thus an insufficient learning signal. Fluctuating rewards are also often used in the studies on replay (e.g. Liu et al., 2021), which informed our initial predictions, and

static rewards might not elicit reactivation processes to the same degree. With these considerations in mind, future work could explicitly examine the impact of static vs. fluctuating reward structures on accelerating changes in information coding, in the service of adaptive allocation of cognitive control. This would not only provide greater insight about how rewards act to reconfigure neural patterns, but would further connect decision-making and cognitive control literatures that are often investigated independently.

To conclude, this thesis bridges two cognitive domains that have traditionally been studied in isolation, cognitive control and decision-making. Across two studies, we did not find converging evidence for our hypothesis that improvement on complex cognitive tasks is linked to gradual reward-guided representational changes that optimise information processing for efficient goal-directed behaviour. Our results call for further exploration of how cognitive performance is affected by changes in the coding of task-relevant information and under which conditions reward affects these processes.

References

Alexander, W. H., & Gershman, S. J. (2022). *Representation learning with reward prediction errors*. arXiv. https://doi.org/10.48550/arXiv.2108.12402

Ambrose, R. E., Pfeiffer, B. E., & Foster, D. J. (2016). Reverse Replay of Hippocampal Place Cells Is Uniquely Modulated by Changing Reward. *Neuron*, *91*(5), 1124–1136. https://doi.org/10.1016/j.neuron.2016.07.047

Ashby, F. G., & O'Brien, J. B. (2007). The effects of positive versus negative feedback on information-integration category learning. *Perception & Psychophysics*, *69*(6), 865–878. https://doi.org/10.3758/BF03193923

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*, *100*(2), 490–509. https://doi.org/10.1016/j.neuron.2018.10.002

Bhattarai, B., Lee, J. W., & Jung, M. W. (2020). Distinct effects of reward and navigation history on hippocampal forward and reverse replays. *Proceedings of the National Academy of Sciences*, *117*(1), 689–697. https://doi.org/10.1073/pnas.1912533117

Botvinick, M., & Braver, T. (2015). Motivation and cognitive control: from behavior to neural mechanism. *Annual Review of Psychology*, *66*, 83–113. https://doi.org/10.1146/annurev-psych-010814-015044

Braver, T. S. (2012). The variable nature of cognitive control: A dual-mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010

D'Mello, A. M., Gabrieli, J. D. E., & Nee, D. E. (2020). Evidence for Hierarchical Cognitive Control in the Human Cerebellum. *Current Biology*, *30*(10), 1881-1892.e3. https://doi.org/10.1016/j.cub.2020.03.028

Diederen, K. M. J., & Schultz, W. (2015). Scaling prediction errors to reward variability benefits error-driven learning in humans. *Journal of Neurophysiology*, *114*(3), 1628–1640. https://doi.org/10.1152/jn.00483.2015

Dixon, M. L., & Christoff, K. (2012). The Decision to Engage Cognitive Control Is Driven by Expected Reward-Value: Neural and Behavioral Evidence. *PLOS ONE*, *7*(12), e51637. https://doi.org/10.1371/journal.pone.0051637

Etzel, J. A., Cole, M. W., Zacks, J. M., Kay, K. N., & Braver, T. S. (2016). Reward Motivation Enhances Task Coding in Frontoparietal Cortex. *Cerebral Cortex*, *26*(4), 1647–1659. https://doi.org/10.1093/cercor/bhu327

Foster, D. J. (2017). Replay Comes of Age. *Annual Review of Neuroscience*, *40*(1), 581–602. https://doi.org/10.1146/annurev-neuro-072116-031538

Fröber, K., & Dreisbach, G. (2016). How performance (non-)contingent reward modulates cognitive control. *Acta Psychologica*, *168*, 65–77. https://doi.org/10.1016/j.actpsy.2016.04.008

Gruber, M. J., Ritchey, M., Wang, S.-F., Doss, M. K., & Ranganath, C. (2016). Post-learning hippocampal dynamics promote preferential retention of rewarding events. *Neuron*, *89*(5), 1110–1120. https://doi.org/10.1016/j.neuron.2016.01.017

