Revealing human sensitivity to a latent temporal structure of changes

Dimitrije Marković, Andrea M.F. Reiter, Stefan J. Kiebel

Abstract

Precisely timed behaviour and accurate time perception plays a critical role in our everyday lives, as our well-being and even survival can depend on well-timed decisions. Although the temporal structure of the world around us is essential for human decision making, we know surprisingly little about how representation of temporal structure of our everyday environment impacts decision making. How does the representation of temporal structure affect our ability to generate well-timed decisions? Here we address this question by using a well-established dynamic probabilistic learning task. Using computational modelling, we found that human subjects’ beliefs about temporal structure are reflected in their choices to either exploit their current knowledge or to explore novel options. The model-based analysis illustrates a large within-group and within-subject heterogeneity. To explain these results, we propose a normative model for how temporal structure is used in decision making, based on the semi-Markov formalism in an active inference framework. We discuss potential key applications of the presented approach to the fields of cognitive phenotyping and computational psychiatry.

Introduction

The passage of time is a fundamental aspect of human experience. Our behaviour is tightly coupled to our estimate of the elapsed time and the expectations about the time remaining to fulfil short or long-term goals. We are highly sensitive to the temporal structure of our everyday environment and capable of forming precise beliefs about the duration of various events (e.g. a theatre play, traffic lights, waiting in a queue). In practice, temporal structure is typically latent (e.g. not reflected in external clocks) and we seem to rely on an internalised timing mechanism, such as various implicit clocking mechanisms (Buhusi and Meck, 2005). This enables us to provide temporal context and an order to events, and to form beliefs about the underlying temporal structure (Eichenbaum, 2014).
It has been proposed that these temporal beliefs are used to make predictions and to adapt our behaviour successfully to ever-changing conditions (Griffiths and Tenenbaum, 2011). Therefore, understanding how we learn and represent the temporal structure of our everyday environment (Kiebel et al., 2008) and use these representations for making decisions (Marković et al., 2019) is essential for understanding human adaptive behaviour (Purcell and Kiani, 2016).

Neuronal and behavioural mechanisms of time perception have been studied in humans and animals, traditionally using interval timing tasks (Eagleman, 2008; Meck, 1996). The key insights of these experiments are that humans and animals integrate the experience of between event duration, in a given context, to form beliefs about possible future duration they might experience. They use these beliefs when estimating or reproducing a newly experienced interval (Jazayeri and Shadlen, 2010); in line with a Bayesian account of decision-making (Shi et al., 2013). However, it is still an open question how we integrate time perception and beliefs about durations into everyday decision making. Recently, distinct but interlinked research fields have illustrated the importance of temporal representations for cognition and decision making in sequential and dynamic tasks (Nobre and Van Ede, 2018; McGuire and Kable, 2012; Vilà-Balló et al., 2017; Eichenbaum, 2014). The sequential neuronal activity in the hippocampus has been suggested to represent elapsed time (Eichenbaum, 2017; Buzsáki and Llinás, 2017; Friston and Buzsáki, 2016), which have led to the postulate of time cells in the hippocampus (Eichenbaum, 2014; MacDonald et al., 2014; Itskov et al., 2011) critical for memory and decision-making. For example, in the research on temporal aspects of attention it has been demonstrated that temporal expectations guide allocation of attentional resources in time (Nobre and Van Ede, 2018). Similarly, inter-temporal choices or one’s willingness to wait for higher reward is strongly influenced by temporal expectations (McGuire and Kable, 2012).

Motivated by the reach literature on temporal representations in the brain, here we focus on the question of how humans form complex temporal representation of their environment. We test how such temporal representations support decisions about whether to explore or to exploit in anticipation of a change in the environment. We introduce a novel computational model of behaviour that describes learning of a latent temporal structure of a dynamic task environment in the context of sequential decision making. The computational model is applicable to any task that can be cast as a dynamic multi-armed bandit problem (Gupta et al., 2011) with semi-Markovian changes or switches in the underlying latent states (Janssen and Limnios, 2013). Here we specifically apply the model to describe learning in a sequential (probabilistic) reversal learning task (Costa et al., 2015; Reiter et al., 2016, 2017; Viñà-Balló et al., 2017). We do so by manipulating temporal contexts in this task: Subjects encountered semi-regular intervals between contingency reversals in one environment. Their behaviour was contrasted with behaviour in another environment where intervals between contingency reversals were irregular.

The proposed behavioural model was based on three components: (i) a set of templates representing possible latent temporal structure of reversals using an implicit representation of between reversal duration (Yu, 2015), (ii) the update of beliefs about states and temporal templates derived via approximate inference (Parr et al., 2019; Yu and Kobayashi, 2003), and (iii) the action selection, that is the planning process, cast as active inference (Friston et al., 2017; Markovic et al., 2021). Together these components allow us to define an efficient and approximate active learning and choice algorithm of latent
temporal structures based on variational inference Blei et al. (2017). Here we extend on our previous investigation of human behaviour in temporally structured dynamic environments (Marković et al., 2019). In this work, we demonstrated that a computational model which infers a between-event duration, can be used to reveal subjects’ beliefs about the latent temporal structure in a dynamic learning task. However, a question that has remained open is how humans acquire temporal structure in the first place. Understanding the learning of temporal structure is critical for revealing between-individual variability in temporal expectations and capturing the evolution of temporal representations within individuals. Critically, with the extended model we present here, we are indeed able to capture the learning of temporal representation and address the non-stationarity of subjects’ temporal representation during the course of the experiment.

Our aim is to address the following questions: (i) Are subjects a priori biased towards expecting regular or irregular temporal structure? (ii) Are subjects able to learn latent temporal structure without explicit instructions? (iii) How does the quality of temporal representation impact their performance? Using simulations we can illustrate the interaction of accurate representation of temporal structure and behaviour, mainly performance on the task and the engagement with exploratory behaviour. Using model-based analysis we demonstrate high diversity between subjects both in their prior beliefs about temporal structure, and their ability to adapt their beliefs to different latent temporal structure. Critically, we link the quality of temporal representation to subjects’ performance both in terms of group-level performance and within-subject variability of their performance during the task.

In what follows we will first briefly describe the experimental task, provide the overall summary of behavioural characteristics, introduce the behavioural model, and finally show results of the model-based analysis of behaviour. The formal details of the approach are described in Methods and Materials.
Results

A typical probabilistic reversal learning task asks subjects to make a binary choice between two options, e.g. A and B, where each option is associated with a probability of receiving a reward or punishment. For example, initially choosing A returns a reward with a high probability $p_H = 0.8$ and choosing B returns a reward with low probability $p_L = 0.2$. Importantly, after several trials the reward contingencies reverse, i.e. switch, such that choosing B returns the reward with high probability $p_H$. However, subjects are not informed about the reversal and they have to infer that a change occurred from the feedback they receive in order to adapt their behaviour. From the point of view of participants, a reversal can be difficult to detect as outcomes are probabilistic. This means that if someone observes a loss after a sequence of gains, e.g. when choosing the option A, this could be caused either by: (i) a true reversal, where now option B is rewarded with the probability $p_H$ or (ii) by an unlucky outcome of an otherwise correct choice. To obtain a more direct information about the subjective uncertainty of participants about the correct choice (i.e. choosing the option with high reward probability, $p_H$) on any given trial, we extended the standard design with an additional third exploratory option. This new option does not result in monetary gain or loss but provides information about the correct choice on a current trial. A high uncertainty about the best choice (current context) can be easily resolved by selecting the epistemic option. We will label all choices of the exploratory options as exploratory, and all other choices as exploitative (note that the outcomes of exploitative options also provide some information about the current context). A trial sequence of the experimental task is shown in Figure 1.

To investigate subjects’ ability to learn latent temporal structure we defined two ex-
peripheral conditions (manipulated in a between-subject design), one with irregular reversals and another with regular reversals (see Figure 2). In the condition with irregular reversals, the moments of reversals are not predictable and between-reversal intervals are drawn from a geometric distribution (Figure 2A). In the condition with regular reversals, the moments of reversal are predictable, and they occur at semi-regular intervals, drawn from a negative binomial distribution (Figure 2B). Subject were randomly assigned to one of the two possible conditions, as illustrated in Figure 2. In the first condition, subjects experience irregular reversal statistics for 800 trials, after which the reversals occur at semi regular intervals for the last 200. In the second condition, subjects experience semi-regular reversal statistics for 800 trials, and then the irregular reversal statistics during the last 200 trials. Note that when changing the temporal statistics we copied the time series of reversals from the initial 200 trials of the different condition.

