Inference over hidden contexts shapes the geometry of conceptual knowledge for flexible behaviour

Casper Kerrén [1*], Stephanie Theves [2], Mikael Johansson [3], Peter Gärdenfors [4], Christian Doeller [1,5],

1 *Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
2 Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany
3 *Department of Psychology, Lund University, Lund 221 00, Sweden*
4 *Cognitive Science, Lund University, Lund 221 00, Sweden*
5 *Kavli Institute for Systems Neuroscience, Centre for Neural Computation, Egil and Pauline Braathen and Fred Kavli Centre for Cortical Microcircuits, Jebsen Centre for Alzheimer's Disease, NTNU Norwegian University of Science and Technology, Trondheim, Norway*

***** *kerren@cbs.mpg.de*

Abstract

Flexible decision-making in uncertain environments requires inferring latent structure and selecting behaviourally relevant information. Here, we tested the hypothesis that internal models support this process by compressing high-dimensional input into lower-dimensional, goal-relevant subspaces. Human participants performed a serial reversal learning task requiring dynamic inference of the currently relevant stimulus dimension(s). Behavioural data showed rapid adaptation to context switches, with improving accuracy and faster responses over trials. A hidden state inference (HSI) model, which maintains probabilistic beliefs over latent contexts, best explained behaviour, outperforming multiple reinforcement learning variants in predicting both choices and inferred context. Eye-tracking revealed that gaze selectively converged on task-relevant features over time, with reduced attentional entropy and increased fixation within relevant regions, hallmarks of belief-guided attention. Crucially, gaze patterns were tightly linked to model dynamics: attentional focus increased when belief updating succeeded and decreased after trials where the model signalled high prediction errors. Guided by these results, we used magnetoencephalography (MEG) to understand the goal-relevant transformations within subspaces and found that neural representations expanded along relevant dimensions and compressed along irrelevant ones, with effects emerging just before decisions. These modulations of space correlated with reaction time, such that participants responded faster when the distance along relevant axes increased. Strikingly, these transformations preceded feedback-locked frontal theta responses modulated by prediction error, suggesting that internal models proactively reconfigure cortical state space in anticipation of learning. Together, these findings highlight a mechanism by which internal models, attention, and neural representations interact to support abstraction and adaptive decision-making.

Introduction

Human beings inhabit environments rich in diverse streams of information, yet the limited capacity of the cognitive system makes it impossible to process all inputs simultaneously. As a result, individuals must prioritise and selectively process information, flexibly adjusting behaviour in accordance with current goals. This adaptive capacity is especially apparent in decision-making, where humans and animals navigate complex environments with remarkable ease, often where classic reinforcement learning (RL) algorithms struggle (Bellman, 1957).

Rather than learning about all possible stimulus–outcome associations, it is often more efficient to represent the environment in a lower-dimensional manifold by focusing on task-relevant features (Konidaris, 2019; Niv, 2019). For example, a cost-conscious car buyer may focus on price and weight, whereas a performance-oriented buyer may prioritise speed, illustrating how contextual goals determine the processing of relevant features. This selective focus is thought to arise through representation learning, which mitigates the curse of dimensionality in RL by identifying a representational subspace aligned with current task goals (Bellman, 1957; Gershman & Niv, 2010; Leong et al., 2017; Niv, 2019; Sutton & Barto, n.d.; Wilson & Niv, 2012). Attentional filters are central to this process (Edgell & Morrissey, 1987; Kruschke, 1996; Nosofsky, 1986), enhancing dimensions deemed relevant by recent experience and internal goals. These filters, measurable via eye-tracking (Blair et al., 2009; Corbetta & Shulman, 2002; Kowler et al., 1995; Moore & Fallah, 2001; Rehder & Hoffman, 2005), dynamically adapt as participants infer which features are currently relevant, offering a behavioural window into structure learning. Through this mechanism, internal models are thought to interact with attention to support flexible, goal-directed behaviour (Bar-Gad et al., 2003; Corbetta & Shulman, 2002; Frank & Badre, 2012; Leong et al., 2017; Niv et al., 2015).

However, classic model-free RL approaches offer limited explanatory power in environments where reward contingencies shift or depend on unobservable variables. In contrast, inference-based approaches such as structure learning posit that agents form and update beliefs about latent states of the environment (Boyen et al., 2013; Braun et al., 2010). These beliefs integrate sensory observations with prior knowledge about the task's causal structure, allowing for inference-guided behaviour under uncertainty. Thus, behaviour is shaped not only by reinforcement history, but also by inferences about hidden causes (Daw et al., 2011; Mishchanchuk et al., 2024; Niv et al., 2015; Starkweather et al., 2018; Vertechi et al., 2020; Wilson et al., 2014).

Neuroscientific evidence increasingly supports the idea that belief-guided dimensionality reduction underlies flexible behaviour. To manage environmental complexity, the brain compresses high-dimensional input into lower-dimensional subspaces aligned with task goals (Behrens et al., 2018; Bellmund et al., 2018; Eichenbaum & Cohen, 2014; Garvert et al., 2017; Knudsen & Wallis, 2021; Leong et al., 2017; Mack et al., 2020; Menghi et al., 2021, 2023; Park et al., 2020; Radulescu et al., 2019; Schuck et al., 2016; Theves et al., 2020; X.-Y. Zhang et al., 2023). Within these subspaces, attention further reshapes representational geometry, enhancing relevant features while suppressing irrelevant ones. This process involves modulating representational axes, expanding dimensions aligned with current goals and compressing those that are not (Duan et al., 2024; Flesch et al., 2022; Kruschke, 1996; Love et al., 2004; Nosofsky, 1986; Park et al., 2023). Although such transformations are typically observed after extensive training (e.g., (Flesch et al., 2022; Park et al., 2023), recent work in non-human primates suggests they can also emerge dynamically (X.-Y. Zhang et al., 2023), supporting flexible behaviour in volatile environments.

In the present pre-registered study, we asked whether participants could dynamically retrieve from memory the relevant latent subspace (context) and modulate representational structure on a trial-by-trial basis. To this end, we designed a serial reversal learning task in which participants judged cars based on three feature dimensions, each with five discrete levels (Fig. 1), controlling for psychological distance in a separate task (Supplemental Figure S1). Crucially, while the stimulus space remained fixed, the task-relevant dimension varied according to an unobservable contextual variable. Participants were instructed to select cars for three different characters, where the value of a car depended on character-specific rules defined over one or two feature dimensions. Feedback was provided on

correct trials, allowing participants to infer the currently relevant dimension through experience. By holding the stimulus space constant while varying task goals, the task enabled us to examine goal-directed representation learning.

To characterise the mechanisms underlying flexible, context-sensitive behaviour, we first analysed participants' behavioural performance to assess their ability to adapt to latent context switches. We then used computational modelling to adjudicate between candidate learning mechanisms, comparing reinforcement learning models to a hidden state inference model that tracks beliefs over unobservable task states. To evaluate whether such belief dynamics guided perceptual processing, we analysed eye-tracking data to test whether attention selectively converged on task-relevant features and further related eye-correlates to the winning model. Finally, we used magnetoencephalography (MEG) to examine whether neural representational geometry reflected belief-guided transformations, specifically expansion along relevant dimensions and compression along irrelevant ones.

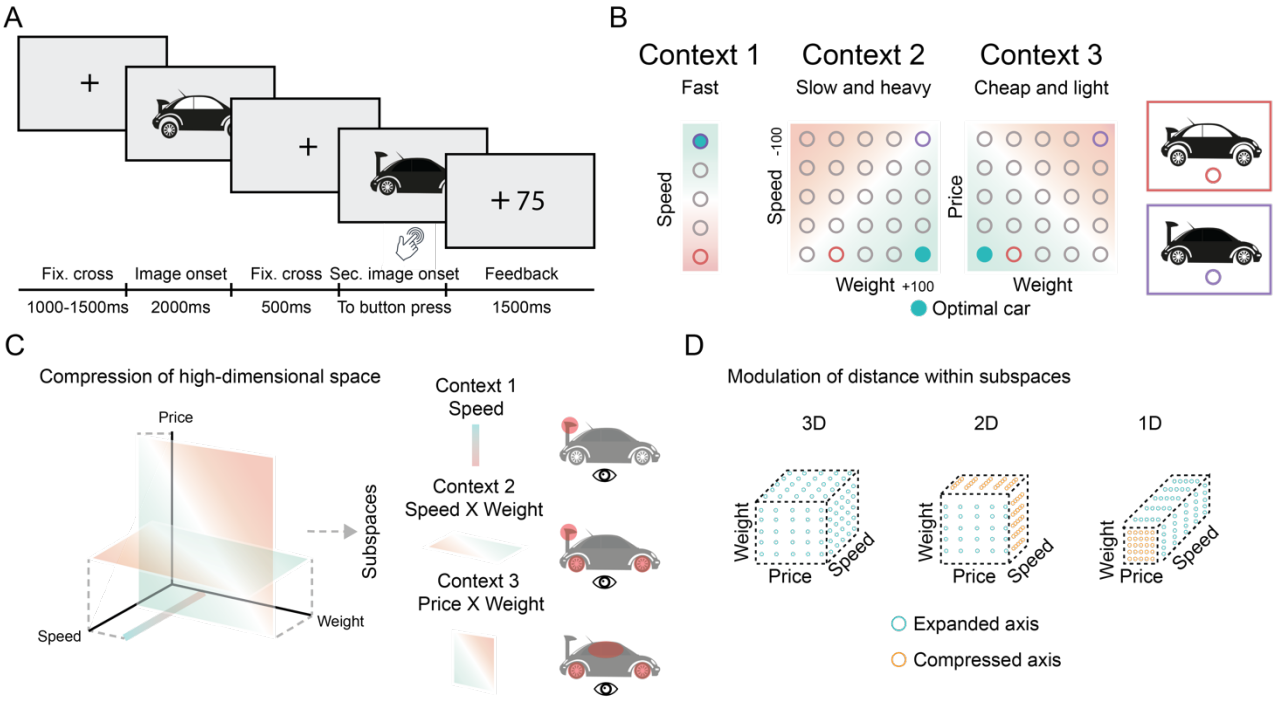

Figure 1. Context-dependent decision-making task. A. Each trial involved the presentation a fixation cross, one car, followed by another fixation cross and a second car. Participants were instructed to indicate which of the two cars had the higher value given the character (context) they thought they were playing. Timeline of a trial shown underneath. **B.** The reward values were defined according to three different stimulus–reward contingency maps. An example of three maps is shown, which across participants were counterbalanced. **C.** Hypothesised compression of a 3D space into subspaces for the different characters participants played. Green to orange gradient showing actual reward in a given context. Cars highlight the relevant dimensions, assessed with eye-tracking. **D.** Neurophysiological hypotheses propose that in an initially full representational space (3D), the distance between levels would be equal across axes, whereas when participants focused on relevant subspaces (2D or 1D), the axes within these subspaces would be expanded (relative to irrelevant ones; shown as green and orange circles, respectively).

Results

Participants rapidly inferred the correct context during the serial-reversal learning task.

We first calculated average reaction times (M = 1.14 secs, SE = 0.06; Fig. 2A) and accuracy (M = 0.77 secs, SE = 0.06; Fig. 2B) across all trials. To examine how performance evolved within each context, we computed reaction time and

accuracy for each trial position (1 to a maximum of 8) within context blocks, for each participant (Fig. 2C–D). We hypothesised a negative slope in reaction times and a positive slope in accuracy across within-context trials. To test these predictions, we fitted a generalised linear model (GLM) to each participant's reaction times and accuracies, and evaluated the resulting slopes using Wilcoxon signed-rank tests (signrank function in MATLAB) against a null hypothesis of zero slope. Consistent with our hypotheses, we observed a highly significant negative slope for reaction time ($z = -5.90$, $p < .01$; Fig. 2C) and a strongly positive slope for accuracy ($z = 6.05$, $p < .01$; Fig. 2D).

To confirm that participants were sensitive to contextual structure, we compared accuracy on each trial to chance level (50%), based on the fact that participants were informed at the end of each context block that a switch was about to occur. As expected, performance was at chance only on the first trial following a switch ($t(49) = 1.32$, $p = .19$; uncorrected), suggesting that participants effectively used feedback to infer the relevant context from the second trial onward. Together, these results provide strong evidence that participants understood the task structure and adapted rapidly to context changes.