Hall-McMaster, S., Muhle-Karbe, P. S., Myers, N. E., & Stokes, M. G. (2019). Reward Boosts Neural Coding of Task Rules to Optimize Cognitive Flexibility. *Journal of Neuroscience*, *39*(43), 8549–8561. https://doi.org/10.1523/JNEUROSCI.0631-19.2019

Hefer, C., & Dreisbach, G. (2017). How performance-contingent reward prospect modulates cognitive control: Increased cue maintenance at the cost of decreased flexibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(10), 1643–1658. https://doi.org/10.1037/xlm0000397

Jimura, K., Locke, H. S., & Braver, T. S. (2010). Prefrontal cortex mediation of cognitive enhancement in rewarding motivational contexts. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(19), 8871–8876. https://doi.org/10.1073/pnas.1002007107

Karayanni, M., & Nelken, I. (2022). Extrinsic rewards, intrinsic rewards, and non-optimal behavior. *Journal of Computational Neuroscience*, *50*(2), 139–143. https://doi.org/10.1007/s10827-022-00813-z

Klink, P. C., Jeurissen, D., Theeuwes, J., Denys, D., & Roelfsema, P. R. (2017). Working memory accuracy for multiple targets is driven by reward expectation and stimulus contrast with different time-courses. *Scientific Reports*, *7*(1), 9082. https://doi.org/10.1038/s41598-017-08608-4

Krebs, R. M., Boehler, C. N., & Woldorff, M. G. (2010). The influence of reward associations on conflict processing in the Stroop task. *Cognition*, *117*(3), 341–347. https://doi.org/10.1016/j.cognition.2010.08.018

Liu, Y., Mattar, M. G., Behrens, T. E. J., Daw, N. D., & Dolan, R. J. (2021). Experience replay is associated with efficient non-local learning. *Science (New York, N.Y.)*, *372*(6544), eabf1357. https://doi.org/10.1126/science.abf1357

Momennejad, I. (2020). Learning Structures: Predictive Representations, Replay, and Generalization. *Current Opinion in Behavioral Sciences*, *32*, 155–166. https://doi.org/10.1016/j.cobeha.2020.02.017

Notebaert, W., & Braem, S. (2016). Parsing the effects of reward on cognitive control. In T. Braver (Ed.), Motivation and cognitive control (pp. 105–122). Routledge/Taylor & Francis Group.

Padmala, S., & Pessoa, L. (2011). Reward reduces conflict by enhancing attentional control and biasing visual cortical processing. *Journal of Cognitive Neuroscience*, *23*(11), 3419–3432. https://doi.org/10.1162/jocn_a_00011

Rule, M. E., O'Leary, T., & Harvey, C. D. (2019). Causes and consequences of representational drift. *Current Opinion in Neurobiology*, *58*, 141–147. https://doi.org/10.1016/j.conb.2019.08.005

Schoonover, C. E., Ohashi, S. N., Axel, R., & Fink, A. J. P. (2021). Representational drift in primary olfactory cortex. *Nature*, *594*(7864), 541–546.

Schultz, W. (2016). Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience*, *18*(1), 23–32. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4826767/

Shen, Y. J., & Chun, M. M. (2011). Increases in rewards promote flexible behavior. *Attention, Perception & Psychophysics*, *73*(3), 938–952. https://doi.org/10.3758/s13414-010-0065-7

Whittington, J. C. R., McCaffary, D., Bakermans, J. J. W., & Behrens, T. E. J. (2022). How to build a cognitive map: insights from models of the hippocampal formation. *ArXiv:2202.01682 [Cs, q-Bio]*.