Importantly, the moments of reversal were fixed within each group (depending on the condition reversals occur always on the same trials), and the same choice by different subjects on any given trial would lead to the same outcome (the outcome statistics were generated only once for each condition and trial, and then replayed to all subjects depending on their choices and the condition they were assigned to).

![Figure 2. Time series of reward probabilities.](image)

**Figure 2. Time series of reward probabilities.** A, Condition with irregular reversals, and B, condition with (semi-)regular reversals. The reward probability of the high-probability stimuli at any time step was set to \( p_H = 0.8 \) and the low-probability stimuli to \( p_L = 0.2 \). Dashed vertical lines shows the moment of change of the latent temporal structure: (i) from irregular to semi-regular statistics in the irregular condition, and (ii) from the semi-regular to irregular statistics in the regular condition. Figures on the right illustrate the generative distribution of the between-reversal intervals \( d \) for each condition. Note that the mean between reversal duration \( \langle d \rangle \) is identical in both conditions.

**Analysis of choice data**

We will first describe the behavioural characteristics of the two groups of subjects exposed to the two different experimental conditions. The two behavioural measures of
interests here are the performance (odds of being correct, i.e. odds of choosing the option with the higher reward probability) and probing (odds of exploring, i.e. odds of choosing the exploratory option). We describe all the behavioural measures in detail in Behavioural measures subsection of Methods and Materials.

Subjects (N = 74) were pseudo-randomly assigned to one of the two experimental conditions, where \( n_r = 41 \) participants were assigned to the condition with regular reversals, and \( n_i = 33 \) to the condition with irregular reversals. Note that some subjects rarely engaged with exploratory option. Out of 50 subjects who where exposed to the variant of the experiment with exploratory option (24 subjects performed a standard version of the task without exploratory option, see Experiment for more details), 5 subjects never engaged with the exploratory option. In Figure 3 we provide a summary of average behavioural measures for individual subjects. We do not find any significant performance differences between the two regularity conditions, see Figure 3A. However, for the subset of subjects which interacted with the exploratory option (45 subjects) we find that the performance is positively correlated with probing (Pearson correlation coefficient \( r = 0.6 \), with \( p < 10^{-4} \)), see Figure 3B. Interestingly, neither of the two behavioural measures (when plotted as a within subject average over the course of experiment), reveals obvious between-condition differences. However, when comparing the temporal profile of these measures over the course of experiment (see Figure 9–Figure Supplement 1), one notices large variability both between subjects but also within a subject over the course of experiment; suggesting ongoing learning of the task structure. In what follows we will classify the heterogeneity of behavioural responses using a model-based analysis.

**Figure 3. Averages of behavioural summary measures.** A, Distribution of the mean performance of subjects with low and high number of exploratory choices, see methods. B, Dependence of the mean performance on the mean probing, where we excluded participants without exploratory choices (count = 0). Note that when computing mean performance and mean probing for each participants, we have excluded the first 400 (initial responses during which the subjects might have still been adjusting to the task) and last 200 (responses after the change in the reversal statistics) responses of each participant, see Methods and Materials – Model inversion for the motivation for the cutoff.

**Behavioural model**

The behavioural model will allow us to investigate the process of learning of the latent temporal structure in different experimental conditions, reveal subjects’ preferences to
engage with exploratory option (collect information), and subjects’ motivation to collect rewards. We achieve this by fitting free model parameters to behavioural responses of each subject (see Methods and Materials for more details). Our aim with the model based analysis is to quantify beliefs about temporal structure of reversals and understand how the belief dynamics influences subjects’ behaviour.

We conceptualised the behavioural model as an active inference agent (Friston et al., 2015, 2016) with hidden semi-Markov models (Yu, 2010), which is capable of representing and inferring latent temporal structure. In active inference, besides defining perception and learning as a Bayesian inference process, action selection is also cast as an inference problem aimed at minimising the expected surprise about future outcomes, that is, the expected free energy (Smith et al., 2022) (see also Eq (18)). Through its dependence on the expected free energy, the action selection has an implicit dual imperative (see possible factorisation of the expected free energy in Eq (18)): The expected free energy combines intrinsic and extrinsic value of a choice, where intrinsic value corresponds to the expected information gain, and the extrinsic value to the expected reward of different choices. The implicit information gain or uncertainty reduction pertains to beliefs about the task’s dynamical structure and choice-outcome mappings, e.g. (Kaplan and Friston, 2018; Schwartenbeck et al., 2013). Therefore, selecting actions that minimise the expected free energy dissolves the exploration-exploitation trade-off, as every action is driven both by expected value and the expected information gain. This is a critical feature of active inference models which allows us to account for exploratory choices (see Figure 1).

We express the agent’s generative model of task dynamics in terms of hidden semi-Markov models (HSMM) (Yu, 2015; Marković et al., 2019). The HSMM framework extends a standard hidden Markov model with an implicit (or explicit) representation of durations between consecutive state changes. HSMM have found numerous applications in the analysis of non-stationary time series in machine learning (Duong et al., 2005; Gales and Young, 2008), and in neuroimaging (Borst and Anderson, 2015; Shappell et al., 2019). HSMM have also been used in decision making for temporal structuring of behavioural policies (Bradtke and Duff, 1994) or in temporal difference learning as a model of dopamine activity when the timing between action and reward is varied between experimental trials (Daw et al., 2002).

Here, we use the semi-Markov representation of task dynamics within the behavioural models to define an agent that can learn latent temporal structure, form beliefs about moments of change, and anticipate state changes. We implemented the learning of the hidden temporal structure of reversals as a variational inference scheme, where we assume that the agent entertains a hierarchical representation of the reversal learning task, with a finite set of models of possible temporal structure of the dynamic environment. In other words, we assume that human brain entertains a set (possibly a very large set) of temporal templates. In Figure 4, we show the graphical representation of the generative model of behaviour, which is described in detail in Behavioural model section. Here we will briefly introduce the relevant parametrisation of the behavioural model, which are critical for understanding the model comparison results presented in the next subsection.

Each temporal template $m$ corresponds to a pair of parameters $m = (\mu, \nu)$ that define the frequency of reversals $\mu$ and the regularity of reversals $\nu$ (the higher the value the
Figure 4. Graphical representation of the generative model and model summary. At the top of the hierarchy is the temporal template variable \( m \). The total number of temporal templates is finite, e.g., \( m \in \{1, \ldots, m_{\text{max}}\} \), and each template \( m \) provides an implicit representation of a prior probability distribution over between-reversal intervals \( d \), parameterised with a pair \( m = (\mu, \nu) \), where \( \mu \) expresses mean between reversal interval, and \( \nu \) plays a role of a precision parameter, that defines regularity of between-reversal intervals. The implicit representation of temporal structure is encoded with probability transition matrices \( L_m \) latent phases \( f \). The number of latent phases depends on precision \( \nu \). The reversal can occur only when end phase is reached \( (f = f + 1) \). Therefore, the phase variable \( f \) controls the transitions probability \( B_{a,f} \) between latent states (the correct choice) of the task denoted as random variable \( s^2 \in \{1, 2\} \). At every trial \( t \) the subject makes a choice \( a_t \), hence decides on an option \( s^1 \) which will generate an outcome \( o_t \). The choices are deterministic, hence \( p(s^1_t | s^2_{t-1}, a_t) = p(s^1_t | a_t) = B_{a} = I_1 \). Finally, the choice-outcome contingencies are treated as latent variables \( R_{s, o} \) which have to be learned over the course of the experiment. We use a vague Dirichlet prior over choice-outcome contingencies.