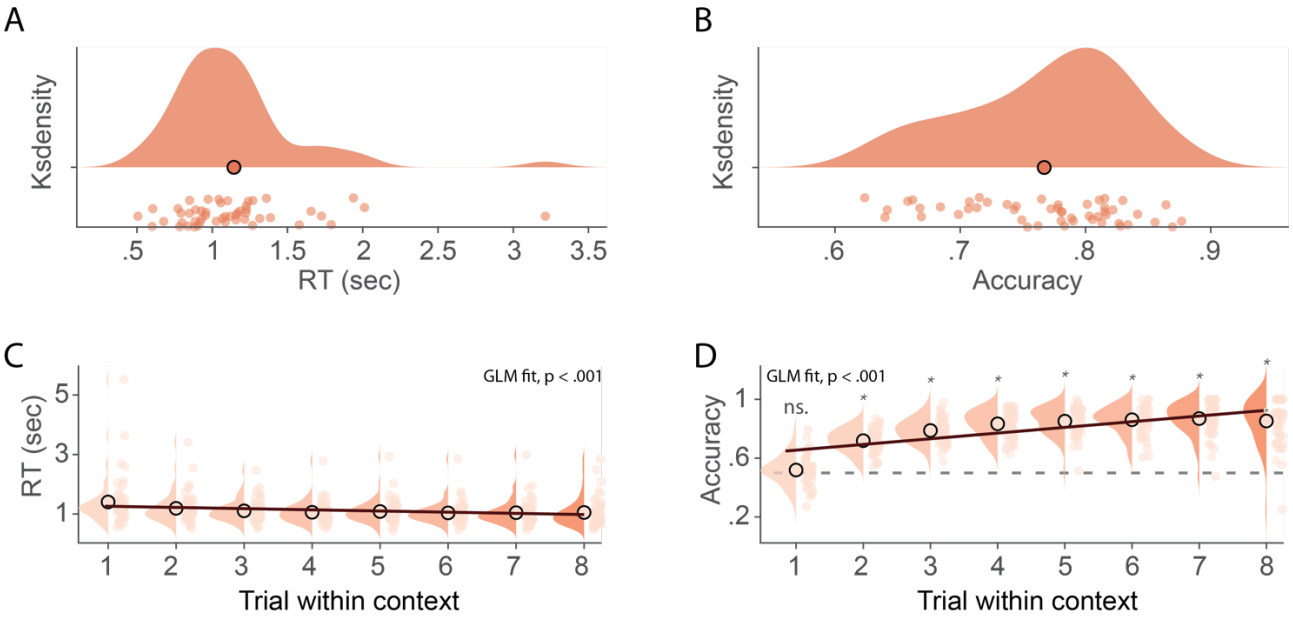

Figure 2. Behavioural results. A. Average reaction times (RTs) across all trials in the serial reversal learning task, with a mean of approximately 1.14 seconds. **B.** Reaction times plotted as a function of trial position within context blocks. A significant negative slope was observed, indicating that participants responded more quickly as they inferred the current context ($z = -5.90$, $p < .01$). **C.** Reaction time plotted as a function of trial position within a context block. A significant negative slope was observed, showing that reaction time decreased across trials ($z = -5.90$, $p < .01$). **D.** Accuracy plotted as a function of trial position within context blocks. A significant positive slope was observed, indicating improved performance with increasing evidence about the current context ($z = 6.05$, $p < .01$).

Participants' behaviour reflects hidden state inference.

To determine the computational mechanisms underlying this adaptive behaviour, we compared candidate learning models that varied in how they handled latent structure. Specifically, we asked whether participants' performance was best explained by reinforcement learning (RL) strategies, which update stimulus–reward values incrementally, or by a hidden state inference (HSI) strategy, which maintains probabilistic beliefs over unobservable contexts.

On each trial, participants had to infer which of three contexts was currently active, using only outcome feedback and cues indicating when a switch was about to occur. On the first trial after a switch, chance performance was expected (Fig. 2D). To learn over time, participants could adopt one of two broad strategies: a reinforcement learning (RL) approach, such as Q-learning, which updates value estimates incrementally based on observed outcomes; or a

hidden state inference (HSI) approach, which constructs and maintains beliefs over a latent context variable that guides both choice and attention.

The four computational models (see Table S1 for model parameters) were selected to capture a range of possible strategies for learning in environments with latent structure. These models differed in how they update stimulus–reward associations and whether they represent hidden states explicitly. First, standard Q-learning (Q) provides a baseline incremental reinforcement learning strategy, in which action values are updated only based on experienced outcomes (Sutton & Barto, 2018; Watkins & Dayan, 1992). Second, Q-learning with forgetting (QF) extends this model by allowing the values of unchosen options to decay over time, capturing a potential recency bias or memory limitation (Kato & Morita, 2016). Third, Q-learning with counterfactual updating (QC) incorporates learning from unchosen outcomes by updating the values of both chosen and unchosen options based on their expected outcomes, thus allowing the model to simulate learning via exploration or internal evaluation (Palminteri et al., 2015). Finally, the hidden state inference (HSI) model assumes that participants infer which latent context is currently active, maintaining and updating probabilistic beliefs over these contexts based on feedback and cues. This model has been shown to account for rodent behaviour in similar tasks with latent reversals (Mishchanchuk et al., 2024) and directly implements the idea that participants may infer task structure rather than relying solely on incremental value updating.

To model choices, we used context-specific stimulus–reward contingency matrices and marginalised over the three possible contexts, weighted by their posterior probabilities. Choice stochasticity was captured via a softmax with a free inverse temperature parameter (β_action). If a participant's belief was weighted toward the correct context, the model was more likely to predict the correct action.

Model fits were evaluated using the Bayesian Information Criterion (BIC). The HSI model provided the best fit to the data (M = 308.10, SEM = 6.38), followed by the counterfactual Q-learning model (QC; M = 464.78, SEM = 7.73), the Q-learning with forgetting model (QF; M = 526.08, SEM = 10.53), and standard Q-learning (Q; M = 495.64, SEM = 11.80). The HSI model outperformed QC by a wide margin ($t(49) = -21.70$, $p < .01$) and was the best-fitting model for all participants (Fig. 3A).

The HSI model also achieved the highest accuracy in predicting context (M = 77.46%, SEM = 8.0%), compared to the second-best model QC (M = 75.35%, SEM = 11.5%), QF (M = 73.68%, SEM = 6.8%), and Q (M = 63.72%, SEM = 15.2%). However, the difference between HSI and QC in predicting context was not statistically significant ($t(49) = 1.55$, $p = .13$; Fig. 3B–C).

A similar pattern was observed for predicting participants' actions. The HSI model again achieved the highest accuracy (M = 74.56%, SEM = 8.3%), followed by QC (M = 64.35%, SEM = 9.5%), QF (M = 61.13%, SEM = 6%), and Q (M = 59.70%, SEM = 7%). The HSI model predicted actions significantly better than the QC model ($t(49) = 18.06$, $p < .01$), with all participants showing higher action prediction accuracy under HSI (Fig. 3D–E).

We also compared model predictions of action to actual behavioural data. The smallest discrepancy between predicted and observed actions was found for the HSI model (M= 3.23%, SEM = 1.39%), followed by QC (M = 13.67%, SEM = 1.58%), QF (M = 16.95%, SEM = 1.63%), and Q (M = 18.19%, SEM = 1.66%; Fig. 3F).

To confirm model identifiability, we conducted a model recovery analysis (Wilson & Collins, 2019) by simulating data with best-fitting parameters for each model and re-fitting all models to each dataset. The resulting confusion matrices showed strong diagonal structure (Fig. 3G), confirming that each model best recovered its own data and that HSI was uniquely identifiable, likely due to its structural divergence from the three RL-based models.

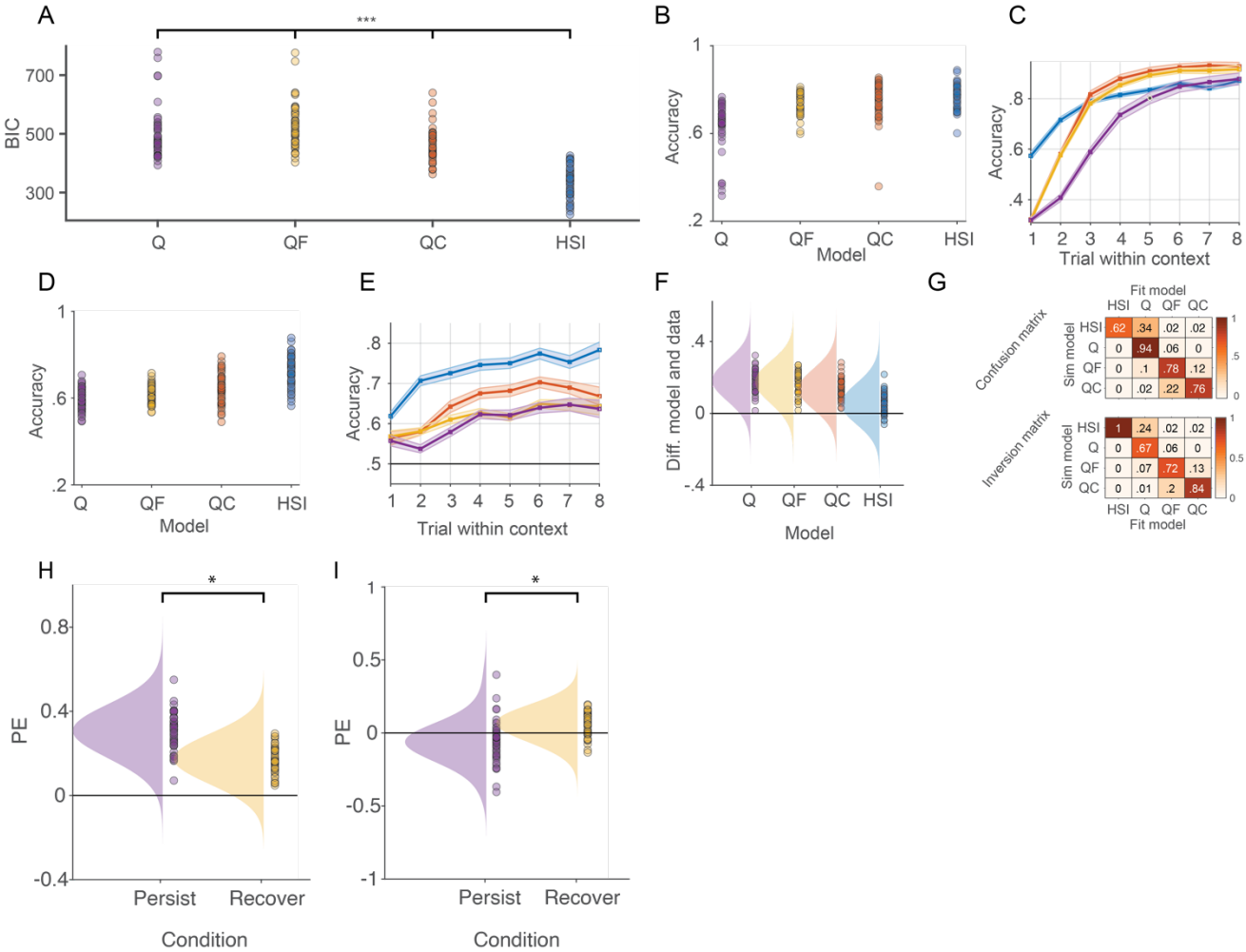

Figure 3. Hidden state inference model best fit the behavioural data. A. Bayesian information criterion (BIC) for Q-learning (Q), Q-learning with forgetting (QF), Q-learning with counterfactual updating (QC) and Hidden state inference model (HSI) showed significantly lower BIC scores for HSI compared to the other models, indicating that model comapared to the other models as the best fit to the data. **B.** Average of fraction of correctly predicted contexts for the four models. **C.** Same as (B), but now showing for each trial within a context. **D.** Average of fraction of correctly predicted button presses for the four models. HSI was significantly better at predicting button presses than QC. **E.** Same as (D), but now showing for each trial within a context. **F.** Comparing predicted button presses to actual button presses showed that HSI was most aligned with actual data. **G.** We assessed model recoverability and identifiability by simulating data from each model and then fit it with each model. Highest values along the diagonal indicates that each model could best recover its own simulated data (top) and that the best-fitting model typically matched the generative model (bottom). **H.** Evaluating the model-based prediction error (PE) on trials following an initial error. When the model persisted to be wrong on the following trial, unsigned PE was significantly higher compared to when it recovered and predicted the right action. **I.** When instead evaluating the signed PE, we found that when the model recovered on the following trial, the PE was mostly positive, significantly more so compared to when it continued to predict the wrong action, where we saw mainly negative PE.

To further probe the relationship between belief updating and model-based prediction error (PE), we computed trial-wise PE as the difference between observed choice accuracy and the model's predicted probability of a correct response. We analysed PE in two complementary ways. First, using unsigned PE to quantify surprise irrespective of outcome valence, we examined whether surprise decreased when the model successfully updated its beliefs following an initial misprediction. On trials where the model was incorrect on trial 1 (i.e., immediately after a context switch), PE on trial 2 was significantly lower when the model recovered and predicted correctly (M = 0.17, SE = 0.01) than when it remained incorrect (M = 0.30, SE = 0.01; $t(49) = 9.30$, $p < .001$; Fig. 3H). This indicates that PE tracked meaningful fluctuations in belief updating and model adaptation.

We repeated the analysis using signed PE to examine outcome valence. When the model recovered on trial 2, signed PE was significantly positive (M = 0.06, SE = 0.01), indicating a better-than-expected outcome. In contrast, when the model remained incorrect, signed PE was significantly negative (M = −0.06, SE = 0.02; t(49) = −4.99, p < .001), reflecting a worse-than-expected outcome (Fig. 3I). These findings suggest that trial-wise PE not only captured the success of latent belief updating (via unsigned PE), but also encoded the valence of outcome evaluation (via signed PE), supporting a role for prediction error in guiding inference, attention, and adaptive behaviour.

Taken together, these results indicate that participants' behaviour was best explained by a model-based inference strategy, with the HSI model outperforming all alternatives in capturing both choices and context tracking.

Gaze selectively tracks task-relevant features

Next, we asked whether participants' eye behaviour reflected a model-based inference strategy. If participants inferred the currently active context, this should guide their attention toward the stimulus dimensions relevant for that context, and this, in turn, should be reflected in their eye movements. To perform the task successfully, participants needed to selectively attend to context-relevant dimensions (Fig. 1C). To assess this process, we analysed eye movements within predefined regions of interest (ROIs; the rear wing, window, and wheels; see Supplemental Fig. 2), which mapped onto either relevant or irrelevant features depending on whether the trial belonged to a 1D or 2D context block.

Previous studies have shown that gaze reliably tracks task-relevant dimensions during learning (e.g., (Braunlich & Love, 2022; Leong et al., 2017; Rehder & Hoffman, 2005)). We replicated and extended this finding: using gaze samples across trials within a context, the total number of attended dimensions decreased, converging toward the actual number of relevant dimensions by the final trials (Fig. 4A, left). We quantified this by segmenting each trial into an early window (from first image onset to fixation cross) and a late window (from second image onset to response). Participants attended to significantly fewer dimensions during the late window (M = 1.36, SE = .04) compared to the early window (M = 1.84, SE = .05; t(49) = 12.57, p < .01; Fig. 4A, right), suggesting a narrowing of attention onto task-relevant features as the trial unfolded.