Wittkuhn, L., Chien, S., Hall-McMaster, S., & Schuck, N. W. (2021). Replay in minds and machines. *Neuroscience & Biobehavioral Reviews*, *129*, 367–388. https://doi.org/10.1016/j.neubiorev.2021.08.002

Yamaguchi, M., & Nishimura, A. (2019). Modulating proactive cognitive control by reward: differential anticipatory effects of performance-contingent and non-contingent rewards. *Psychological Research*, *83*(2), 258–274. https://doi.org/10.1007/s00426-018-1027-2

Appendix

A1. Study 1: Binary Trials Analyses

Preferences following non-rewarded blocks were random

In the non-rewarded blocks, each end position of the sequence was equally beneficial (resulting in a 'correct!f 'eedback). As a sanity-check, we compared how often, on average, animal at each position was selected (when it was possible) to chance level of 0.5 (e.g we have compared average likelihood of selecting animal in position 1 to the likelihood that would result from a random selection). We have ran a one sample t-test comparing data for how often each position was selected to a sample with a mean 0.5 and unknown variance (i.e. a sample that would result from random choices). The mean of was not significantly different from 0.5 for any of the positions ($p>0.05$) (for position 1: $M = 0.444$, $SD = 0.482$, 2: $M = 0.521$, $SD = 0.455$, 3: 0.490, $SD = 0.565$, 4: $M = 0.550$, $SD = 0.451$, 5: $M = 0.489$, $SD = 0.462$). For all positions, the likelihood of it being chosen was close to the chance level of 0.5 and the variability in participants choices was high.

Preference for high reward locations did not increase over time

We further hypothesised that participants were most likely to select animals associated with the high reward location. For this analysis, we selected trials where a choice was made between: *high reward vs low reward* and *high reward vs no reward* locations. Overall, participants selected the high reward location on 57.0% of trials ($SD = 0.494$) where it was possible. We ran a repeated measures ANOVA with block number as a factor (1 through 4). The analysis did not find a significant effect of block on proportion of times participants selected the high reward location ($F(3,30) = 1.401$, $p = 0.253$, $\eta_p^2 = 0.076$). Therefore, the analysis suggests that proportion of times participants selected the high rewarded location remained stable over time (for block1: $M = 0.540$, $SD = 0.289$, 2: $M =$

0.500, *SD* = 0.328, 3: *M* = 0.630, *SD* = 0.246, 4: *M* = 0.640, *SD* = 0.234) (Fig.10b). We did not find expected improvement in selecting high-reward locations over time. However, participants selected animals linked to high-reward location significantly above chance on block 3 ($t(17) = 2.233$, $p = 0.039$) and 4 ($t(17) = 2.557$, $p = 0.024$).

Preference for high reward vs. low reward locations did not reliably increase over time

Further, we were interested in how often participants chose to select the animal linked to the high reward location when the alternative was a low reward animal. Across all blocks, participants selected the high reward animal over the low reward animal on 58.7% of trials, significantly above chance (*SD* = 0.327). ($t(17) = 2.283$, $p = 0.025$). Visual inspection of the means suggest increased preference for the higher reward location in the last two blocks (for block 1: *M* = 0.54, *SD* = 0.323, 2: *M* = 0.500, *SD* = 0.353, 3: *M* = 0.653, *SD* = 0.298, 4:*M* = 0.653, *SD* = 0.298)(Fig.A1). We ran an ANOVA with the same factors as the previous section. The ANOVA did not detect significant effect of block on high reward location preference ($F(3,29) = 1.065$, $p = 0.343$, $\eta_p^2 = 0.059$). Therefore, the numerical increase in preference for high reward locations was not statistically significant.

Preference for low reward vs. no reward locations did not change over time

We analysed the proportion of times participants selected low reward locations when alternative was not rewarded. The analysis selected trials where participants made a choice between low *reward vs. no reward* (the middle) location. Overall, participants selected locations associated with low reward on 49% (SD = 0.506) of the trials. A one-sample t-test did not find difference between the mean of our sample and a sample that would result from randomly selecting animals ($t(71) = -0.781$, $p = 0.478$). An ANOVA with the same factors as the previous section found no significant change in preference for low reward animal across blocks ($F(3,43) = 0.601$, $p = 0.592$, $\eta_p^2 = 0.034$). Therefore, in our sample there was no detectable difference in preference for low reward

location compared to the middle, unrewarded node (for block 1: $M = 0.389$, $SD = 0.404$, 2: $M = 0.528$, $SD = 0.436$, 3: $M = 0.472$, $SD = 0.401$, 4:$M = 0.472$, $SD = 0.436$)(Fig.A1).