Inverting the generative model of outcomes using variational inference defines the inference and learning component of the behavioural model. In turn, marginal beliefs about latent states \( s^1_t, s^2_t \) and parameters \( R_{s, o} \) are used to define action selection, that is compute the choice likelihoods using the expected free energy (Eq (18)).

more regular the changes are). In Figure 12 we illustrate three of these templates, which differ in their regularity parameter \( \nu \), but all have the same frequency parameter \( \mu \). It is important to note that when \( \nu = 1 \) (the lowest value) the temporal templates correspond to the hidden Markov model (HMM) representation. HMM representation implies that the moments of reversals are unpredictable, or maximally irregular. Here we use the HMM representation as a reference point for determining whether participants were able to learn latent temporal structure of reversals, and whether they a priori expected
predictable moments of reversal.

When simulating behaviour and fitting the model to participants’ choices, we use a prior probability \( p(m) \) over temporal templates \( m \) to restrict otherwise rich set of all possible temporal templates \( m = (\mu, \nu) \), that span all combinations of \( \mu \in \{5, \ldots, 45\} \) and \( \nu \in \{1, \ldots, 10\} \). Hence, template prior \( p(m) \) reflects prior expectations of an agent at the beginning of the experiment about the possible temporal structure of the task dynamics. Therefore, to capture a wide range of prior beliefs we require a flexible prior \( p(m) \) that can reflect subjects with different prior expectations about temporal structure. Posterior estimates of the most likely parametrisations of the temporal prior, allows us to infer from the behavioural data if participants’ beliefs are a priori precise and biased towards expecting irregular reversals, or are imprecise and accommodate a wide range of possible latent temporal structures. In the model, we use the following prior over temporal templates:

\[
p(m|_{\max}) = p(\mu, \nu|_{\max}) = p(\mu)p(\nu|_{\max})
\]

\[
p(\mu) = \frac{1}{40}
\]

\[
p(\nu|_{\max}) = \begin{cases} \frac{1}{v_{\max}} & \text{for } 1 \leq \nu \leq v_{\max} \\ 0 & \text{otherwise} \end{cases}
\]

where \( v_{\max} \in \{1, \ldots, 10\} \). Note that the prior regularity parameter \( v_{\max} \) reflects Bayesian prior expectations about the maximal precision of between-reversal intervals. In other words, \( v_{\max} \) captures the agent’s expectations about the maximal regularity of reversals, and hence their predictability. Thus, with this parametrisation we assume that subjects, at the beginning of the experiment, have uniform beliefs about a possible mean duration between reversal interval, but might differ in their propensity to represent high or low regularity of between-reversal intervals. For example, some subjects could hold precise beliefs that reversals were not under their control and were therefore inherently unpredictable (corresponding to \( v_{\max} = 1 \)). Such a subject would fail to learn – or accumulate evidence for – the regularity of reversals in the regular condition. Conversely, some participants may have imprecise prior beliefs about regularity (\( v_{\max} > 1 \)); enabling them to learn that reversals were regular, thus predictable, in the appropriate condition.

The beliefs about temporal templates, influence the beliefs about the reversal probability on any given trial (i.e., how likely is that a reversal occurs in the next trial), and consequently modulate beliefs about the latent state of the task (i.e., which card is associated with high reward probability) and corresponding outcome probabilities. In turn, the beliefs about the latent state influence the choices. As mentioned above, choices are defined as the minimisers of the expected free energy (surprise about future outcomes), typically denoted by \( G \). Given the expected free energy \( G_{a} [P_{o}, v_{\max}, t] \) of action \( a \) on trial \( t \) we define the choice likelihood as

\[
a_{t} \sim p(a) \propto e^{\gamma G_{a}[P_{o}, v_{\max}, t]}.
\]
In active inference (Friston et al., 2017) prior preference parameter $P_o$ defines motivation to collect rewards (generate correct choices) and collect information (engage with the exploratory option). In more familiar terminology of reinforcement learning, the logarithm of prior preferences $\ln P_o$ assigns a subjective value to possible outcomes, and the expectation of log-preferences defines the expected value of different actions (see Eq (18)). Importantly, Eq (2) acts in two different ways: (i) as a mapping from beliefs into actions which we used to simulate behavioural choices, and (ii) as a choice likelihood which we use for inverting the model when fitting the model to subjects’ choices to derive the posterior estimates of free model parameters $(\gamma, p_\gamma, p_\lambda, \gamma_{\text{max}})$, individually for each subject. Details of the model inversion procedure are described in Model inversion.

Simulating the behavioural effect of prior expectations over temporal templates

By simulating the model’s behaviour given different values of temporal regularity parameter $\gamma_{\text{max}}$, we aimed to demonstrate that the agent can acquire a correct representation of the latent temporal structure in different experimental conditions, and that $\gamma_{\text{max}}$ influences the dynamics of both performance and probing. Importantly, different values of $\gamma_{\text{max}}$ should lead to sufficiently distinct behaviour, if we hope to accurately associate subjects’ behaviour with underlying model parametrisation.

![Figure 5. Model dependent dynamics of behavioural measures for varying $\gamma_{\text{max}}$.](image)

Each line corresponds to an average over $n = 50$ simulated trajectories with $\gamma = 5$, and $P_o = (0.1, 0.6, 0.15, 0.15)$. **A**, Performance estimated as odds of generating a correct choice within a 200 trials long time window centred at trial index. **B**, probing, computed as odds of selecting the exploratory option within the 200 trials long time window.

The temporal regularity parameter $\gamma_{\text{max}}$ is the key parameter in the model to understand how learning about temporal structure comes about. As $\gamma_{\text{max}}$ constrains the maximal temporal regularity the agent expects in the task, it is a measure of subjects’ sensitivity to the latent temporal structure. Importantly, we find that varying $\gamma_{\text{max}}$ results in simulated behaviour with distinct behavioural patterns in our two experimental conditions as shown in Figure 5. As we increase $\gamma_{\text{max}}$ the behavioural performance increases,
in both conditions. In contrast, as we increase $v_{\text{max}}$ the probing decreases, as the agent is more certain about the moment of reversal and requires information provided by exploratory option less often. Note that different values of $v_{\text{max}}$ induce stronger differences in both performance and probing in the regular condition, compared to the irregular condition. Practically, this means that we can infer $v_{\text{max}}$ from behavioural data with higher precision in regular than in irregular condition. We validate the classification accuracy of $v_{\text{max}}$ based on posterior estimates given simulated data in the form of confusion matrix as shown in Figure 8–Figure Supplement 1. Note that even in the ideal case when behaviour is generated exactly from the behavioural model, classification accuracy with regard to $v_{\text{max}}$ is substantially lower in irregular compared to irregular condition. We will clarify the impact of this in the next subsection when discussing the the results of model-based analysis.

Demonstrating the learnability of latent temporal structure

As a next step we will illustrate that the agent with the highest value of temporal prior ($v_{\text{max}} = 10$)—that is, the agent with the most adaptable beliefs about the latent temporal structure—is capable of accurately inferring the correct temporal template $m$, and that the rate at which agent learns correct representation of the temporal structure depends on the given temporal context. Hence we expect that human subjects should also be capable of learning the correct statistics with comparable prior expectations about temporal structure. In Figure 6 we show posterior beliefs over temporal templates in the form of marginal posterior beliefs about the mean $\mu$ and the regularity $\nu$ at each time step of the experiment. We see that the agent quickly learns the correct mean between-reversal duration (already after 200 trials the highest posterior probability is close to $\mu = 19$), but it takes longer (more than 400 trials) to form precise beliefs about the level of temporal regularity. In contrast, in the irregular condition, learning the correct mean between-reversal-interval (fixed to $\mu = 19$ in both conditions) takes more time and is less precise, but the posterior estimates over the precision parameter ($\nu$) converge faster to the correct values (already after 200 trials). Note that having the correct representation of both mean and precision parameters is more important in the regular condition as one can achieve higher improvements in the performance compared to the irregular condition, as we demonstrated previously in Marković et al. (2019).