To characterise how this attentional tuning evolved across trials, we computed the proportion of gaze samples landing on relevant versus irrelevant ROIs separately for each trial position (1–8) and participant, within the late time window. Overlapping features were counted as irrelevant only when they did not belong to the currently relevant feature set. These proportions were then averaged across participants and collapsed across 1D and 2D contexts. As shown in Fig. 4B (left), gaze allocation toward relevant features increased systematically with trial progression, while gaze to irrelevant features declined. Collapsing across trials, participants devoted significantly more gaze to relevant (M = .68, SE = .02) than to irrelevant features (M = .16, SE = .01; t(49) = 16.50, p < .01; Fig. 4B, right).

Next, we asked whether gaze to relevant features predicted correct performance. We fit a generalised linear model (GLM) to the difference in gaze allocation between correct and incorrect trials using gaze density measures from the late window. Wilcoxon signed-rank tests on model coefficients revealed significantly greater fixation within relevant ROIs on correct trials (z = 2.86, p = .004; Fig. 4C), suggesting a tight link between attentional allocation and task success.

To probe this link further, we focused on inference trials, a subset of trials where participants had been incorrect on the first trial of a new context but responded correctly on the current trial (Courellis et al., 2024). These trials are informative because they index the moment participants revise their beliefs about the relevant feature. We compared these to trials where participants were incorrect on both the first and current trial within a new context. A time-resolved analysis, corrected using cluster-based permutation testing, revealed a buildup of gaze toward the relevant dimensions starting at first image onset, with significant differentiation emerging around the second image onset ($p_{cluster} < .05$; Fig. 4D).

Finally, we assessed attentional uncertainty using gaze entropy, a measure of disorganisation in gaze patterns reflecting how consistently across time within a trial participants focused on relevant features (Shiferaw et al., 2019), with low entropy values reflecting more stable focus. We predicted that entropy would be highest immediately after a context switch and decline as participants inferred the relevant feature. We computed binary time series indicating whether gaze landed on relevant features and calculated entropy separately for correct and incorrect trials. After z-scoring entropy vectors, we fit GLMs to assess changes across trial position and tested slope values against zero. As predicted, entropy declined significantly across correct trials ($z = -6.15$, $p < .01$), while entropy increased across incorrect trials ($z = 5.19$, $p < .01$; Fig. 4E), suggesting that attentional stabilisation was tightly coupled with successful inference.

Taken together, these results show that participants' gaze dynamically tracked the task's latent structure, converging on the correct features as their internal model of context solidified. This goal-directed modulation of attention highlights the role of inferring the hidden context not just in action selection, but in guiding perception.

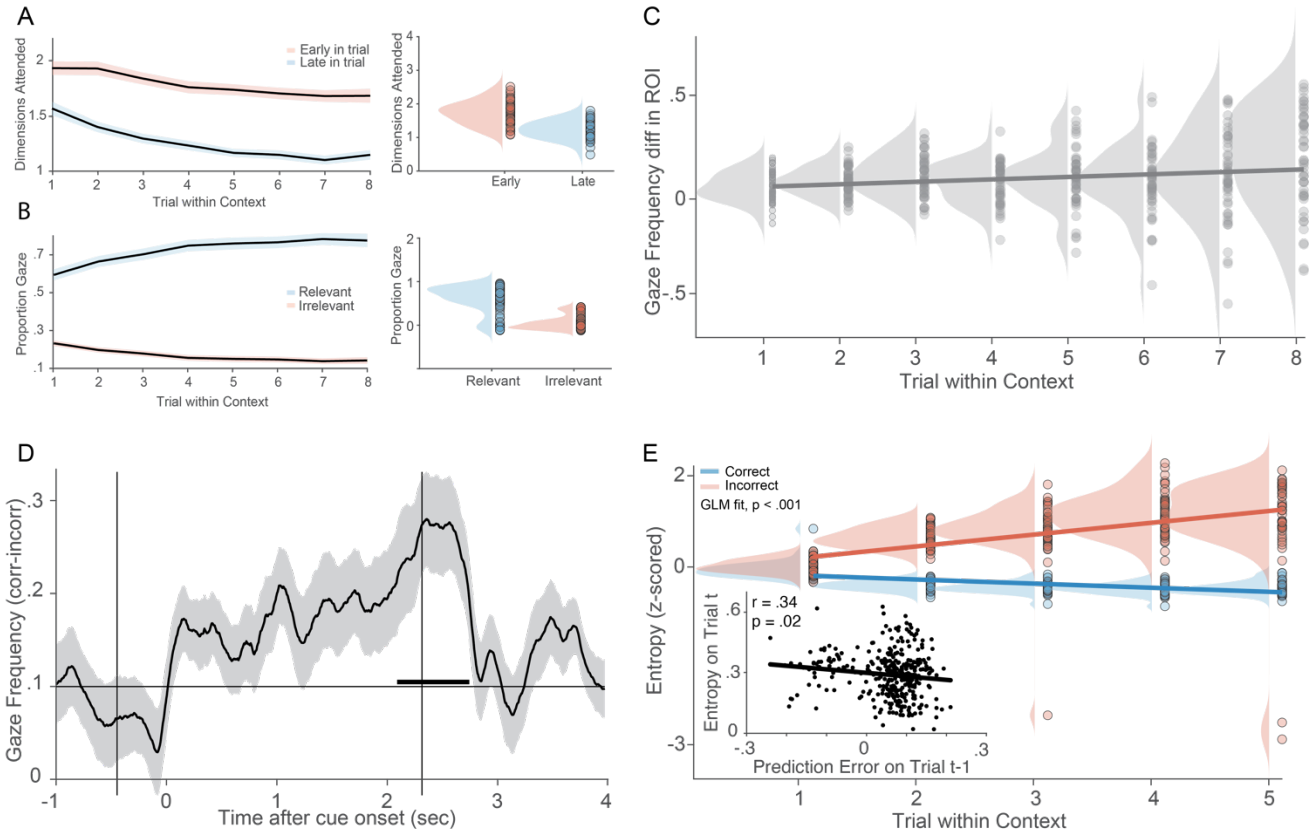

Figure 4. Context-guided eye-behaviour. A. Dimensions attended to for early (after first image onset to fixation cross; red) and late time points in trial (after second image onset to reaction time; blue), per trial within a context (left) and average across trials within contexts (right). At late time points participants focused more on the relevant dimensions (1 dimension on y-axis reflects 1 dimension for 1D trials and 2 dimensions for 2D trials). **B.** The proportion of gaze frequency for all trials between relevant (blue) and irrelevant (red) dimensions, shown for each trial within a context (left) and averaged across trials (right). Participants focused significantly more at the relevant dimensions as compared to the irrelevant dimensions **C.** Gaze frequency difference between correct and incorrect trials showed a significant positive slope, indicating that when participants where correct, their gaze was focused more within the relevant regions. **D.** When participants made an inference they shifted their gaze to the relevant ROI, as compared to when they did not make an inference, with a significance onset following second image onset. **E.** Entropy as a function of trial within a context. When participants were correct, entropy significantly decreased across trials (blue), whereas when they were incorrect, entropy significantly increased (red). **Inset in E**. Signed PE predicted gaze entropy such that more negative PEs were related with more broadened visual sampling, whereas positive PEs were related to more focused attention.

Linking internal models to visual attention

We showed that participants' eye movements reflected selective attention to context-relevant stimulus dimensions, consistent with an inference-based strategy. As a next step, we asked whether trial-by-trial variation in gaze behaviour tracked internal model dynamics. Specifically, if gaze reflects belief-guided attention, then trials with more focused gaze should coincide with stronger model predictions (i.e., higher certainty about the correct action according to the HSI model).

Gaze entropy was significantly lower on trials where the model correctly predicted the participant's button press ($M = 0.30$, $SE = 0.01$) compared to trials where it did not ($M = 0.33$, $SE = 0.01$; $t(49) = 4.05$, $p < .01$), indicating that more focused visual attention was associated with more consistent, model-aligned behaviour.

We next tested whether signed PE modulated subsequent gaze behaviour. A linear mixed-effects model revealed that prediction error (PE) on trial t−1 significantly predicted gaze entropy on trial t ($\beta = -0.17$, $p = .02$), such that more negative PEs were associated with more diffuse gaze distributions, while more positive PEs were followed by more focused attention (Fig. 4E, inset). Rather than selectively attending to a single feature dimension, participants appeared to broaden their attentional sampling in response to negative outcomes, consistent with exploratory behaviour guided by latent state inference.

Motivated by this link between belief dynamics and gaze, we then incorporated gaze metrics directly into the HSI model, allowing gaze entropy and gaze distribution to modulate belief updating and/or action selection. However, none of these augmented models outperformed the original HSI model in terms of BIC or choice prediction accuracy.

In sum, gaze data mirrored model dynamics: as beliefs about the latent context stabilised, attention became increasingly focused on task-relevant features. Although gaze patterns were consistent with model-predicted belief states, incorporating gaze into the model did not improve predictive performance. This suggests that gaze and choice were jointly governed by an underlying internal model, rather than gaze providing additional information beyond what was already captured by inference over hidden states.

Having established that belief dynamics were reflected in both behaviour and gaze, we next turned to our neurophysiological data to test whether these internal models also shaped brain activity, both in terms of ongoing inference and outcome processing.

Frontal theta power is elevated on the first trial of each context block

Before turning to model-derived prediction errors, we asked whether frontal theta power was sensitive to basic task structure. Specifically, whether it was modulated by trial position within each context block. Frontal midline theta (FMT) activity has been consistently linked to cognitive control and conflict monitoring processes, with a prominent role in signalling the need for behavioural adjustment (Cavanagh & Frank, 2014; Cavanagh & Shackman, 2015). The first trial of each block is uniquely informative: participants receive a contextual switch cue but must infer the new context from feedback alone. If frontal midline theta reflects belief updating or decision uncertainty, as suggested by prior work (Cavanagh & Frank, 2014), it should be elevated on these early trials.

We compared stimulus-locked theta power on the first trial of each context block versus the rest of the trials. This analysis revealed significantly greater theta power over midfrontal electrodes on trial 1 compared to subsequent trials ($p_{cluster}<.05$; Supplemental Figure 3), consistent with elevated uncertainty and cognitive control demands during early context inference.

This finding establishes that frontal theta is sensitive to broad structural features of the task. Specifically, theta is elevated on the first trial of a context block when uncertainty and the need for inference are highest. However, this analysis does not speak to how participants update their beliefs based on specific outcomes. If frontal midline theta reflects not only general control demands but also the dynamics of belief revision, then it should vary on a trial-by-trial basis as a function of outcome-driven prediction errors.

Frontal theta reflects trial-wise belief updating

We therefore next asked whether theta power tracked model-derived prediction errors from the HSI model, which quantifies the extent to which an outcome deviates from a participant's current belief about the active context. In reinforcement learning contexts, FMT has been associated with trial-wise prediction error (PE) signals, reflecting violations of expectation that guide learning. Specifically, several studies have shown that FMT power increases with larger unsigned PEs, consistent with a role in surprise or salience detection (Cavanagh & Frank, 2014). Some evidence also points to valence sensitivity in FMT, with greater theta power following positive outcomes or better-than-expected events (Cockburn & Holroyd, 2018; Heydari & Holroyd, 2016; Lopez-Gamundi et al., 2024). These findings raise the possibility that FMT may encode not only the magnitude of surprise but also its direction; whether an outcome was better or worse than expected.

In our study, we aimed to characterise how model-derived signed prediction errors (PEs) from the HSI model relate to frontal theta activity during outcome processing. By regressing trial-wise signed PEs against post-outcome EEG amplitude over midfrontal electrodes, we tested whether frontal midline theta (FMT) reflects surprise, outcome valence, or both, as indexed by the model's internal belief updates. This analysis revealed a significant cluster over frontal electrodes, indicating that FMT power systematically varied with both the sign and magnitude of model-derived PEs, consistent with a role in outcome evaluation and belief updating (Fig. 5A).

To further probe this relationship, we grouped trials by signed PE into negative, neutral, and positive outcomes, using the 33rd and 67th percentile, and compared theta power across groups. This revealed a valence-sensitive pattern: theta power was significantly higher following positive PEs (M = .04, SE = .003) compared to negative PEs (M = .02, SE = .004), $t(49) = 3.39$, $p = .0014$ (Fig. 5B). This suggests that FMT increases not only with the magnitude of belief update (surprise), but also with its direction, supporting a model in which FMT encodes whether an outcome was better or worse than expected.

Together, these findings align with prior evidence linking FMT to outcome valence and reward positivity (Cockburn & Holroyd, 2018; Heydari & Holroyd, 2016; Lopez-Gamundi et al., 2024), and extend them by showing that such effects emerge even when PE is estimated from a latent-state inference model rather than overt feedback alone.

Lastly, we asked whether the extent to which participants' frontal theta tracked prediction errors was related to behavioural performance. To test this, we computed, for each participant, the slope (β-value) of the regression between signed PE and midfrontal theta amplitude, and correlated it with their average accuracy in the serial reversal learning task. This analysis revealed a significant positive correlation ($r = .34$, $p = .02$), suggesting that participants who more strongly encoded prediction errors in frontal theta tended to perform better. This supports the idea that theta-mediated outcome monitoring facilitates belief updating and promotes adaptive behaviour.