Finally, we were interested in how often participants selected animals located in the top nodes. If the sequence began at the top node it was impossible to the reach rewarded locations in four steps. We looked at summary statistics for proportion of times top location was selected, when it was not associated with high reward. Across all blocks participants selected top location on 41.0% of the trials (SD =0.405). We did not find significant difference between participant's preferences and random selection ($t(17) = -1.890$, $p = 0.062$), when running a one-sample t-test with a population mean of 0.5 and unknown variance.

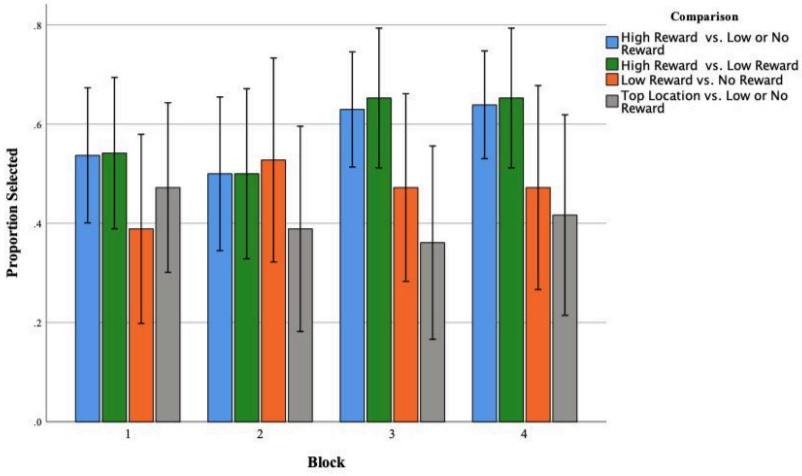

Figure A1. Proportion each zoo location was selected during binary choice trials, separated by specific reward comparison: *high reward vs. other, high vs. low reward*, *low vs. no reward*, t*op location vs. low or no reward.* Error bas represent SEM. For the analysis, experimental blocks 1-8 were separated into the 1st, 2nd, 3rd and 4th rewarded block and the 1st, 2nd, 3rd and 4th unrewarded block.

A2. Study 2: Binary Trials Analyses

Preference for high reward locations did not increase over time

We analysed whether participants had a greater preference for high reward locations. Overall, participants selected goal location on 52.3% of trials (*SD* = 0.241) where it was possible (for block 1: *M* = 0.507, *SD* = 0.276, 2. *M* = 0.554, *SD* = 0.243, 3. *M* = 0.522, *SD* = 0.223, 4. *M* = 0.507, *SD* = 0.232) (Fig. A2). This level of preference was not significantly differently from chance (t(91) = 0.996, p = 0.370). We repeated the procedure from the Study 1 to look into variation in preferences across blocks. Similarly, an ANOVA did not detect significant improvement in selection of goal location across time (*F*(3, 40) = 0.298, *p* = 0.748, η_p^2 = 0.012).

Preference for high reward vs. low reward locations did not increase over time

We were interested in exploring preferences for high vs. low reward locations. In this version of the experiment we increased the difference between high and low reward locations relative to Study 1. The bonus for reaching high and low reward locations was between €50- €100 and €15- €10, respectively. Participants selected high reward location over low reward location on 52.3%, (*SD* = 0.242) of the trials (for block: 1: *M* = 0.507, *SD* = .276, 2: *M* =.560,SD = .243, 3: *M* = .522, SD = .224, 4: *M* = .507, *SD* = .231)(Fig. A2). An ANOVA with the same factors as in the above section did not detect significant effect of block number on proportion of times participants chose high over low reward location (*F*(3, 41) = 0.269, *p* = 0.748, η_p^2 = 0.012).