Simulating the behavioural effect of prior preferences over outcomes

As mentioned above, the prior preference over outcomes $P_a$ parameterise agents’ motivation to collect rewards (generate correct choices) and collect information (engage with the exploratory option). Therefore, it is important to understand how prior preferences interact with performance and probing. We show that the more an agent engages with the exploratory options, (i.e., the higher its preference for choice cues), the better its representation of latent temporal structure, and consequently the higher agent’s performance. This is because selecting exploratory options maximally reduces the uncertainty about the latent state (which option has higher reward probability) which in turn allows an agent to learn a more accurate representation of the latent task dynamics. We visualise these dependencies in Figure 7, where we show what impact changing $p_+$ and $p_-$ have on performance, probing, and the quality of temporal representation after 800 trials. In the supplementary figure Figure 7–Figure Supplement 1 we show same dependencies but with respect to changing $p_+$ and $p_c$, hopefully helping the reader to build an
intuition about interactions between prior preference parameter and behaviour. Note
that in both figures we only consider cases in which $p_+ \geq p_-$ as this reflects higher
prior preference for gains than for losses in the agent, which we would expect to hold for all
subjects.

Model-based analysis of subjects’ choices
Using the model-based analysis of the choice data we next ask whether human subjects
can learn latent temporal regularities in the reversal learning tasks? An individual’s capacity
to learn correct temporal regularity corresponds to their behaviour being associated
with a less precise prior over temporal templates (Eq (1)), that is, larger $\nu_{\text{max}}$. An agent
with imprecise prior over temporal templates is able to learn an accurate representation
of a distribution of between-reversal-intervals, and to form expectations about the moment
of reversals (see Figure 6 and Figure 7) in both conditions. Thus, we anticipated
that between-subject variability in performance and probing would be reflected in dif-
ferent posterior estimates of the most likely $\nu_{\text{max}}$ value associated with the behaviour of
individual subjects.

Therefore, we first classify subjects based on the maximum a-posteriori estimate over
possible values of $\nu_{\text{max}} \in \{1, \ldots, 10\}$, as shown in Figure 8. For each subject we compute
a posterior probability over $\nu_{\text{max}}$ and assign the subject the value of the temporal prior
$\nu_{\text{max}}$ corresponding to the value with the highest exceedance probability (see Model in-
version). Using this procedure we find that 8 out of 41 subjects in the regular condition,
Figure 7. Dependence of performance, probing, and the quality of temporal representations on prior preferences over outcomes. Each dot in the plot corresponds to a single run with fixed prior preferences $P_s = (p_-, p_+, \frac{1}{2} p_+, \frac{1}{2} p_+)$, with temporal prior set to $\nu_{\text{max}} = 10$, and with decision precision set to $\gamma = 5$. For each possible pair $(p_-, p_+) \in \{0.05, 0.06, \ldots, 0.25\} \times \{0.5, 0.55, 0.6\}$ we have repeated $n = 50$ simulations in each condition. Here $\text{RMSE}_x = \sqrt{\sum p(x_i)(x_i - x^*)^2}$ stands for the root mean square error (RMSE) of corresponding parameters $\mu, \nu$ that define temporal template. The RMSE is computed using posterior probabilities $q(\mu, \nu)$ obtained at trial $t = 800$. The performance and the probing are computed as averages over responses from trial $t = 400$ until trial $t = 800$. Note that probing is increasing as we reduce $p_-$ and keep $p_+$ fixed (circles of the same colour), and as we reduce $p_+$ and keep $p_-$ fixed, as the larger the sum $(p_+ + p_-)$ is, the lower is the preference for choice cues $p_c$, and hence the tendency of the agent to engage with the epistemic option. Higher probing (larger circle size) results in higher performance in both conditions (top row). Similarly, both $\text{RMSE}_\mu$ and $\text{RMSE}_\nu$ are lower for larger exploration odds, with the exception of $\text{RMSE}_\nu$ in irregular condition. Note that forming an accurate temporal representation is especially important in the regular condition, where forming correct anticipatory beliefs can substantially improve behavioural adaptation and simplify the problem of balancing between exploratory and exploitative choices. In contrast, in the irregular condition, having a precise representation of temporal structure does not impact performance substantially, and the agent performs better when engaging with the epistemic option more often.

Figure 7–Figure supplement 1. Dependence of performance, probing, and the quality of temporal representations on prior preference over cues $p_c$.

and 17 out of 33 subjects in the irregular condition are assigned to the group with temporal prior $\nu_{\text{max}} > 1$. For the subjects in the regular condition this result suggests that they have learned to anticipate reversals to certain extent. However, it is critical to note that the classification of subjects is not perfect, especially in irregular condition as shown in Figure 8–Figure Supplement 1. However, as our aim is not to identify precisely participants’ temporal prior, but simply to distinguish between subjects that learn temporal regularities ($\nu_{\text{max}} > 1$) from those that do not ($\nu_{\text{max}} = 1$), limiting the analysis to binary
classification leads to the following classification accuracy in simulated data: (i) in the regular condition $v_{\text{max}} = 1$, $\text{ACC} = 1$, and $v_{\text{max}} > 1$, $\text{ACC} = 0.91$, (ii) in the irregular condition $v_{\text{max}} = 1$, $\text{ACC} = 0.4$ and $v_{\text{max}} > 1$, $\text{ACC} = 0.8$. Low classification accuracy in the irregular condition is the reason for substantially higher number of subjects assigned to the $v_{\text{max}} > 1$ model class.

**Figure 8.** Posterior probability over temporal prior $v_{\text{max}}$. Posterior probability of possible $v_{\text{max}}$ values for each subject, reflecting a subject's flexibility to learn latent temporal structure. On the right hand side, we combine posterior estimates into two classes, one for the limiting case $v_{\text{max}} = 1$, and another for all other options $v_{\text{max}} > 1$. This split differentiates subjects not sensitive to temporal regularities from the ones who a priori expected a regular temporal structure of reversals. Note that lighter colours correspond to higher posterior probability.

The posterior estimates of model parameters shown in **Figure 8** show that 65% of the participants were assigned to the model class corresponding to the simplest HMM representation ($v_{\text{max}} = 1$) which assumes maximal irregularity. Importantly, when we plot the time course of both performance and probing, as shown in **Figure 9**, we find the similar trajectory of behavioural measures over the course of the experiment as what we see in simulated data. Namely, that the performance is higher and the probing is lower in the group of participants associated with larger $v_{\text{max}}$ (compare with **Figure 5** and **Figure 5–??)). The result is notably more robust for the regular condition than the irregular. The behavioural trajectories of individual participants are shown in **Figure 9–Figure Supplement 1**.

These findings show a good correspondence between simulated behaviour for different parametrisation of the model ($v_{\text{max}} = 1$ vs $v_{\text{max}} > 1$ in **Figure 5**), and the participants’ behaviour associated with different model classes (**Figure 9**). There are two possible explanations for this: (i) the model inversion accurately captures the participants behaviour
Figure 9. Category based mean estimate of behavioural measures. Each line corresponds to a model class average over behavioural trajectories of subjects assigned to that model class. Note the similarity of the trajectory profiles to the simulated trajectories in Figure 5.

Figure 9–Figure supplement 1. Subject specific trajectories of behavioural measures.

and between-participant sensitivity to temporal regularities of the task, (ii) the group differences come from other free model parameters and do not correspond to differences in sensitivity to temporal structure. To exclude the second option we show in Figure 10 the mean of the posterior estimates of free model parameters $\gamma$, $p_-$ and $p_+$. Note that in both experimental conditions we see a lack of separation between free model parameters associated with each model class.

There are a couple of interesting observations to be made from the posterior expectations of the other model parameters. First, we find in most participants rather large posterior estimates of choice precision $\gamma$, close to $\gamma = 5$, suggesting that choice stochasticity is rather low in most participants. Low choice stochasticity means that choices are well aligned with the choice likelihoods encoded in terms of expected free energy (Eq (18)) and the model is rather accurate in predicting behavioural responses. Second, the posterior estimates of outcome preference parameters $p_-$ and $p_+$ split subjects in two distinct groups, which correspond to their preference for receiving informative cues when selecting exploratory option. The 29 subjects who never engaged with the exploratory option have a higher preference for losses than for informative cues, hence $p_- \geq p_c$. We marked with the dashed grey line the limiting case of $p_- = p_c = \frac{1-p_+}{2}$, which separates the subjects which did not interact with the exploratory option (above the dashed line) and subjects that were relying on exploratory option to reduce their belief uncertainty (bellow the dashed line). Similarly, the participants who prefer informative cues over gains would have prior preferences over either of the cues in the following region $p_c \geq p_+$. The dotted grey line marks the limiting case of $p_+ = p_c = \frac{1-p_-}{2}$.