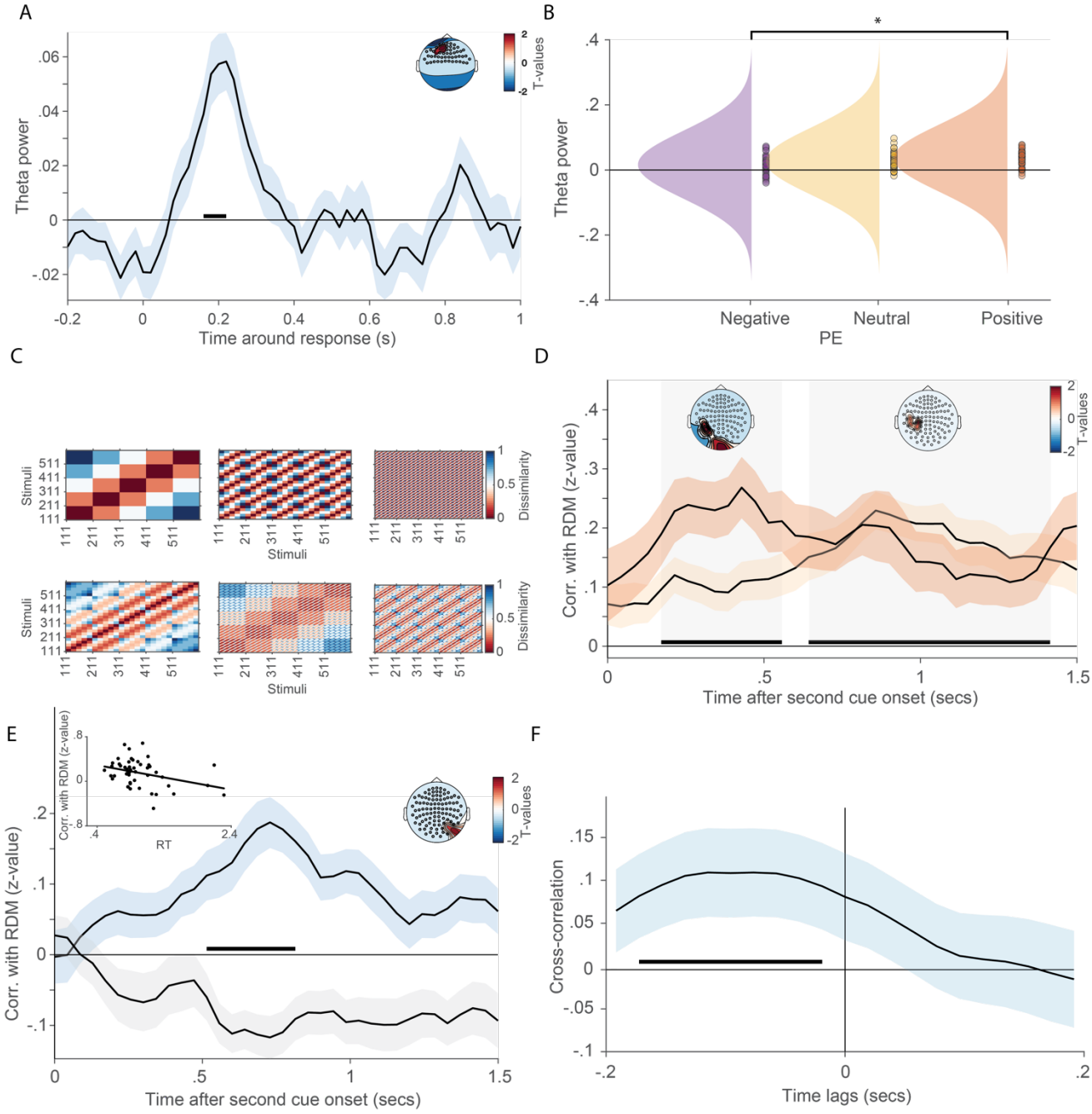

Figure 5. Frontal midline theta, PE, and representational dimensionality expansion and compression. A. Model-derived signed prediction error (PE) regressed against frontal midline theta (FMT) revealed a significant cluster approximately 200ms after response onset. **B.** PE grouped into negative, neutral and positive, using the 33th and 67th percentiles, showed that FMT increased was mainly seen for positive PE. **C.** Representational dissimilarity matrices (RDMs) for one and two dimensions. **D.** When correlating correct (blue) and incorrect (red) trials with DSMs for relevant dimensions for 1- and 2D trials collapsed, we observed an expansion for correct and a compression for incorrect with a significance commencing early after second image onset. This effect had a left centro-parietal and occipito-parietal topography. **E.** When instead correlating correct trials with RDMs for relevant and irrelevant dimensions for 1D trials, we observed a similar temporal pattern, with a right-lateralised posterior activation. **Inset E.** This difference was negatively correlated with reaction time, such that the larger the difference the faster the reaction time. **F.** Cross-correlation between FMT-PE and dimensionality change for correct-incorrect-trials revealed a significant negative time lag with a positive amplitude, suggesting increased difference between correct and incorrect trials preceded increased FMT-PE.

Representational expansion and compression of relevant and irrelevant dimensions.

Having established a signature of dimensionality reduction via latent state inference and eye-behaviour, we next asked how the brain dynamically manipulates these inferred subspaces (Fig. 1D). Specifically, whether neural representations show expansion along task-relevant dimensions and compression along irrelevant ones. To test this, we examined the structure and temporal evolution of the neural representational geometry. We constructed a set of model representational dissimilarity matrices (RDMs; Fig. 5C) designed to capture larger Euclidean distances between stimulus levels along relevant dimensions, relative to irrelevant ones (see Methods).

For 1D and 2D trials collapsed, we focused on one contrast: relevant dimension on correct versus incorrect trials. For 1D only trials, we focused on relevant versus irrelevant dimensions on correct trials only. For 2D trials, we were limited to contrasting correct and incorrect trials, as all three dimensions were task-relevant at different times, making it infeasible to construct RDMs involving two unattended dimensions.

Neural RDMs were computed by estimating the dissimilarity between the first and second image of each trial. We then correlated the neural RDMs at second image onset with the model RDMs using Spearman's partial correlation. Crucially, to determine whether the resulting correlations reflected true representational expansion or compression beyond chance, we generated a null distribution by shuffling the neural RDMs prior to correlation. This process was repeated 250 times per participant, after which we subtracted the mean of the shuffled correlations from the empirical values and divided by the standard deviation of the shuffled baseline, yielding z-scored values for each condition and participant. Positive z-values reflected stronger correlation with the model RDMs than what would be expected by chance and therefore indicated an expansion of the distance between levels, whereas negative z-values indicated compression. Statistical inference was performed using cluster-based permutation tests.

When participants attended to the relevant dimensions, we observed positive correlations between the neural RDMs and the model RDMs on correct trials, which was significantly higher than for incorrect trials (two clusters, both $p < .05$). This significant difference exhibited a left centro-parietal and occipito-parietal topography and emerged around 150 ms after the onset of the second image ($p_{cluster} < .05$; Fig. 5D). In both significant time windows, correct trials were significantly greater than zero ($t(49) = 8.45$, $p < .01$; $t(49) = 6.68$, $p < .01$), while incorrect trials were not significantly below zero ($p > .1$), supporting the presence of an expansion of the neural trace when participants made a correct choice as compared to incorrect.

When instead contrasting relevant and irrelevant dimensions on correct 1D trials, we observed a similar temporal pattern (Fig. 5E): neural signals correlated positively with the model RDM for relevant dimensions and negatively for irrelevant dimensions, suggesting both an expansion and a compression, respectively ($p_{cluster} < .05$). In the significant time window, relevant dimensions were significantly above zero ($t(49) = 4.52$, $p < .01$), while irrelevant dimensions were significantly below zero ($t(49) = -3.97$, $p < .01$). The magnitude of this expansion (i.e., the difference in correlation between relevant and irrelevant dimensions) was negatively correlated with reaction time ($r_{spearman} = -.30$, $p = .03$), such that participants responded more quickly when neural representations more strongly differentiated the task-relevant dimension (Fig. 5E, inset).

Finally, to examine whether frontal midline theta modulated by prediction error (FMT-PE) was related to changes in neural representational geometry, we performed a cross-correlation between each participant's time-resolved dimensionality change and their FMT-PE signal. This analysis was conducted separately for the two significant correct-minus-incorrect dimensionality time courses in Figure 5D. Because the FMT-PE signal was time-locked to response onset and reflected post-decision neural activity, we hypothesised that changes in representational geometry, particularly the expansion difference observed in correct-incorrect trials, might precede FMT-PE, consistent with a model in which dimensionality change supports anticipatory inference, while FMT reflects reactive outcome evaluation.

As previously shown, FMT was not only sensitive to the magnitude of surprise, but also tracked the valence of the prediction error, with stronger theta power following better-than-expected outcomes. We therefore hypothesised that increased differences in representational geometry between correct and incorrect trials would precede

increased FMT-PE, reflecting a temporally coordinated cascade from latent state inference to outcome-based updating.

Performing the cross-correlation analysis against a null distribution of 1000 phase-shuffled surrogates between dimensionality change and frontal midline theta modulated by prediction error (FMT-PE) revealed a significant positive relationship, with a peak at a negative time lag. Specifically, dimensionality change in the early significant time window, indexed by increased Euclidean distance between neural representations for correct versus incorrect trials, preceded the FMT-PE signal ($p_{cluster} < .05$, two-sided; Fig. 5F), peaking at –120 ms. This suggests that changes in representational geometry occurred before the emergence of theta activity related to outcome evaluation. A similar, though non-significant, trend was observed for the later dimensionality change effect.

In sum, our findings support an adaptive neural mechanism by which relevant information is expanded on correct as compared to incorrect trials, while irrelevant information, is actively compressed. This, in turn, precedes the FMT power modulated by prediction error. This supports the interpretation that representational expansion reflects an anticipatory inference process that precedes and possibly facilitates the subsequent engagement of outcome-related theta responses.

Discussion

In uncertain and dynamic environments, adaptive behaviour depends on the ability to infer latent structure and selectively attend to task-relevant information. Here, we show that participants rapidly adapted to context changes in a serial reversal learning task by dynamically updating internal models, selectively focusing visual attention, and reorganising neural representations. Crucially, these adaptations were driven not by observable stimuli or reinforcement history, but by inference over latent structure. Together, these findings demonstrate that inference over hidden states supports flexible, goal-directed dimensionality reduction across behavioural, attentional, and representational domains (Niv et al., 2015; Wilson et al., 2014).

Participants quickly inferred the current context, as evidenced by systematic improvements in accuracy and reaction time across within-context trials. A hidden state inference (HSI) model captured these behavioural adaptations more accurately than multiple reinforcement learning (RL) variants by explicitly maintaining probabilistic beliefs over latent task states (Boyen et al., 2013; Braun et al., 2010). The HSI model not only predicted choices and inferred contexts more accurately than its RL counterparts, but also showed high model recoverability and participant-level identifiability (Fig. 3). That human behaviour aligned most closely with this structure-sensitive model, even in a task with three contexts and frequent switches, supports the view that participants used internal models to infer latent structure, not just to accumulate reward histories (Daw et al., 2011; Vertechi et al., 2020). These results parallel recent findings in rodents (Costa et al., 2015; Mishchanchuk et al., 2024; Vertechi et al., 2020) extending the cross-species evidence for inference as a core mechanism of adaptive control.

Crucially, this inferential process was reflected in participants' eye movements. Over trials, gaze selectively converged on task-relevant dimensions, and this attentional narrowing was more pronounced on correct trials. Gaze entropy declined over time within context blocks, indicating an increasingly structured sampling policy guided by contextual knowledge (Najemnik & Geisler, 2005; Nelson & Cottrell, 2007; Rehder & Hoffman, 2005; Yang et al., 2016); Fig. 4A–B). Moreover, trial-level fluctuations in gaze patterns aligned with model-inferred belief dynamics: entropy was lower when the model correctly predicted behaviour, and increased following high prediction error (PE), consistent with exploratory sampling in the face of uncertainty (Fig. 3E, inset). These results suggest that visual attention did not passively reflect learned associations but instead tracked inferred relevance (Corbetta & Shulman, 2002; Moore & Fallah, 2001).

However, while attention and belief were tightly coupled, their relationship appeared mechanistically integrated rather than sequential or additive. Incorporating gaze entropy or spatial distribution into the HSI model did not improve choice prediction, implying that gaze and choice were governed by a shared internal belief state, rather

than gaze providing independent information. In this sense, attention was not an external modulator of inference, but an overt behavioural expression of it, resembling a readout of the evolving internal model (Bar-Gad et al., 2003).

To test whether these belief-guided adaptations extended beyond behaviour and gaze, we next examined whether neural representations were similarly modulated by inferred relevance. A well-researched neural response to conflict monitoring and uncertainty is the frontal midline theta power (FMT; Cavanagh & Frank, 2014). In parallell, FMT has also been implicated in reinforcement learning processes, such that the FMT tracks participants' prediction error (PE) on a trial-by-trial basis. In our study, the first trial in a new context was arguably where participants were most uncertain as they did not know which context they were playing. Contrasting these first trials to the rest we found an increased FMT, peaking in a similar time-window as has been previously shown (Supplemental Figure 3; Cavanagh & Shackman, 2015). Crucially, when testing whether FMT tracked the model-derived PE, we found that FMT power systematically varied with both the sign and magnitude of model-derived PEs, consistent with a role in outcome evaluation and belief updating (Fig. 5A). Interstingly, theta power was significantly higher following positive PEs as compared to negative PEs (Fig. 5B), suggesting that FMT increases not only with the magnitude of belief update (surprise), but also with its direction, supporting a model in which FMT encodes whether an outcome was better or worse than expected. Together, these findings align with prior evidence linking FMT to outcome valence and reward positivity (Cockburn & Holroyd, 2018; Heydari & Holroyd, 2016; Lopez-Gamundi et al., 2024), and extend them by showing that such effects emerge even when PE is estimated from a latent-state inference model rather than overt feedback alone.