Preference for low reward vs. no reward locations did not change over time

Finally, we were interested in how often participants selected *low reward vs. no reward* location. Participants selected locations associated with lower reward on 49.0% (SD = 0.358) of the trials (for

block 1: $M = 0.522$, $SD = 0.502$, 2: $M = 0.424$, $SD = 0.495$, 3: $M = 0.565$, $SD = 0.498$, 4: $M = 0.457$, $SD = 0.501$)(Fig. 11B). We ran an ANOVA to explore whether the proportion of times participants chose high reward location changed over time. ANOVA did not detect significant effect of block on selection of low reward location($F(3, 66) = 1.1194$, $p = 0.318$, $\eta_p^2 = 0.051$), reflecting stable preferences over time oscillating around 0.5.

We also explored participants preference for animals located at the top nodes. If the sequence began at the top node it was impossible to reach rewarded locations. We looked into proportion of times top location was selected, when it was not associated with high reward. Across all blocks participants selected top location on 46.7% of the trials ($SD = 0.380$). We did not find significant difference between participant's preferences and preferences that would arise from random choice ($t(22) = -0.827$, $p = 0.14$), when running a one-sample t-test with a population mean of 0.5 and unknown variance.

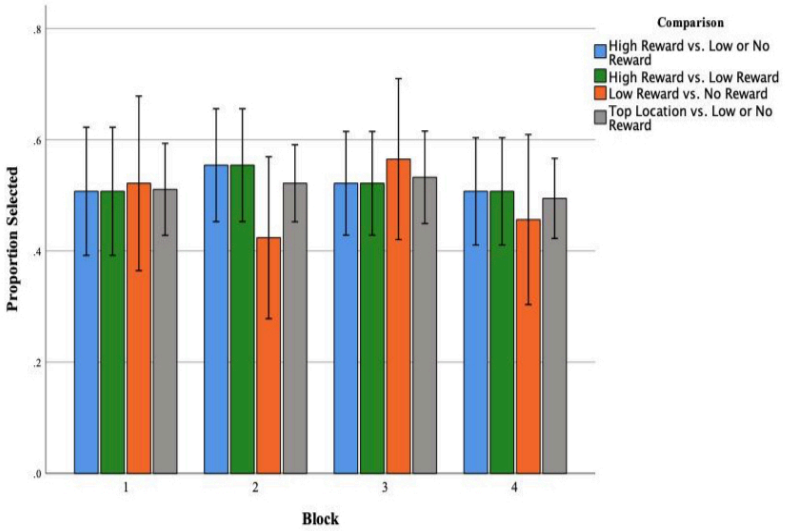

Figure A2. Proportion where location was selected, separated by specific comparison: *high reward vs. other, high vs. low reward*, *low vs. no reward*, t*op location vs. low or no reward.* Error bas represent SEM. For the analysis, experimental blocks 1-8 were separated into the 1st, 2nd, 3rd and 4th rewarded block and the 1st, 2nd, 3rd and 4th unrewarded block.

A3. Stimuli sources

Table A1. Stimuli sources - target images

Target Image	Source
Shapes	Drawn directly in matlab
Wave	Claudio Guglieri via The Pattern Library
Stripes	rawpixel.com via freepick.com
Dots	webvillapl via freepick.com
Plaid	henspec via pixabay.com
zig-zag	Glamazon via pixabay.com

Table A2. Stimuli sources - animal videos

Animal Video	Source
Lion	Shah Jahan via Pexels
Elephant	Salim Mediacity via Pexels
Zebra	Mikhail Nilov via Pexels
Giraffe	Magda Ehlers via Pexles
Parrot	Mikhail Nilov via Pexels
Seal	Ruvim Miksanskiy via Pexels
Penguin	Taryn Elliott via Pexels
Otter	Magda Ehlers via Pexles
Scottish highland cow	Matthias Groeneveld via Pexels
Lemour	Mikhail Nilov via Pexels