Discussion

Sequential activity of neuronal assemblies is one of principled neuronal operations that support higher level cognitive functions Eichenbaum (2014); Buzsáki and Llinás (2017)
Figure 10. Posterior median of continuous model parameters. Each dot corresponds to the median of the $n = 1000$ samples from the posterior distribution of $\gamma$, $p_-$, and $p_+$ for each participant. The dashed grey lines denotes equality between preferences for losses and epistemic cues, that is, when $p_+ = \frac{p_-}{2}$ then $p_- = (1 - p_+)/2$. In turn, the dotted grey line indicates equality between preferences for gains and epistemic cues, that is, when $p_+ = \frac{p_-}{2}$ then $p_- = 1 - 3p_+$. Note that we use the same colour coding as in the previous figure to denote classification of participants into different model classes.

and allow humans to form complex spatio-temporal representation of our every day environment Frölich et al. (2021). Akin to grid cells known to support representation of both spatial and nonspatial task states Fu et al. (2021), time cells have been linked to temporal representation of state sequences critical for memory and decision-making Eichenbaum (2014). Importantly, in spite of these fruitful experimental findings we have no clear computational understanding of how humans learn temporal structure in the service of successfully behavioural adaptation.

Here we introduced a novel computational model of behaviour capable of learning latent temporal structure of a probabilistic reversal learning task with multiple reversals (Costa et al., 2015; Reiter et al., 2016, 2017; Vilà-Balló et al., 2017). The computational model combines hidden semi-Markov framework for representing latent temporal structure (Yu, 2015) and active inference for resolving exploration-exploitation trade-off (Friston et al., 2015, 2016). Importantly the model be used for investigating decision making in changing environments in any behavioural task that can be cast as a dynamic multi-armed bandit problem (Markovic et al., 2021; Gupta et al., 2011); of which the reversal learning tasks is a special case corresponding to a specific type of two-armed bandit problem.

The probabilistic reversal learning task, which we utilised to demonstrate flexibility of proposed model, is one of the most established paradigms for investigating human behaviour in changing environments and quantifying cognitive disorders. We used model-based analysis of behavioural data to infer temporal expectations of subjects exposed to one of the two task variants: (i) with regular intervals between reversals, (ii) with irregular intervals between reversals. Notably, being able to form expectations about the
moment of reversal is critical for achieving high performance in the probabilistic reversal learning task, which we illustrate using simulations. We demonstrated that participants behaviour is highly heterogeneous reflecting the differences in participants expectations about temporal regularities. Importantly, the participants expectations about temporal regularities influence their ability to correctly learn latent temporal structure (especially relevant in the condition with regular between reversal intervals), and is reflected in their performance throughout the experiment.

Importantly, we have extended the standard reversal learning task and incorporated an explicit exploratory option in addition to the two standard options whose choice results in monetary gain or loss. This exploratory option informs the participant about the currently correct choice. The additional behavioural response provides us with more direct access to the individual uncertainty about a correct choice and improves model selection. Interestingly, in addition to participants’ diversity of temporal representation, we find stark differences in their preferences to engage with the exploratory option, suggesting individual differences for the value of information \(\text{(Niv and Chan, 2011)}\) and utilised strategies for resolving the exploration-exploitation trade-off. Critically, their epistemic preferences are not obviously correlated with the quality of the learned temporal structure, as in both groups participants show heterogeneous prior expectations about temporal regularities limiting the available temporal templates, hence the accuracy of temporal representations. However, the willingness to engage with the epistemic options does influence participants’ performance, where higher engagement results in better performance. Therefore, these joint findings reveal distinct components of the computational mechanisms that underlie adaptive behaviour in dynamic environments.

In spite of improved from of the reversal learning, the proposed method still has a limited capability of precisely identifying temporal priors from participants’ responses (see Figure 8–Figure Supplement 1). This introduces uncertainty into the cognitive phenotyping, and constrains how specific we can be about the sources of the participant behavioural variability. We expect that a different experimental paradigm is required to separate participants ability of forming accurate temporal representation from the lack of motivation to invest cognitive resources for improved performance. Particularly, in the current reversal learning task it is not possible to separate a range of factors that could contribute to behavioural differences: (i) participants not being able to learn correct latent temporal structure, (ii) participant not expecting that learning higher order statistics can improve performance, and (iii) participants not being motivated enough to invest cognitive resources for performing the task. Being able to decouple contributions of these factors will be critical for applying the behavioural model to cognitive phenotyping and making the approach useful to the research in computational psychiatry and developmental psychology.

To accurately predict future it is critical not only to know that change might be coming but also when the change will occur. To anticipate the changes in our-everyday environments and adapt our behaviour accordingly it is critical accurately estimate and represent elapsed time between relevant events. A range of experimental findings has linked timing of events and hence forecasting the future to underlying Bayesian inference mechanisms \(\text{(Jazayeri and Shadlen, 2010; Griffiths and Tenenbaum, 2011)}\). Most recently \text{Ma-heu et al.} \text{(2022)} has linked sequence learning and prediction in human subjects to an underlying hierarchical Bayesian inference model with distinct hypothesis spaces for statis-
tics and rules corresponding to a set of deterministic temporal templates. The authors conclude that the hierarchical Bayesian inference mechanism underlies human ability to process sequence, similar to hierarchical semi-Markov framework proposed here.

Furthermore, in recent years, various neuroimaging studies have linked different neuro-cognitive domains, such as attention and working memory, to specific spatio-temporal expectations about underlying dynamics of the environment (Nobre and Van Ede, 2018). Interestingly, the human ability to estimate and reproduce elapsed time was also previously linked to reward discounting and intertemporal choice behaviour (Bermudez and Schultz, 2014; Retz Lucci, 2013; Ray and Bossaerts, 2011). For example, McGuire and Kable (2015) demonstrated that “impulsivity” (reluctance to wait for a better reward), depends on the hidden statistics of delays — between an initial bad offer and a later but more valuable offer — which human participants experienced. In Mikhael and Gershman (2019) authors have linked time perception and dopaminergic neuronal activity, demonstrating the role of value-based prediction errors in time representation. Furthermore, time perception and timed behaviour have been linked to all major neuromodulatory systems (Meck, 1996) either directly using neuropharmacological manipulations (Crockett and Fehr, 2014) or indirectly using neurological disorders (Story et al., 2016) and ageing research (Read and Read, 2004).

Together these findings provide important evidence for the role of temporal expectations in goal-directed decision making and let one speculate whether a range of aberrant behaviours might be related to an erroneous representation of the temporal structure of the task. Importantly, a computational behavioural models that we introduced here can emulate the learning of temporal structure, hence can become a potent tool linking aberrant behaviour found in cognitive disorders to erroneous prior beliefs about the rules that govern the dynamics of the environment, as suggested by the active inference account of human behaviour (Friston et al., 2017, 2016, 2015).

To conclude, the results presented here provide novel insights into computational mechanism underlying the human ability to learn hidden temporal structure of the environment and the computational principles they utilise for making decisions based on temporal representations. The fact that we find behavioural heterogeneity in a population of healthy young adults suggests a potential use of the proposed design and behavioural model for cognitive phenotyping and for revealing causes of aberrant behaviour in clinical populations.
Methods and Materials

Ethics statement
All subjects provided written informed consent and were paid on an hourly basis. The Ethical Board of Technical University Dresden approved the study.

Experiment
Probabilistic reversal learning
In the experimental task subjects were deciding between two cards shown on a screen, each showing a different stimulus (a geometric shape, e.g. rectangle, triangle or a question mark) as shown in Figure 1. The reward probabilities associated with the two choice options were anti-correlated on all trials: whenever reward probability of choice A was high ($p_H = 0.8$) the reward probability of choice B was low ($p_L = 0.2$), and vice versa. Note that $p_H = 1 - p_L$ on all trials. The location of each stimulus on the screen (right or left side) was kept fixed over trials. After each choice the stimulus was highlighted and depicted for 1.5 s minus the reaction time. The feedback in the form of a gain or a loss was shown for 0.5 s. Similarly, the feedback after an exploratory choice was also shown for 0.5 s. If no response occurred during the decision window of 3 s, the message “too slow” was presented, and no outcome was delivered.