Next, using magnetoencephalography (MEG), we found that the representational geometry of the decision space changed dynamically as a function of task demands. When participants correctly identified the relevant dimension, neural representational distances expanded along task-relevant axes and compressed along irrelevant ones (Fig. 5C-F). These transformations emerged shortly after the onset of the second stimulus, with a parietal topography, and were predictive of faster responses. Moreover, expansion was significantly greater on correct trials and was correlated with attentional focus, as measured by gaze (Duan et al., 2024; Park et al., 2023). These results provide compelling evidence that internal models drive dynamic reorganisation of cortical representations in line with inferred goals (M. Zhang & Yu, 2024).

To probe the temporal relationship between belief-guided representational change and outcome-related theta activity, we performed a cross-correlation analysis between time-resolved dimensionality change and FMT-PE. Strikingly, we found that early increases in neural dissimilarity between correct and incorrect trials reliably preceded FMT responses modulated by prediction error. This temporal precedence, peaking around 120 ms before the theta response, suggests that representational expansion may reflect anticipatory inference, setting the stage for subsequent outcome evaluation and learning. That is, before outcomes are even processed, internal beliefs have already shaped the structure of neural representations, potentially preparing the system to register and integrate feedback more efficiently. These findings add a mechanistic temporal bridge between representational geometry and feedback-related control signals.

The interactions between belief, attention, and behaviour suggest a more general principle: that internal models guide sampling policies across multiple timescales and modalities. This idea is formalised in the Sampling Emergent Attention (SEA) model (Braunlich & Love, 2022), which proposes that attention emerges through belief-guided sampling of information sources to maximise expected utility. As in SEA, we observed that participants dynamically narrowed their attention to task-relevant dimensions over time, with gaze increasingly focused on informative features as internal models of context stabilised. Crucially, however, our results extend this framework by showing that selective sampling was guided not by fixed stimulus features, but by inferred latent contexts, adding a hierarchical layer to SEA's account. Trial-wise fluctuations in prediction error modulated subsequent gaze patterns, supporting SEA's bidirectional relationship between learning and attention. Furthermore, we found converging neural evidence: on correct trials, relevant information was expanded and irrelevant information compressed in representational space, consistent with the attentional warping proposed by SEA. These findings suggest that SEA provides a compelling account of context-sensitive attention, and may be fruitfully extended to accommodate inference over hidden structure.

While previous work has demonstrated that task-relevant features become increasingly emphasised in neural representations during categorisation (Duan et al., 2024; Mack et al., 2020), our study advances this literature in several key ways. Prior approaches often attribute representational compression to explicit task demands or data-driven feature selection. In contrast, we show that inference over latent structure, not stimulus visibility or feature informativeness, drives representational transformation. Compared to (Mack et al., 2020), which tracks dimensionality compression over learning blocks, we uncover trial-by-trial representational expansion and compression aligned with internal beliefs about latent context, with neural geometry adapting within hundreds of milliseconds. Unlike (Duan et al., 2024), which maps stimulus feature visibility onto task-specific manifolds, we hold stimuli constant and instead manipulate goal structure via hidden contexts, allowing us to disentangle relevance from visibility.

Relatedly, (Flesch et al., 2022) found that extensive training led to orthogonal neural manifolds aligned with context-specific feature dimensions in fronto-parietal cortex. While their study revealed how task-specific representations emerge through rich learning, our findings demonstrate that similar transformations can occur rapidly and reversibly, without prolonged training, via dynamic belief updating. Moreover, our use of computational modelling demonstrates that beliefs over hidden states not only drive attention and decision-making, but also shape representational geometry in real time. In addition to linking belief dynamics with static representational geometry, we examined how outcome processing signals relate to these neural transformations. Frontal midline theta (FMT), a neural signature implicated in cognitive control and feedback monitoring (Cavanagh & Frank, 2014), tracked model-derived prediction errors (PEs) computed from our latent-state inference model (Fig. 5A-B). Specifically, theta power increased as a function of signed PEs, with stronger activity following better-than-expected outcomes (Cockburn & Holroyd, 2018; Heydari & Holroyd, 2016; Lopez-Gamundi et al., 2024). This pattern suggests that FMT reflects not just surprise or salience, but also the valence of belief updates. Moreover, participants who exhibited stronger PE sensitivity in FMT achieved higher task accuracy, supporting the idea that FMT contributes to adaptive belief updating and behavioural flexibility. By linking model-inferred belief dynamics with gaze entropy and MEG representational structure, our work provides a unified, mechanistic account of how latent inference sculpts attention and cortical representations to support abstraction and flexible control.

Together, these findings support a mechanistic account of abstraction in which internal models, attention, and neural representation interact to support flexible control. Rather than encoding the environment in a static or fully detailed manner, participants compressed the task space into context-appropriate subspaces, expanding dimensions aligned with the current goal and suppressing irrelevant features (Konidaris, 2019; Mack et al., 2020; Niv, 2019). This form of belief-guided dimensionality reduction enables efficient learning and decision-making in high-dimensional, non-stationary environments (Leong et al., 2017; Radulescu et al., 2019).

An open question is whether attention merely reflects the outcome of inference, or whether it also serves as a learning signal that shapes representational change. Our findings are compatible with both interpretations. On one hand, attention closely tracked belief dynamics, consistent with a downstream manifestation of inference. On the other, increased focus on relevant dimensions may facilitate the neural differentiation of those dimensions, suggesting a reciprocal loop between inference and attention (Kruschke, 1996; Love et al., 2004; Nosofsky, 1986). This view resonates with recent work showing that hippocampal representations disentangle latent causes from stimulus identity to support generalisation (Courellis et al., 2024). Here, we extend this perspective by linking inference to real-time attentional selection and rapid reconfiguration of cortical representational geometry (Flesch et al., 2022).

In sum, we propose that latent state inference underpins flexible decision-making by shaping both what we look at and how we internally represent the task. By continuously aligning attention and representation with inferred relevance, the brain reduces the effective dimensionality of the problem space, supporting abstraction, adaptability, and efficient control in complex environments (Bellmund et al., 2018; Gershman & Niv, 2010; Schuck et al., 2016).

Acknowledgments

The authors would like to thank Burkhard Mess and Yvonne Wolff-Rosier for help in collecting data and customise scripts for optimal data collection and Max Schulz for translation of instructions. They would also like to thank Daniel Carlström Schad, Evie Coxon, Nicholas Menghi, Daniel Reznik and Simone Vigano for their constructive input and discussions on this work.

Materials and methods

Subjects

Fifty participants (23 females, 27 males; mean age = 27.9 years, SD = 4.36 years, range = 19–36) took part in the study in exchange for financial compensation. All participants had normal or corrected-to-normal vision and reported no history of neurological or psychiatric disorders.

All experimental procedures, including behavioural piloting, were approved by the Ethics Committee of the Faculty of Medicine at Leipzig University, Germany (approval number: 211/23-EK), and were conducted in accordance with the Declaration of Helsinki. For the online piloting phase, participants provided informed consent by electronically ticking a consent box. For the MEG experiment, written informed consent was obtained prior to participation.

Stimuli

The stimulus set consisted of 125 images of a car, each varying along three visual features (hereafter referred to as dimensions). These dimensions were: (i) the height of the rear wing, (ii) the distance between the car body and the wheels, and (iii) the brightness of the windows. These visual features were designed as perceptual proxies for abstract car attributes: wing height represented speed, wheel-body distance represented weight, and window brightness represented cost. Each dimension varied across five levels, resulting in a fully crossed set of 125 unique stimuli (5 × 5 × 5). Stimulus selection was partly motivated by prior research on geometric representations of similarity spaces (Gärdenfors, 2000), allowing us to manipulate dimensionality while focusing on goal-directed learning processes.

To ensure that perceptual spacing across levels was uniform within each dimension, we conducted a behavioural study employing maximum likelihood difference scaling (MLDS). This method estimates latent perceptual scale values based on participants' dominance judgments across triads of stimuli (e.g., judgments of whether stimulus b is more similar to a or c). Participants were presented with a sufficiently large number of triads, and the resulting data were fit using maximum likelihood estimation (Supplemental Figure S1).

This scaling task was piloted online using Pavlovia (via Prolific), with the experiment implemented in PsychoPy. Thirty participants took part (mean age = 29.23 years, SD = 4.77 years; 14 females). A repeated- measures ANOVA revealed no significant effect of relevant dimension on accuracy, $F(2,58) = 1.86$, $p = .17$, indicating that participants performed similarly across the weight, speed, and price dimensions. An unstandardised profile confirmed that the minimum and maximum scale values were comparable across dimensions, suggesting similar levels of difficulty and perceptual span for the three features. Crucially, this ensured that an equidistant cube, where each dimension spans an equal range from minimum to maximum, could reliably be used in the main task.

To enable potential future analysis of grid-like coding, we restricted stimulus pairs to specific spatial trajectories. These included cardinal angles (0°, 90°, 180°, 270°), oblique angles (45°, 135°, 225°, 315°), and symmetrical directions offset from cardinal axes by 20° and 60°, yielding a total of 16 angular trajectories. While the current study did not directly investigate grid-like representations, this design enables future analysis.

Task

Prior to each phase of the experiment, participants received detailed instructions followed by training to ensure full understanding of the task requirements. The experiment comprised five phases: (1) a localiser task; (2) learning of

the three stimulus dimensions; (3) learning the contexts' reward-contingency preferences; (4) the serial reversal learning task; and (5) a post-task localiser. The primary focus of the current study was Phase 4 (the serial reversal learning task) in which participants were required to infer the current context and select actions accordingly.

Localiser task

In the localiser task, participants were presented with individual images of cars and were instructed to press the space bar on the keyboard whenever the currently displayed car was identical to the one shown on the previous trial. These repeat trials, referred to as catch trials, served as an attention check and occurred approximately every 10 trials. Each trial began with a fixation cross (500 ms), followed by the presentation of a car image (1500 ms). All 125 unique car images were shown once, and 14 of them were repeated, resulting in a total of 139 trials. The localiser phase lasted approximately 6 minutes, including instructions and excluding breaks.

Learning of different dimensions

In the second phase of the experiment, participants learned to discriminate between the different dimensions (i.e., visual features) of the car stimuli. They were informed that each dimension was a proxy for a hidden characteristic of the car (see Stimuli), but were not told how many levels each dimension contained. The goal of this phase was to implicitly learn the ordinal structure within each dimension through repeated exposure and feedback.

Before the first trial of each dimension block and then every 20th trial thereafter participants received a brief instruction screen indicating which feature to attend to (e.g., "Your task is to choose the fastest of the two cars!", which referred to the rear wing). These instructions were displayed for 2000 ms. Each trial involved a two-alternative forced choice between two cars. Participants responded using the left and right arrow keys to select the first or second car, respectively. Trials followed the following structure: a fixation cross (750 ms), the first car (1500 ms), a second fixation cross (500 ms), and the second car, which remained onscreen until a response was made or for a maximum of 2000 ms. Every 20th trial, participants received feedback on their performance in the form of accuracy and the percentage of completed trials (presented for 2000 ms).

Participants completed 100 trials for each dimension, resulting in a total of 300 trials. If a participant achieved more than 80% accuracy and completed more than 80% of the trials within a dimension, they progressed to the next dimension. This phase lasted approximately 20 minutes, including two short breaks, assuming full completion of all 300 trials. Average performance was 97.22% (SE = .32%).

Learning context preferences

In the third phase of the experiment, participants learned the car preferences of three fictional car-dealer characters, each representing a distinct context. The reward preferences were counterbalanced across participants to ensure that each dimension was relevant in isolation and in combination with another dimension across the three contexts. Specifically, for a given participant, one context exhibited preferences based on a single dimension (1D), while the other two contexts relied on combinations of two dimensions (2Da and 2Db). The relevant dimensions were partially overlapping across contexts. For example, if Context 1 rewards high speed (dimension 1), another context might reward low speed on the same dimension, resulting in anti-correlated reward contingencies. Alternatively, if the second context rewards a distinct combination like low speed and high cost (dimensions 1 and 3), this could reduce representational overlap and even approximate orthogonality in the reward structure.

Participants were informed, before the first and then every 20th trial, which dimension(s) were relevant for making correct decisions (e.g., "Your task is to choose the fastest and most expensive car!"). These instructions were presented for 2000 ms.

Each trial followed this structure: a fixation cross (750 ms), presentation of the first car (2500 ms), a second fixation cross (500 ms), and the second car, which remained onscreen until a response was made. Participants responded

using the left and right arrow keys to select the first or second car, respectively. Feedback was given after each trial for 1500 ms if the participant made the correct choice. Unbeknownst to participants, feedback was determined by the underlying stimulus–reward contingency maps specific to each context (see Fig. 1). Thus, the goal was to implicitly learn the reward structure associated with each context's preferences. After every 20th trial, participants received feedback on the percentage of the current phase completed.

In total, participants completed up to 375 trials (125 per context). However, if they reached a performance threshold of over 85% accuracy and completed at least 70% of the trials (i.e., 88 trials), the current context phase ended early, and after a short break, training resumed with a new context. This adaptive thresholding procedure limited the total duration of this phase to approximately 30–40 minutes, including two breaks. Average performance was 91.32% (SE = .51%).

Serial reversal-learning task

In the fourth and main phase of the experiment, participants were tested on the knowledge acquired during the previous phases. On each trial, participants were required to make a decision without knowing in advance which context they were currently playing. Instead, they had to infer the correct context through trial and error based on the feedback received after each decision. Unbeknownst to participants, they remained in the same context for between 5 and 8 trials before switching to another. These block lengths were chosen to maximise the number of trials and context switches while keeping the overall task duration within reasonable limits to minimise participant fatigue. Each trial began with a fixation cross of random duration (1000, 1250, or 1500 ms), followed by presentation of the first car image (2000 ms), a second fixation cross (500 ms), and then the second car, which remained onscreen until a response was made. Participants responded using the left or right arrow key to choose the first or second car, respectively. Feedback was provided for 1500 ms if the choice was correct.

Due to confusion and low behavioural performance in a separate piloting task, we added instructions when a context was about to change. Accordingly, when a context switch was about to occur, participants were shown an instruction screen reading: "There is now a high probability that there will be a shift in the character you are playing!" (German: "Es besteht nun eine hohe Wahrscheinlichkeit, dass sich der Charakter, den du spielst, vera¨ndert!"). This warning served two purposes: (i) it ensured that the first trial of a new context was behaviourally equivalent across switches (i.e., 50% chance accuracy), and (ii) it allowed us to control for dimensionality and behavioural uncertainty in our subsequent analyses (see Analysis section for details).

After approximately every 30 trials (timed to avoid disrupting context switches), participants received information about the percentage of the phase completed. In total, participants completed 351 trials, including 54 context switches. This phase lasted approximately 40 minutes, including short breaks.

Experimental setup

Data acquisition

The piloting task was implemented online using custom-written code in Python and JavaScript in PsychoPy, and tested via Pavlovia (https://pavlovia.org). Participants were recruited through the online platform Prolific (https://www.prolific.com). The main experiment was implemented using custom-written code in MATLAB (MathWorks), with PsychToolbox (Brainard, 1997) and the EyeLink Toolbox (Cornelissen et al., 2002) for stimulus presentation and eye-tracking integration.

In the main task, stimuli were presented relative to the monitor's refresh rate to ensure precise timing. During the main experiment, participants were seated in front of a monitor with a resolution of 1280 × 1024 pixels at a viewing distance of approximately 90 cm, with their heads positioned within the MEG helmet to minimise head movements. Eye movements were recorded continuously using an EyeLink 1000 Plus system (SR Research) at a sampling rate of 1000 Hz. The eye tracker was positioned 58 cm from the participants' eyes. Before each experimental phase, the eye

tracker was calibrated using the standard 9-point calibration and validation procedures provided by the EyeLink software.

Analysis

The general hypotheses and analysis plan were pre-registered prior to data collection and are available on the Open Science Framework (OSF; https://osf.io/6x42q/?view_only=4e3c9567d6a84ea2863e609677c76627). Data preprocessing was conducted using the MNE-Python toolbox (Gramfort et al., 2013) and custom-written scripts in Python. The main time-resolved neural analyses were carried out using the FieldTrip toolbox (Oostenveld et al., 2011) and custom-written MATLAB code. A total of 50 participants took part in the MEG experiment. For MEG analyses, no participants were excluded due to excessive artefacts or poor data quality, resulting in a final sample of 50 participants included in the reported analyses.

Statistical analysis

In accordance with the pre-registration (available on OSF), all main statistical tests that were mentioned in the pre-registration were one-sided and conducted at the group level. Non-parametric cluster-based permutation tests were used to assess significance when applicable.

MEG data analysis

MEG data were recorded at the Max Planck Institute for Human Cognitive and Brain Sciences (MPI CBS) in Leipzig, Germany, using an Elekta Neuromag TRIUX system (Elekta, Stockholm, Sweden) comprising 306 channels; 204 planar gradiometers and 102 magnetometers, sampled at 1000 Hz.

Eye tracking analysis

Eye-tracking data were converted from edf to MATLAB using the Edf2Mat toolbox (https://github.com/uzh/edf-converter) and downsampled to 250Hz, to conform to the same sampling rate as the MEG signal. We interpolated time points where the eye-tracking signal was not available (e.g., due to blinks) and set time points where participants looked beyond the field of view of the eye-tracker, automatically detected by the Eyelink, to NaNs. We then set each time point after each trial's reaction time to NaN. For all our eye-tracking analyses, we constructed three regions of interest (ROIs; the wheels, the windows, and the rear wing; see Supplemental Fig S2), in which we assessed gaze and fixation.

To quantify selective attention dynamics across learning, we examined the proportion of gaze time allocated to attended versus unattended feature dimensions over the course of trials within each context block. We analysed gaze behaviour separately for 1D and 2D trials, and then combined the results (Fig. 4A).

For each subject, gaze time was extracted from the late phase of the trial (2.5–4 seconds post-stimulus onset). Gaze samples were binarised within predefined regions of interest (ROIs) corresponding to the attended and unattended dimensions, based on the reward structure of each trial. To account for overlapping features across context pairs (e.g., 1 and 3 and 2 and 3), fixation within an unattended ROI was only counted when it reflected unique information not included in the attended dimensions. Similar pattern as in 3A was observed when instead using fixations within ROI (early: $M = 1.28$ dimensions, $SE = .02$; late: $M = 1.17$ dimensions, $SE = .01$; $t(49) = 7.99$, $p < .01$)

Gaze proportions were computed for each trial position (1–8) within a context block. For each trial, we calculated the total number of gaze samples falling within attended ROIs and separately for each of the two unattended dimensions. We then computed the proportion of total gaze time attributed to attended and unattended dimensions, averaging across all valid trials at each position.

This was done separately for 1D and 2D trials, and the resulting matrices of attended and unattended gaze proportions were concatenated across subjects. For visualisation, we computed the mean and standard error of the mean (SEM) across participants, and plotted the resulting time courses (Fig. 4B).

The same analysis as in 3B was conducted but now splitting the data into correct and incorrect trials, and calculate the gaze proportion within relevant dimensions (Fig. 4C).

Our main pre-registered hypothesis was to assess the entropy across trials within a context. To assess this, we counted the probability of ones and zeros (present/absent) within the ROIs and estimated entropy for correct and incorrect trials separately and for each trial within a context using the formula:

$$H = -\sum_{i, p_i > 0} p_i \log_2 p_i \tag{1}$$

where H represents entropy, p_i is the probability of gaze presence in an ROI. The summation runs over all nonzero probabilities ($p_i > 0$), as probabilities of zero do not contribute to entropy.

Trial 1-5 within a context contained enough trials for both correct and incorrect responses to reliably estimate entropy.

Preprocessing

For each participant, bad channels were annotated in a separate document, which was incorporated into the main data structure prior to preprocessing. To attenuate environmental noise and correct for head movements, we applied Signal Space Separation (SSS) using the Maxwell filter. The origin of the head position was set to auto with respect to the head coordinate frame. An internal expansion order of 11 and an external expansion order of 2 were used. Magnetic scaling was automatically determined by the software (mag scale = auto). Site-specific calibration and crosstalk correction files were applied throughout the filtering process.

Channels were automatically rejected if they exceeded specific amplitude thresholds: 400 fT for magnetometers 4000 fT/cm for gradiometers, and 150 μV for EOG channels. Channels identified as bad during preprocessing were excluded from the SSS computation to avoid introducing artefacts into the spatial filter. Head movement correction was performed using a sliding window algorithm with 200 ms windows and 10 ms steps. The HPI coil fit was constrained by an error threshold of 5 mm and a g-value of 0.98. Following movement correction, the MEG signals were virtually re-aligned to the participant's mean head position during the initial localiser task to ensure comparability of sensor-level data across all experimental phases. After applying the MAX filter, all data were visually inspected to confirm that artefacts had been sufficiently attenuated and that no signal distortions had been introduced. The cleaned data were then band-pass filtered with a high-pass cut-off of 0.1 Hz and a low-pass cut-off of 200 Hz. Line noise was attenuated using notch filters at 50, 100, and 150 Hz.

Artefact rejection was further refined using independent component analysis (ICA). To optimise the decomposition, a copy of the data was high-pass filtered at 1 Hz and downsampled to 250 Hz. ICA was performed using the *fastica* algorithm from the MNE-Python toolbox, with the maximum number of components set to 100, a maximum of 300 iterations, and a fixed random state for reproducibility. Bad channels were excluded from ICA. Components reflecting eye movements and cardiac activity were manually identified and removed. The cleaned data were then back-projected to the original signal space.

Next, the continuous data were epoched into trials. We used the *AutoReject* algorithm (Jas et al., 2017) as implemented in MNE-Python to automatically detect and reject bad trials and repair noisy channels. Each participant's preprocessed data were then visually inspected to confirm the quality of the trial rejection and channel

interpolation. On average, 327.28 trials (SEM = 2.99; range: 253–351) out of the total 351 trials were retained for the serial reversal learning phase.

Finally, the cleaned and epoched data were transferred to MATLAB and analysed using the FieldTrip toolbox for time-resolved and multivariate analyses.

Frontal midline theta

We performed time–frequency decomposition on MEG signals using complex Morlet wavelets. Preprocessed sensor-level data were epoched from −1 to 5 seconds and analysed separately per trial. We focused on frontocentral gradiometers (all gradiometers anterior of midline). For each trial, we estimated spectral power from 2 to 20 Hz in 1-Hz steps, using variable-width wavelets with a minimum of three cycles and ~500 ms window length. To reduce inter-trial and inter-frequency variability, we baseline-corrected power using a pre-stimulus window (-.5–-.1 s), expressed as relative change. Theta-band power (3–7 Hz) was squared and log-transformed and then averaged across frequency bins, and z-scored across time within trial. This yielded a time-resolved measure of FMT amplitude per trial and sensor. Next, we grouped the trials into 1) first trials of a context and 2) all other trials and performed a non-parametric cluster-based permutation (Fig. S3).

Frontal midline theta and prediction error

To examine trial-wise modulations in frontal theta activity, we processed the data as in previous section, but instead epoched trials from −1 to 5 seconds around outcome onset and analysed separately per trial. We again focused on frontocentral gradiometers (all gradiometers anterior of midline), based on prior work implicating frontal midline theta (FMT) in feedback and outcome processing (Cavanagh & Frank, 2014); estimated spectral power from 2 to 20 Hz in 1-Hz steps, using variable-width wavelets with a minimum of three cycles and ~500 ms window length; and baseline-corrected power using a pre-outcome window (2–2.4 s), expressed as relative change, in line with (Lopez-Gamundi et al., 2024). Theta-band power (3–7 Hz) was squared and log-transformed and then averaged across frequency bins, and z-scored across time within trial. This yielded a time-resolved measure of FMT amplitude per trial and sensor, suitable for regression against model-derived prediction error.

To test whether frontal theta activity varied as a function of model-derived signed prediction error (PE), we regressed single-trial frontal theta amplitude against trial-wise signed PE estimates from the Hierarchical State Inference (HSI) model. Trials were aligned to response onset, and time windows spanning from 200 ms before to 1 s after response were extracted per trial. For each participant, we performed linear regression at each sensor and time point, using signed PE as the predictor and post-response theta amplitude as the dependent variable. The resulting beta maps reflected how strongly theta power tracked the magnitude and direction of the PE. To identify significant spatiotemporal clusters, we used a cluster-based permutation test as implemented in the Fieldtrip toolbox (Fig. 5A).

To further probe this effect, we grouped trials into percentiles (0-33, 34-66, 67-100) based on signed PE (negative, neutral, positive) and averaged theta power over the significant cluster (Fig. 5B).

Representational similarity analysis

To investigate the temporal dynamics of representational compression and expansion, we constructed six representational dissimilarity matrices (RDMs): three for 1D representations and three for 2D representations. For model RDMs we calculated the Euclidean distance between items, using *pdist* in MATLAB. The RDMs were calculated per dimension, meaning that when a dimension was relevant, there was a linear increase in distance between levels of each dimension, independent of the values of the other dimensions. A high value meant a higher dissimilarity. This reflected the piloting study, showing an equal distance between dimensions. Crucially, to assess whether the correlation values reflected expansion or compression, we randomly shuffled the model RDMs before correlating it with the neural RDMs 250 times. This way, we assured that the values used for creating the model RDMs were not

influencing the resultant correlational values. For instance, in a trial where participants only had to pay attention to the weight of the car, a car with the smallest distance between wheels and the car body compared to a car with the largest distance (independent of the other two dimensions) would be maximally dissimilar in the relevant RDM. However, in the other two model RDMs, they would be maximally similar. On the other hand, a similar contrast (heaviest vs. lightest), but now also changing the values on the other two dimensions between the first and second car, for example a medium fast car that was medium expensive compared to a very fast car that was cheap, would still result in a maximally dissimilar representation in the relevant model RDM, but would in the other model RDMs result in a medium and low similarity, respectively.

The neural data were first smoothed using a 200 ms moving average (implemented via smoothdata in MATLAB). We then split the dataset based on task demands into trials where participants focused on one dimension (1D) and those where they focused on two dimensions (2D).

Prior to computing dissimilarities, we applied multivariate noise normalisation and cocktail blank removal to increase the signal-to-noise ratio, following methods established in previous multivariate work (Guggenmos et al., 2018). In line with the pre-registration, and to estimate spectral power, we convolved the time series with complex Morlet wavelets. Power was calculated using linearly spaced wavelets from 3 to 100 Hz: all integer frequencies between 3 and 30 Hz, and every fifth frequency between 40 and 100 Hz. The continuous-time wavelet transforms were squared and log-transformed. To normalise for power differences across frequency bands, we z-scored power values separately for each frequency using that frequency's mean and standard deviation across all trials. These were then averaged within and across canonical frequency bands: theta (3–7 Hz), alpha/beta (8–30 Hz), and gamma (40–100 Hz), resulting in a single signal representing the average across all frequencies.

The data were segmented into two different time windows: an early window from the onset of the first stimulus to the following fixation cross (0–2 s), and a late window from the onset of the second stimulus to the end of the trial epoch (2.5–4 s). To reduce computational load, we employed a sliding window approach with 200 ms windows and 80% overlap. Within each window, we computed the 1− Euclidean distance between neural patterns elicited by the first and second stimuli. The dissimilarities during the first stimulus period (0–2 s) were averaged, yielding a dynamic neural dissimilarity time series for the second stimulus period (2.5–4 s). At each time point, we used partial Spearman's correlation to assess the relationship between model RDMs and neural RDMs.

For each condition, we subtracted the mean of the shuffled baseline distribution from the empirical values and divided by the standard deviation, yielding a z-score at each time point. Positive z-scores indicate greater representational dissimilarity, interpreted as expansion, while negative z-scores indicate compression.