All subjects underwent a training session during which they had the opportunity to learn the statistics of the rewards associated with high $p_H$ and low $p_L$ reward probability choices. The set of stimuli used in the training phase differed from the one used during the testing phase. Subjects were instructed that they could either win or lose 10 cents on each trial, and that they will be paid the total amount of money they gained during the testing phase. Each subject performed 40 training trials with a single reversal after the 20th trial. Before the start of the testing phase subjects were told that the reward probabilities might change at regular intervals (in both conditions) over the course of the experiment. No other information about reversals or the correlation of choices and outcomes was provided. Thus, the subjects had no explicitly instructed knowledge about the anti-correlated reward probabilities or between-reversal-intervals before the experiment.

Note that, out of $n = 74$ participants $n_p = 24$ were exposed to the variant of the reversal learning task without epistemic option. This group of subjects belongs to an initial pilot study that used the standard two-choice task design. In the pilot study 14 subjects were assigned to the regular condition and 10 to the irregular condition. We decided to include the subjects from the pilot into the analysis, as we noticed that almost 30% of subjects, in the post pilot group, choose not to interact at all with the exploratory option, even when that was a possibility. We will not explore this finding here in more detail, but we can exclude their misunderstanding of the task as a potential confound, as we provided a detailed instructions and training before they performed the task (see Experiment for more details).

Behavioural measures
To quantify behaviour we have used two summary measures: (i) performance, defined as odds of making a correct choice, and (ii) probing, defined as odds of making an exploratory choice.
Figure 11. Computing behavioural performance. The process of computing performance. A. We label subjects’ responses as either correct or incorrect, depending on whether a card with higher reward probability was selected or not. B We compute a probability of making a correct choice within a 200 trial window, centred at the current trial number $t$. C For each trial we compute performance as odds of being correct.

The process of computing performance is illustrated in Figure 11. We first label subjects’ responses as either correct or incorrect, depending on whether a card with higher reward probability was selected or not, see Figure 11A. Then we compute a probability of making a correct choice within a 201 trial window, centred at the current trial number $t$, see Figure 11B. Finally, for each trial we compute performance as odds of being correct, see Figure 11C.

The probing is computed in similar manner to performance, with the only difference that we label choices as either exploratory or exploitative depending on whether subjects have chosen the exploratory option (middle card in Figure 1), or not. Probing is defined as the odds of selecting the exploratory option within a 200 trials time window.

**Behavioural model**

To introduce the generative model of task dynamics, and subsequently derive the behavioural model via model inversion methods, we will consider the following features of the task. At any trial the task environment is in one of the two possible states, defined as the configuration of reward contingencies. For example, state one corresponds to stimulus A being associated with a high reward probability $p_H$, and state two to stimulus B being associated with a low reward probability $p_L$. Subjects do not know in advance how likely rewards and losses are when making a correct choice compared to making an incorrect choice, and this is something they have to learn during the course of experiment. In other words, we also treat reward probabilities ($p_H$, and $p_L$) as latent variables. Between trials the state can change, i.e. when a reversal occurs but only after a certain
minimum number of trials has elapsed since the last state change. Depending on the experimental condition the between reversal duration will either be semi-regular (occurring every 20 trials with small variability) or irregular (occurring every 20 trials, but with maximal variability).

The explicit representation of state duration enables us to associate changes in state transition probabilities with the current trial and the moment of the last change. The dependence of state transition probability on the number of trials since the last change corresponds to the formalism of hidden semi-Markov models (HSMM) (Yu, 2010; Murphy, 2002), which allows mapping complex dynamics of non-stationary time series to a hierarchical, time aware, hidden Markov model. However, using an explicit representation of context duration is inefficient, as it requires an enormous state space representation. Here, we will instead adopt a phase-type representation of duration distribution (Varmazyar et al., 2019) which substitutes duration variable $d \in \{1, \ldots, \infty\}$ with a phase variable $f \in \{1, \ldots, f_{\text{max}}\}$, allowing for a finite state representation of an infinite duration state space.

In what follows we will define the components of the generative model (observation likelihood, the dynamics of latent variables, and the parametrisation of the dynamics) and derive the corresponding update rules for latent variables and state, hence enabling the learning of different temporal contexts during the experiment. The graphical representation of the generative model is shown in Figure 4.

Practically we introduce four latent states, to describe the task on any trial:

- First, the configuration of reward contingencies can be in one of the two possible states. Hence, $s^1_t \in \{1, 2\}$ which describes which card is associated with high reward probability and which with low reward probability.
- Second, choosing one of the options on a given trial corresponds to setting the task in one of the three possible choice states $s^2_t \in \{1, 2, 3\}$ (chosen left card, chosen middle card – exploratory option, and chosen right card) corresponding to the chosen option. The choice of the option is deterministic and this state is always known with certainty after the choice is made.
- Third, current phase $f_t \in \{1, \ldots, \nu + 1\}$ of the task dynamics. The phase latent variable controls transitions of latent state $s^1_t$, where the change of state is only possible if the end phase ($f_t = \nu + 1$) is active on the current trial. Note that the larger the number of phases is (parameter $\nu \in 1, \ldots$) the more regular is the occurrence of reversals. We have limited here the number of phases by setting $\nu = 10$, as this is sufficiently large for accurate representation of reversal dynamics in regular condition.
- Fourth, temporal template $m$. Latent temporal template defines the frequency of reversals, $\mu$ (mean between-reversal duration) and the number of latent phases $\nu$, that is the regularity of reversals.

**Observation likelihood**
The observation likelihood links latent states ($s^1_t$, and $s^2_t$) with probabilities of observing different possible outcomes in those states.

In the temporal reversal learning task there are four possible outcomes: (1) loss of 10 Eurocents, (2) gain of 10 Eurocents, (3) the correct card is left card, or (4) the correct card is the right card. Therefore, we define the observation likelihood as a categorical
distribution

\[ p \left( o_i | \rho, s_1^i, s_2^i \right) = \prod_{i=1}^{4} \delta_{o_i, s_1^i, s_2^i} \]  

(3)

where \( i \) denotes the outcome type, \( o_i \in \{1, \ldots, 4\} \). The probabilities of different outcomes are parameterised via \( \rho_{s_1^i, s_2^i} \), where each state tuple \((s_1^i, s_2^i)\) corresponds to a unique probability of observing any of four possible outcomes. We define prior beliefs about outcome probabilities in the form of a product of Dirichlet distributions

\[ p(\rho) = \prod_{s_1=1}^{4} \prod_{s_2=1}^{3} \text{Dir}(\rho_{s_1, s_2} | \alpha_{s_1, s_2}^0) \]  

(4)

We set the parameters of Dirichlet priors to the following values:

\[
\begin{array}{c|ccc}
  o_i \backslash s^2 & 1 & 2 & 3 \\
\hline
  1 & 6 & 1 & 32 \\
  2 & 32 & 1 & 6 \\
  3 & 1 & 1000 & 1 \\
  4 & 1 & 1 & 1 \\
\end{array}
\quad
\begin{array}{c|ccc}
  o_i \backslash s^2 & 1 & 2 & 3 \\
\hline
  1 & 32 & 1 & 6 \\
  2 & 6 & 1 & 32 \\
  3 & 1 & 1 & 1 \\
  4 & 1 & 1000 & 1 \\
\end{array}
\]

(5)

The above configuration for the parametrisation of prior Dirichlet probabilities reflects an assumption that the participants have formed during training an initial – vague beliefs – about reward probabilities associated with different actions in different states. We assume that participants are highly certain that selecting the epistemic option does not return gain or loss (high value of \( \alpha_{s_1=1, s_2=2}^0 \) for the corresponding outcome in both states). Furthermore, we assume that participants have formed good expectations gain/loss probabilities \( \langle p_H \rangle = \frac{32}{40} = 0.8 \), and \( \langle p_L \rangle = \frac{6}{40} = 0.15 \), but that they are still uncertain about the exact values. Weak priors about outcome probabilities allow for ongoing adaptation of beliefs during the course of experiment.
Hidden state dynamics

To formalise the presence of sequential reversals, we define the phase dependent state transition probability as follows

$$p(s_t^1 | s_{t-1}^1, f_{t-1}) = \begin{cases} I_2, & \text{if } f_{t-1} = v + 1, \\ J_2 - I_2, & \text{if } f_{t-1} \leq v, \end{cases} \quad (6)$$

where $I_2$ denotes the $2 \times 2$ identity matrix and $J_2$ denotes the $2 \times 2$ all-ones matrix. The above relations describe a simple deterministic process for which the current state $s_t^1$ remains unchanged as long as the phase variable $f_{t-1}$ remains below the end phase, $v + 1$. The transition between states occurs with certainty (e.g. if $s_{t-1}^1 = 1$ then $s_t^1 = 2$) once the end phase is reached, that is, when $f_{t-1} = v + 1$.