Our conditions of interest were: 1D and 2D trials collapsed, assessing the correlation between model RDMs and neural RDMs for correct and incorrect trials for the relevant dimensions (assessing whether correct trials were related to an expansion and incorrect trials related to a compression). And for 1D trials, assessing the correlation between model RDMs and neural RDMs for relevant and irrelevant dimensions for correct trials (assessing whether relevant dimensions were related to an expansion, whereas irrelevant dimensions were related to a compression). Statistical comparisons between conditions were performed using non-parametric cluster-based permutation testing, as described previously.

Cross-correlation between FMT-PE and RSA

Finally, to assess the temporal relationship between frontal midline theta modulated by prediction error (FMT-PE) and neural representational expansion, we performed a cross-correlation analysis. As the two signals differed in temporal resolution, we linearly interpolated them to a common time vector using MATLAB's *interp1* function. For each participant, we then generated 1,000 surrogate datasets to construct a null distribution. To preserve the temporal autocorrelation structure of the signals, we employed Iterative Amplitude Adjusted Fourier Transforms (IAAFT) (Schreiber & Schmitz, 1996), which randomise phase values rather than time points. The observed group-average cross-correlation was z-scored against the distribution of permuted cross-correlations at each time lag. To

correct for multiple comparisons across lags (ranging from –200 to +200 ms), we applied a two-sided cluster-based permutation test as implemented in the FieldTrip toolbox (Oostenveld et al., 2011).

Computational modelling

Model Description

We evaluated four computational models to account for participants' behaviour in the serial reversal learning task. Three of these models were variants of Q-learning, a family of reinforcement learning (RL) models based on value updating. The fourth model was a hidden state inference (HSI) model, which has recently shown promise in a similar task with rodents. The modelling objective was to predict participants' inferences about the current context and their button-press decisions.

Q-learning

We implemented a context-sensitive Q-learning model that simultaneously tracked beliefs about context identity and action values. On each trial, the model maintained context-specific Q-values and used softmax inference to estimate the current context. Let ct ∈ {1,2,3} denote the latent context on trial t. Q-values were initialised uniformly:

$$Q_0(c) = \frac{1}{3} \quad \text{for all } c \in \{1, 2, 3\}. \tag{1}$$

The probability of inferring context c was computed via a softmax decision rule:

$$P(c_t = c) = \frac{\exp\left(\beta Q_t(c)\right)}{\sum_{c'} \exp\left(\beta Q_t(c')\right)} \tag{2}$$

After observing the true context and reward on each trial, the model updated the Q-value for the true context using a reward prediction error

$$\delta_t = r_t - Q_t(c_t), \quad Q_{t+1}(c_t) = Q_t(c_t) + \alpha\, \delta_t \tag{3}$$

Q-learning with Forgetting

A second model variant extended standard Q-learning by incorporating forgetting, allowing the model to gradually reset the value of non-chosen contexts toward the average expected value. This mechanism accounts for the idea that unused associations decay over time unless reinforced. After updating the Q-value for the true context using the standard Q-learning rule:

$$Q_{t+1}(c_{\text{true}}) = Q_t(c_{\text{true}}) + \alpha\big(r_t - Q_t(c_{\text{true}})\big), \tag{4}$$

the model adjusted the Q-values of the non-chosen contexts based on their deviation from the average value Qt. Specifically, for each non-chosen context $c' \neq c_{\text{true}}$, the Q-value was updated according to

$$Q_{t+1}(c') = \begin{cases} (1-\delta)\cdot Q_t(c') & \text{if } Q_t(c') > \overline{Q_t} \text{ and } (1-\delta)\cdot Q_t(c') \geq \overline{Q_t}, \\ (1+\delta)\cdot Q_t(c') & \text{if } Q_t(c') < \overline{Q_t} \text{ and } (1+\delta)\cdot Q_t(c') \leq \overline{Q_t}, \\ \overline{Q_t} & \text{otherwise.} \end{cases} \tag{5}$$

Q-learning with Counterfactual Updating

A third model extended standard Q-learning by incorporating counterfactual updating. In this model, not only is the value of the chosen context updated based on received feedback, but the values of the unchosen contexts are also adjusted in the opposite direction. This mechanism captures the idea that selecting one context and receiving a particular outcome can inform beliefs about the unselected alternatives. For the chosen context c_{true}, the update rule depended on the received reward r_t:

$$Q_{t+1}(c_{\text{true}}) = \begin{cases} Q_t(c_{\text{true}}) + \alpha_r\big(r_t - Q_t(c_{\text{true}})\big), & \text{if } r_t = 1, \\ Q_t(c_{\text{true}}) + \alpha_{nr}\big(r_t - Q_t(c_{\text{true}})\big), & \text{if } r_t = 0. \end{cases} \tag{6}$$

For each unchosen context $c' \neq c_{true}$, the update was based on the counterfactual outcome $1 - r_t$, again with separate learning rates for rewarded and unrewarded trials:

$$Q_{t+1}(c') = \begin{cases} Q_t(c') + \alpha_r\big(0 - Q_t(c')\big), & \text{if } r_t = 1, \\ Q_t(c') + \alpha_{nr}\big(1 - Q_t(c')\big), & \text{if } r_t = 0. \end{cases} \tag{7}$$

Hidden State Inference Model

The Hidden State Inference (HSI) model adopts a Bayesian framework, in which agents infer the most likely latent state (i.e., context) of the environment on each trial. The agent maintains a belief distribution $b^t(s) = p(s_t \mid o^{t-1})$ over possible states $s \in \{1,2,3\}$, conditioned on past observations o^{t-1}. Each observation is defined as an action–reward pair:

$$o_t = \{a_t, r_t\}$$

On each trial, beliefs are updated via Bayes' rule, combining the transition model and the observation likelihood:

$$p(s_{t+1} \mid o^t) \;=\; \frac{p(o_t \mid s_t)\, p(s_t \mid o^{t-1})\, p(s_{t+1} \mid s_t)}{\sum_{s'} p(o_t \mid s'_t)\, p(s'_t \mid o^{t-1})\, p(s_{t+1} \mid s'_t)}. \tag{8}$$

State transitions follow a Markov process with persistence parameter $\gamma \in [0, 1]$. The probability of remaining in the same state is

$$\chi_s \;=\; 0.5 + 0.5\gamma,$$

and the remaining probability $1-\chi_s = 0.5-0.5\gamma$ is split evenly across the two alternative states, yielding

$$p(s_{t+1} \mid s_t) = \begin{bmatrix} 0.5+0.5\gamma & \frac{0.5-0.5\gamma}{2} & \frac{0.5-0.5\gamma}{2} \\ \frac{0.5-0.5\gamma}{2} & 0.5+0.5\gamma & \frac{0.5-0.5\gamma}{2} \\ \frac{0.5-0.5\gamma}{2} & \frac{0.5-0.5\gamma}{2} & 0.5+0.5\gamma \end{bmatrix}. \tag{9}$$

Let $a^*(s)$ denote the action that is rewarded in state s (the state's "correct" action in the task). We introduce two sensitivity parameters $c, d \in [0, 1]$ controlling how strongly the likelihood favours feedback that is consistent or inconsistent with the current state's contingency. The likelihood $p(o_t \mid s)$ depends on whether the taken action matches $a^*(s)$ and on the received reward r_t:

$$p(o_t \mid s) \;=\; \begin{cases} 0.5+0.5\,c, & \text{if } a_t = a^*(s) \text{ and } r_t = 1, \\ 0.5-0.5\,c, & \text{if } a_t = a^*(s) \text{ and } r_t = 0, \\ 0.5-0.5\,d, & \text{if } a_t \neq a^*(s) \text{ and } r_t = 1, \\ 0.5+0.5\,d, & \text{if } a_t \neq a^*(s) \text{ and } r_t = 0. \end{cases} \tag{10}$$

Action Selection Based on Context Inference

Although each model used a different mechanism to infer the current context, via value updating, decay, counterfactual reasoning, or Bayesian inference, the final step of action selection was shared across all models. On each trial, once the model computed a probability distribution over contexts $P(c_t)$ or beliefs over latent states $b_t(s)$, it used these to estimate the expected reward for each action. Specifically, expected reward values for the left and right choices were computed as the weighted sum of the rewards in each context:

$$Q_L = \sum_c P(c_t = c) \cdot r_L^c \tag{11}$$

$$Q_R = \sum_c P(c_t = c) \cdot r_R^c \tag{12}$$

Here, r_L^c and r_R^c denote the reward associated with the left and right choices, respectively, in context c (or latent state s in the case of the HSI model).

The probability of selecting each action was then determined using a separate softmax function:

$$P_L = \frac{\exp(\beta_{\text{action}} Q_L)}{\exp(\beta_{\text{action}} Q_L) + \exp(\beta_{\text{action}} Q_R)}, \quad P_R = 1 - P_L. \tag{13}$$

The model then simulated a button press based on these action probabilities. These predicted responses were used to compute trial-by-trial likelihoods and to evaluate how well each model captured participants' observed behaviour. Thus, the model was more confident in selecting the correct action when the inferred probability for the corresponding context was high, as this increased the weighting of the relevant reward contingency in the expected value computation.

Model Fitting

To estimate the parameter values that best characterise the behavioural data, we employed a likelihood maximisation approach to model fitting. For each behavioural session, we estimated the probability of individual choices based on a given model (*m*), its parameters (Θ_m), and the history of choices and outcomes within that session.

The logarithms of the choice probabilities corresponding to the participants' choices on each trial were then summed. The MATLAB function *fmincon* was used to determine the parameter values that minimised the negative log-likelihood of the data (*LL*) given the model parameters:

$$LL = \log p(d_{1:T} \mid \Theta_m, m) = \sum_{t=1}^{T} \log p(c_t \mid d_{1:t-1}, \Theta_m, m) \tag{14}$$

To prevent convergence to local minima in the minimisation process, model fitting was repeated 50 times using randomly selected initial values within predefined bounds for each parameter. The log-likelihood of the best-fitting model was recorded for each run, and the final parameter estimates were taken from the run with the highest log-likelihood value.

To identify the model that provided the most parsimonious fit to the data, we compared the model fits for each individual session using the Bayesian Information Criterion (BIC). The BIC incorporates an explicit penalty for the number of free parameters (k_m) in the model m, thereby controlling for overfitting:

$$BIC = -2 \log \hat{L} + k_m \log(N) \tag{15}$$

where *L* denotes the log-likelihood value at the best-fitting parameters, and *N* represents the number of trials within a session. To compare model fits for each participant, we calculated the difference in BIC scores (*ΔBIC*) between each model fit and the most parsimonious model for the same session.

Model verification

To assess how reliably the best-fitting model from our model comparison procedure reflects the true generative process, we performed a model recovery analysis (Wilson & Collins, 2019). This involved a comprehensive all-by-all comparison: for each model, we simulated behavioural data and then fit all models to each simulated dataset.

We quantified the proportion of datasets generated by one model that were best fit by each of the candidate models (i.e., fit model and simulated model), and summarised the results in a confusion matrix. If each model perfectly recovered itself, the confusion matrix would be an identity matrix, indicating perfect recoverability.

We also computed the corresponding inversion matrix, which reflects the probability that a given model generated the data, given that it provided the best fit (i.e., simulated model and fit model). This allows us to estimate how confidently we can infer participants' underlying strategies based on model fits and specifically, how likely it is that the best-fitting model also generated the observed behaviour.

Supplementary material

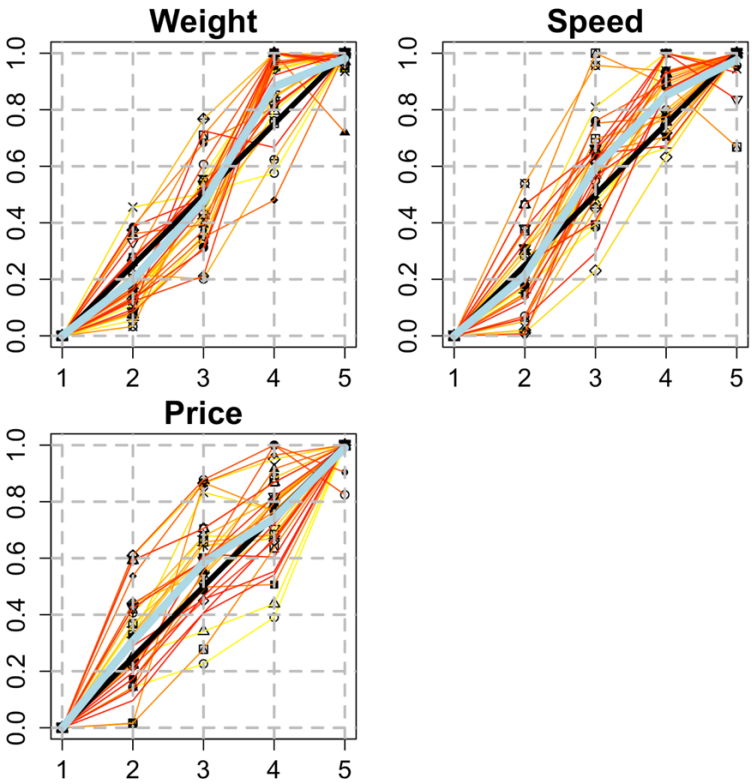

Figure S1. Maximum likelihood difference scaling of different dimensions. A pilot study comprising 30 participants assured that the psychological distance between levels within each dimension was equal. The light blue diagonal shows the average for all participants for any given dimension, whereas the black shows the optimal performance. Lines in different colours show each participant's performance. If the line is above the diagonal, it means that the psychological distance was judged larger, whereas when the line was below the diagonal, the distance was judged to be smaller.

Figure S2. Regions of interest for eye-tracking analyses and gaze heatmap of one participant. Highlighted as white circles and boxes are the regions of interest used for eye-tracking analyses. Yellow highlights heatmap for one example participant.