Although it is possible to condition state changes on a duration variable $d$, as demonstrated in Marković et al. (2019), such an explicit representation is inefficient as it requires large state spaces (Yu and Kobayashi, 2003; Vaseghi, 1995). Here we adopt the discrete phase-type (DPH) representation of duration distribution (Varmazyar et al., 2019). The DPH representation defines transitions between phase variables $f_t$ and the following parametrisation of phase transition probabilities corresponds to the DPH representation of the negative binomial distribution

$$p(f_t | f_{t-1}, m) = \begin{cases} \delta_m, & \text{if } f_{t-1} \leq v, \text{ and } f_t = f_{t-1} + 1 \\ 1 - \delta_m, & \text{if } f_{t-1} \leq v, \text{ and } f_t = f_{t-1} \\ \pi_{f_t}^m, & \text{if } f_{t-1} = v + 1, \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

where $\pi_i^m = \binom{v}{i-1} (1 - \delta_m)^{v-i-1} \delta_m^{i-1}$ for $i < v + 1$, and $\pi_{v+1}^m = 1 - \sum_{i=1}^{v} \pi_i^m$.

The corresponding negative binomial distribution of between-reversal duration can be expressed as follows

$$p_m(d) = \binom{d + v - 2}{d - 1} (1 - \delta_m)^{d-1} \delta_m^v; \quad d \in \{1, 2, \ldots\} \quad (8)$$

Figure 12. Negative binomial distribution. We illustrate here the changes in the negative binomial distribution as a function of shape parameter $v$ which is inversely proportional to the variance of between reversal durations. Note that for higher values of $v$ the distribution peaks around its expected value (dashed line). As the variance increases (green) the mode shifts toward zero. The limiting case of the negative binomial distribution in the form of geometric distribution (red) corresponds to $v = 1$. For all three cases we fixed the mean duration to the same value.
where the expected duration corresponds to

\[ E_{\mu m}[d] = \frac{\nu (1 - \delta_m)}{\delta_m} + 1 = \mu + 1; \quad \delta_m = \frac{\nu}{\mu + \nu}, \]  

(9)

and variance, hence uncertainty about duration regularity, to

\[ Var_{\mu m}[d] = \mu + \frac{\mu^2}{\nu}. \]  

(10)

Note that the parameter \( \nu \) of the negative binomial distribution, acts as a precision parameter. We illustrate this in Figure 12.

The choice of prior beliefs about the between-reversal interval \( d \) in the form of a negative binomial distribution has interesting consequences on the dynamics of the marginal probability that a reversal will occur at some future point

\[ m[t] = p(s_{1t}^1|s_{t-1}^1 = 2, f_{t-1} = \nu + 1, m) \]

\[ = \sum_{f_{t},... f_{t+1}} \sum_{s_{1t}} p(s_{1t}^1, f_{t}, f_{t-1} = \nu + 1, m) \prod_{k=t+1}^{t+r} p(s_{k}^1, f_{k} | s_{k-1}^1, f_{k-1}, m) \]  

(11)

In Figure 13 we show the dependence of the future reversal probability \( \delta_m[r] \) on the precision parameter \( \nu \), given a fixed mean duration \( E_{\mu m}[d] = 20 \). Note that for \( \nu = 1 \) we get a constant transition probability, which corresponds to the expectations of change probabilities found in hidden Markov models. In contrast, for larger values of \( \nu \) one obtains a trial-dependent, effective transition probability with values alternating between low and high probabilities in a periodic manner. This temporal dependence of the transition probability will affect the inference process. The agent will become insensitive to subsequent reversals occurring a few trials after the previous reversal, and highly sensitive to reversals occurring twenty to thirty trials after the previous reversal.

Finally, the choice states \( s_{1t}^2 \) are fully dependent on the current choice \( a_t \in \{1, 2, 3\} \), and we express the state transition probability as

\[ p(s_{1t}^2 | s_{t-1}^2, a_t) = p(s_{1t}^2 | a_t) = \delta_{s_{1t}^2, a_t}. \]  

(12)

In practice this means that the agent is always certain about the choice it made and how that choice impacted the state of the task. Therefore, the posterior estimate over \( s_{1t}^2 \) can be trivially expressed as

\[ q(s_{1t}^2 | a_t) = \delta_{s_{1t}^2, a_t}. \]
Active Inference

In active inference, agents form posterior beliefs both about latent states of the environment and about their own actions. In other words, both perception and action selection are cast as inference problems (Botvinick and Toussaint, 2012; Attias, 2003). Practically, we will use variational inference for defining update rules for beliefs (Friston et al., 2017; Blei et al., 2017). In what follows we will first introduce perception as minimisation of the variational free energy (upper bound on log-marginal likelihood) with respect to posterior beliefs over latent states, and after that introduce action selection as minimisation of the expected free energy (Smith et al., 2022), that is, expected surprisal about future outcomes.

We write the generative model of outcomes $o_t$ on trial $t$ as

$$
\bar{p} \left( o_t, s_t^1, s_t^2, f_t, m, \rho \right) = p \left( o_t | s_t^1, s_t^2, \rho \right) \tilde{p}_i \left( s_t^1 | f_t \right) p \left( s_t^2 | a_t \right) \tilde{p}_i \left( f_t | m \right) \tilde{p}_i \left( m \right) \tilde{p}_i \left( \rho \right),
$$

(13)

where we use $\tilde{p}_i(\cdot)$ to denote prior beliefs conditioned on a sequence of past outcomes, $o_{1:t-1} = (o_1, \ldots, o_{t-1})$ and choices $a_{1:t-1} = (a_1, \ldots, a_{t-1})$. Given a choice $a_t^i$ and an observed outcome $o_t$ at trial $t$, the approximate posterior belief $q_{t}(x)$ over latent states $x = (s_t^1, s_t^2, f_t, \rho, m)$ is obtained in two steps:

- We first compute the marginal likelihood with respect to $\tilde{p}_i(\rho)$, and obtain the exact marginal posterior over discrete states using the Bayes rule

$$
q_{t} \left( s_t^1, s_t^2, f_t, m \right) = \frac{\tilde{p}_i \left( o_t, s_t^1, s_t^2, f_t, m \right)}{\tilde{p}_i \left( o_t \right)}.
$$

(14)

- Given the marginal posterior $q_{t} \left( s_t^1, s_t^2 \right) = \sum_{f_t, m} q_{t} \left( s_t^1, s_t^2, f_t, m \right)$ we compute the posterior over outcome probabilities using the variational message passing update

$$
q_{t} \left( \rho \right) \propto \tilde{p}_i \left( \rho \right) \sum_{s_t^1, s_t^2} q_{t} \left( s_t^1, s_t^2 \right) \ln p \left( o_t | s_t^1, s_t^2, \rho \right).
$$

(15)

As we initially defined the prior over outcome probabilities in the form of a Dirichlet distribution with parameters $a^0$, we can express the posterior estimate on every trial in the same functional form. Hence,

$$
q_{t} \left( \rho \right) = \prod_{s_t^1} \prod_{s_t^2} \text{Dir} \left( \rho_{s_t^1, s_t^2} | a_{s_t^1, s_t^2}^t \right)
$$

(16)

where

$$
a_{s_t^1, s_t^2}^t = a_{s_t^1, s_t^2}^{t-1} + \delta_{a_t^i} \cdot q \left( s_t^1 = s_t^1 \right) + a_{s_t^1, s_t^2 = a_t^i}^{t-1},
$$

(17)

and $\tilde{p}_i(\rho) = q_{t-1}(\rho)$. The above belief updating scheme corresponds to the variational surprise minimization learning algorithm (Liakoni et al., 2021; Markovic et al., 2021) adapted to the categorical likelihood and the Dirichlet prior.