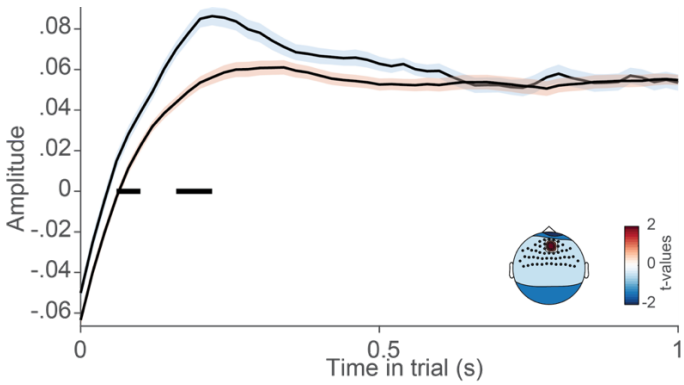

Figure S3. Frontal midline theta. When contrasting the first trial within a new context and all other trials, we observed increased theta power over midfrontal regions.

Table S1. Parameter values (Mean ± SD) from model fitting for each model.

Model	Alpha	Alpha nr	Beta	Delta	Beta-action	C	Gamma
QN	.62033 ± .1402	N/A	2.6417 ± .70688	N/A	14.5887 ± 11.0483	N/A	N/A
QF	.27291 ±.081964	N/A	8.9654 ± 1.3827	0.60101 ± 0.27275	17.5727 ± 9.8558	N/A	N/A
QC	.87645 ± .1768	0.085772 ± 0.11839	2.3789 ± .57669	N/A	12.3439 ± 9.5323	N/A	N/A
HSI	N/A	N/A	.48681 ± .10367	N/A	7.3397 ± 2.067	.83811 ± .1944	.28805 ± .2532

References

Bar-Gad, I., Morris, G., & Bergman, H. (2003). Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. *Progress in Neurobiology*, *71*(6), 439–473. https://doi.org/10.1016/j.pneurobio.2003.12.001

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

Bellman, R. (1957). *Dynamic Programming*. Princeton University Press.

Bellmund, J. L. S., Gärdenfors, P., Moser, E. I., & Doeller, C. F. (2018). Navigating cognition: Spatial codes for human thinking. *Science*, *362*(6415), eaat6766. https://doi.org/10.1126/science.aat6766

Blair, M. R., Watson, M. R., Walshe, R. C., & Maj, F. (2009). Extremely selective attention: Eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1196–1206. https://doi.org/10.1037/a0016272

Boyen, X., Friedman, N., & Koller, D. (2013). *Discovering the Hidden Structure of Complex Dynamic Systems* (arXiv:1301.6683). arXiv. https://doi.org/10.48550/arXiv.1301.6683

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357

Braun, D. A., Mehring, C., & Wolpert, D. M. (2010). Structure learning in action. *Behavioural Brain Research*, *206*(2), 157–165. https://doi.org/10.1016/j.bbr.2009.08.031

Braunlich, K., & Love, B. C. (2022). Bidirectional influences of information sampling and concept learning. *Psychological Review*, *129*(2), 213–234. https://doi.org/10.1037/rev0000287

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421. https://doi.org/10.1016/j.tics.2014.04.012

Cavanagh, J. F., & Shackman, A. J. (2015). Frontal midline theta reflects anxiety and cognitive control: Meta-analytic evidence. *Journal of Physiology, Paris*, *109*(1–3), 3–15. https://doi.org/10.1016/j.jphysparis.2014.04.003

Cockburn, J., & Holroyd, C. B. (2018). Feedback information and the reward positivity. *International Journal of Psychophysiology*, *132*, 243–251. https://doi.org/10.1016/j.ijpsycho.2017.11.017

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), Article 3. https://doi.org/10.1038/nrn755

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, *34*(4), 613–617. https://doi.org/10.3758/BF03195489

Costa, V. D., Tran, V. L., Turchi, J., & Averbeck, B. B. (2015). Reversal learning and dopamine: A bayesian perspective. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(6), 2407–2416. https://doi.org/10.1523/JNEUROSCI.1989-14.2015

Courellis, H. S., Minxha, J., Cardenas, A. R., Kimmel, D. L., Reed, C. M., Valiante, T. A., Salzman, C. D., Mamelak, A. N., Fusi, S., & Rutishauser, U. (2024). Abstract representations emerge in human hippocampal neurons during inference. *Nature*, 1–9. https://doi.org/10.1038/s41586-024-07799-x

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-Based Influences on Humans' Choices and Striatal Prediction Errors. *Neuron*, *69*(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

Duan, Y., Zhan, J., Gross, J., Ince, R. A. A., & Schyns, P. G. (2024). Pre-frontal cortex guides dimension-reducing transformations in the occipito-ventral pathway for categorization behaviors. *Current Biology*, *34*(15), 3392-3404.e5. https://doi.org/10.1016/j.cub.2024.06.050

Edgell, S. E., & Morrissey, J. M. (1987). Delayed exposure to additional relevant information in nonmetric multiple-cue probability learning. *Organizational Behavior and Human Decision Processes*, *40*(1), 22–38. https://doi.org/10.1016/0749-5978(87)90003-3

Eichenbaum, H., & Cohen, N. J. (2014). Can We Reconcile the Declarative Memory and Spatial Navigation Views on Hippocampal Function? *Neuron*, *83*(4), 764–770. https://doi.org/10.1016/j.neuron.2014.07.032

Flesch, T., Juechems, K., Dumbalska, T., Saxe, A., & Summerfield, C. (2022). Orthogonal representations for robust context-dependent task performance in brains and neural networks. *Neuron*, *110*(7), 1258-1270.e11. https://doi.org/10.1016/j.neuron.2022.01.005

Frank, M. J., & Badre, D. (2012). Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: Computational analysis. *Cerebral Cortex (New York, N.Y.: 1991)*, *22*(3), 509–526. https://doi.org/10.1093/cercor/bhr114

Gärdenfors, P. (2000). *Conceptual Spaces: The Geometry of Thought*. The MIT Press. https://doi.org/10.7551/mitpress/2076.001.0001

Garvert, M. M., Dolan, R. J., & Behrens, T. E. (2017). A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *ELife*, *6*, e17086. https://doi.org/10.7554/eLife.17086

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: Carving nature at its joints. *Current Opinion in Neurobiology*, *20*(2), 251–256. https://doi.org/10.1016/j.conb.2010.02.008

Heydari, S., & Holroyd, C. B. (2016). Reward positivity: Reward prediction error or salience prediction error? *Psychophysiology*, *53*(8), 1185–1192. https://doi.org/10.1111/psyp.12673

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, *159*, 417–429. https://doi.org/10.1016/j.neuroimage.2017.06.030

Kato, A., & Morita, K. (2016). Forgetting in Reinforcement Learning Links Sustained Dopamine Signals to Motivation. *PLoS Computational Biology*, *12*(10), e1005145. https://doi.org/10.1371/journal.pcbi.1005145

Knudsen, E. B., & Wallis, J. D. (2021). Hippocampal neurons construct a map of an abstract value space. *Cell*, *184*(18), 4640-4650.e10. https://doi.org/10.1016/j.cell.2021.07.010

Konidaris, G. (2019). On the necessity of abstraction. *Current Opinion in Behavioral Sciences*, *29*, 1–7. https://doi.org/10.1016/j.cobeha.2018.11.005

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, *35*(13), 1897–1916. https://doi.org/10.1016/0042-6989(94)00279-u

Kruschke, J. K. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 3–26. https://doi.org/10.1037/0278-7393.22.1.3

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. *Neuron*, *93*(2), 451–463. https://doi.org/10.1016/j.neuron.2016.12.040

Lopez-Gamundi, P., Mas-Herrero, E., & Marco-Pallares, J. (2024). Disentangling effort from probability of success: Temporal dynamics of frontal midline theta in effort-based reward processing. *Cortex*, *176*, 94–112. https://doi.org/10.1016/j.cortex.2024.03.014

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A Network Model of Category Learning. *Psychological Review*, *111*(2), 309–332. https://doi.org/10.1037/0033-295X.111.2.309

Mack, M. L., Preston, A. R., & Love, B. C. (2020). Ventromedial prefrontal cortex compression during concept learning. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-019-13930-8

Menghi, N., Kacar, K., & Penny, W. (2021). Multitask learning over shared subspaces. *PLOS Computational Biology*, *17*(7), e1009092. https://doi.org/10.1371/journal.pcbi.1009092

Menghi, N., Silvestrin, F., Pascolini, L., & Penny, W. (2023). The emergence of task-relevant representations in a nonlinear decision-making task. *Neurobiology of Learning and Memory*, *206*, 107860. https://doi.org/10.1016/j.nlm.2023.107860

Mishchanchuk, K., Gregoriou, G., Qü, A., Kastler, A., Huys, Q. J. M., Wilbrecht, L., & MacAskill, A. F. (2024). Hidden state inference requires abstract contextual representations in the ventral hippocampus. *Science*, *386*(6724), 926–932. https://doi.org/10.1126/science.adq5874

Moore, T., & Fallah, M. (2001). Control of eye movements and spatial attention. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(3), 1273–1276. https://doi.org/10.1073/pnas.98.3.1273

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387–391. https://doi.org/10.1038/nature03390

Nelson, J. D., & Cottrell, G. W. (2007). A probabilistic model of eye movements in concept formation. *Neurocomputing*, *70*(13), 2256–2272. https://doi.org/10.1016/j.neucom.2006.02.026

Niv, Y. (2019). Learning task-state representations. *Nature Neuroscience*, *22*(10), 1544–1553. https://doi.org/10.1038/s41593-019-0470-8

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement Learning in Multidimensional Environments Relies on Attention Mechanisms. *The Journal of Neuroscience*, *35*(21), 8145–8157. https://doi.org/10.1523/JNEUROSCI.2978-14.2015

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. https://doi.org/10.1155/2011/156869

Palminteri, S., Khamassi, M., Joffily, M., & Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning. *Nature Communications*, *6*(1), 8096. https://doi.org/10.1038/ncomms9096

Park, S. A., Miller, D. S., Nili, H., Ranganath, C., & Boorman, E. D. (2020). Map Making: Constructing, Combining, and Inferring on Abstract Cognitive Maps. *Neuron*, *107*(6), 1226-1238.e8. https://doi.org/10.1016/j.neuron.2020.06.030

Park, S. A., Zolfaghar, M., Russin, J., Miller, D. S., O'Reilly, R. C., & Boorman, E. D. (2023). *The representational geometry of cognitive maps under dynamic cognitive control* (p. 2023.02.04.527142). bioRxiv. https://doi.org/10.1101/2023.02.04.527142

Radulescu, A., Niv, Y., & Ballard, I. (2019). Holistic Reinforcement Learning: The Role of Structure and Attention. *Trends in Cognitive Sciences*, *23*(4), 278–292. https://doi.org/10.1016/j.tics.2019.01.010

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*(1), 1–41. https://doi.org/10.1016/j.cogpsych.2004.11.001

Schreiber, T., & Schmitz, A. (1996). Improved Surrogate Data for Nonlinearity Tests. *Physical Review Letters*, *77*(4), 635–638. https://doi.org/10.1103/PhysRevLett.77.635

Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron*, *91*(6), 1402–1412. https://doi.org/10.1016/j.neuron.2016.08.019

Shiferaw, B., Downey, L., & Crewther, D. (2019). A review of gaze entropy as a measure of visual scanning efficiency. *Neuroscience and Biobehavioral Reviews*, *96*, 353–366. https://doi.org/10.1016/j.neubiorev.2018.12.007

Starkweather, C. K., Gershman, S. J., & Uchida, N. (2018). The Medial Prefrontal Cortex Shapes Dopamine Reward Prediction Errors under State Uncertainty. *Neuron*, *98*(3), 616-629.e6. https://doi.org/10.1016/j.neuron.2018.03.036

Sutton, R. S., & Barto, A. G. (n.d.). *Reinforcement Learning: An Introduction*. 352.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction, 2nd ed* (pp. xxii, 526). The MIT Press.

Theves, S., Fernández, G., & Doeller, C. F. (2020). The Hippocampus Maps Concept Space, Not Feature Space. *The Journal of Neuroscience*, *40*(38), 7318–7325. https://doi.org/10.1523/JNEUROSCI.0494-20.2020

Vertechi, P., Lottem, E., Sarra, D., Godinho, B., Treves, I., Quendera, T., Oude Lohuis, M. N., & Mainen, Z. F. (2020). Inference-Based Decisions in a Hidden State Foraging Task: Differential Contributions of Prefrontal Cortical Areas. *Neuron*, *106*(1), 166-176.e6. https://doi.org/10.1016/j.neuron.2020.01.017

Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning*, *8*(3), 279–292. https://doi.org/10.1007/BF00992698

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *ELife*, *8*, e49547. https://doi.org/10.7554/eLife.49547

Wilson, R. C., & Niv, Y. (2012). Inferring Relevance in a Changing World. *Frontiers in Human Neuroscience*, *5*, 189. https://doi.org/10.3389/fnhum.2011.00189

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, *81*(2), 267–279. https://doi.org/10.1016/j.neuron.2013.11.005

Yang, S. C.-H., Lengyel, M., & Wolpert, D. M. (2016). Active sensing in the categorization of visual patterns. *ELife*, *5*, e12215. https://doi.org/10.7554/eLife.12215

Zhang, M., & Yu, Q. (2024). The representation of abstract goals in working memory is supported by task-congruent neural geometry. *PLoS Biology*, *22*(12), e3002461. https://doi.org/10.1371/journal.pbio.3002461

Zhang, X.-Y., Bobadilla-Suarez, S., Luo, X., Lemonari, M., Brincat, S. L., Siegel, M., Miller, E. K., & Love, B. C. (2023). *Adaptive stretching of representations across brain regions and deep learning model layers* [Preprint]. Neuroscience. https://doi.org/10.1101/2023.12.01.569615