Action selection

In active inference, decision strategies (behavioural policies) are chosen based on a single optimisation principle: minimising expected surprisal about observed and future outcomes, that is, the expected free energy (Smith et al., 2022; Schwartenbeck et al., 2019).
Here, we will express the expected free energy of a choice $a$ on trial $t$ as

$$G_a = D_{KL}(\tilde{p}(o_t|a)||P(o_t)) + E_{\tilde{p}(s^1_t|o_t)}[H[o_t|\rho, s^1_t, s^2_t = a]]$$

$$\approx -E_{\tilde{p}(o_t|a)}[\ln P(o_t)]$$

$$- E_{\tilde{p}(o_t|a)}[D_{KL}(q_t(s^1_t, s^2_t, f_t|o_t, a) || \tilde{p}_t(s^1_t, s^2_t, f_t|a))]$$

$$- E_{\tilde{p}(o_t|a)}[D_{KL}(q_t(o_t|a) || \tilde{p}_t(o)) + D_{KL}(q_t(m|a) || \tilde{p}_t(m))]$$

where $P(o_t)$ denotes prior preferences over outcomes, $H[o_t|\rho, s^1_t, s^2_t]$ the entropy of outcome likelihood $p(o_t|\rho, s^1_t, s^2_t)$, and $D_{KL}(p||q)$, stands for the Kullback-Leibler divergence between two probability densities: $p$ and $q$. Note that action selection based on minimisation of expected free energy would have an implicit dual imperative (see the different factorisations in Eq (18)): On one hand, the expected free energy combines ambiguity and risk. On the other hand, it consists of information gain (epistemic value + novelty) and extrinsic value. Therefore, selecting actions that minimise the expected free energy dissolves the exploration-exploitation trade-off, as every action contains both expected value and information gain. This is a critical feature of action selection which allows us to account for epistemic choices as used in our experimental paradigm (see Figure 1).

At any trial $t$ choice $a_t$ is sampled from choice beliefs $p(a_t)$ (cf. planning as inference Botvinick and Toussaint (2012); Attias (2003)) defined as

$$a_t \sim p(a_t|\gamma, P_o, \nu_{max}) \propto e^{-\gamma G_a[p_o, \nu_{max}, t]}$$

where parameter $\gamma$ corresponds to choice precision, which we will attribute to empirical choice behaviour of participants. Therefore, for describing participants’ behaviour we assume that the action selection process is corrupted by external sources of noise; e.g. mental processes irrelevant for the task at hand. In our simulations we will fix $\gamma$ to a reasonably large value, to achieve approximate free energy minimisation as the following relation will be satisfied

$$a_t \approx \arg\min_a G_a, \text{ when } \gamma \gg 1.$$  

(20)

Notably, here we consider the simplest form of active inference in which expected free energy is computed from a one-step-ahead prediction. This is a standard simplification for environments in which actions cannot interfere with the state transitions, as is the case in typical dynamic multi-armed bandit problems (Markovic et al., 2021).

To express the expected free energy, $G(a_t)$, in terms of beliefs about arm-specific reward probabilities, we will first constrain the prior preference to the following categorical distribution

$$P(o_t) = \prod_{o_t} [P_o]^\delta_{o_t}, \quad P_o = \left(p_-, p_+, \frac{1}{2} p_c, \frac{1}{2} p_c \right)$$

(21)
In active inference, prior preferences determine whether a particular outcome is attractive, that is, rewarding. Here we assume that all agents prefer gains ($o = 2$) over losses ($o = 1$). Hence, we constrain parameter values such that $p_+ > p_-$ holds always. The ratio $\frac{p_+}{p_-} = \lambda$ determines the balance between epistemic and pragmatic imperatives. When prior preferences for gains are very precise, corresponding to large $\lambda$, the agent becomes risk sensitive and will tend to forgo exploration if the risk is high; see Eq. 18. Conversely, a low lambda corresponds to an agent which is less sensitive to risk and will engage in exploratory, epistemic behaviour, until it has familiarised itself with the environment.

Given the following expressions for the marginal predictive likelihood,

\[
\tilde{p}_t(o_t|a) = \sum_{s_1, s_2} \int p(o_t|\theta, s_1^t, s_2^t) \tilde{p}_t(s_1^t) p(s_2^t|a) \tilde{p}_t(\theta) d\theta
\]

we get the following expressions for the expected free energy

\[
G_t(a) = \sum_o \tilde{p}_o^{-1} \ln \frac{\tilde{p}_o^{-1}}{p_o} - \sum_s \tilde{p}_t(s_1^t = s) \tilde{p}_t^{-1} \sum_o [\mu_o^{-1}]_{s_1, a, o} \left( \psi \left( \frac{\alpha_o^{-1}}{\alpha_o} + 1 \right) - \psi \left( 1 + \sum_j a_j \right) \right)
\]

(23)

Above we have used the following relation

\[
\int dx \text{Dir}(x|\alpha) x_i \ln x_i = \frac{\alpha_i}{\sum_j \alpha_j} \left( \psi(\alpha_i + 1) - \psi(1 + \sum_j \alpha_j) \right)
\]

(24)

for computing ambiguity term in Eq (18).

**Model inversion**

Given a sequence of subjects’ responses $A^n = (a_1^n, \ldots, a_T^n)$, where $n$ denotes subject index and $T = 1000$ denotes the total number of trials, the response likelihood is defined as

\[
P(A^n|y, P_o, v_{max}) = \prod_t p(a_t = a_t^n|y, P_o, v_{max}).
\]

(25)

Note that for estimating the posterior over model parameters $(y, P_o, v_{max})$ we ignore the first 400 responses from the likelihood. We expect that during these first trials, subjects are still getting used to the task, and potentially use additional strategies for representing the task and making choices. As we do not model all possible task representations, exclusion of initial trials reduces the noise in model comparison. Importantly, we do use the entire set of responses for computing belief trajectories of the active inference agents, that is, we expose the agent to the complete sequence of individual responses and the corresponding outcomes.
For the posterior estimates from the behavioural data we used the following parametric generative model of behaviour, a mixture model:

\[
\begin{align*}
\rho^n_v &\sim \text{Dir} \left( \rho | \alpha \right), \quad \alpha_v = \frac{1}{2}, \quad v \in \{1, \ldots, 10\}, \\
p^n_i &\sim \text{Dir} \left( \rho | \beta \right), \quad \beta_i = 1, \quad i \in \{1, 2, 3\}, \\
P^n_0 &= \left( \frac{p^n_1}{2} \cdot \frac{p^n_1}{2} + \frac{p^n_2}{2} \cdot \frac{p^n_3}{2} \right), \\
\gamma^n &\sim \Gamma^{-1} \left( 2, 2 \right), \\
A^n &\sim \sum_v \prod_{t=400}^{1000} p \left( a_t | \gamma^n, P^n_0, \nu_{\text{max}} = v \right).
\end{align*}
\] (26)

This mixture model unifies parameter estimation with model comparison, as marginal posterior estimates \( p \left( r^n_v | A^n \right) \) corresponds to the posterior probability of the precision parameter \( \nu_{\text{max}} \). To classify subjects’ behaviour in terms of adaptability of temporal representations we use the exceedance probability of the marginal posterior defined as

\[
\begin{align*}
in,s^n &= \arg\max_v r^n_v, \\
X^n_i &= \frac{1}{S} \sum_{s=1}^{S} \delta_i, in,s^n, \\
\end{align*}
\] (27)

where \( r^n_v \sim p \left( r^n_v | A^n \right) \) denotes a sample from the posterior. Finally, the most likely precision parameter \( \nu_{\text{max}}^n \) of the \( n \)th subject corresponds to

\[
\nu_{\text{max}}^n = \arg\max_i X^n_i.
\]

We used the probabilistic programming library Numpyro (Phan et al., 2019) for implementing the full generative model. Numpyro library provides an interface to multiple state-of-the-art inference schemes. For drawing samples from the posterior we have used Numpyro’s implementation of the No-U-Turn sampler (NUTS) (Hoffman et al., 2014). NUTS is an self-tuning version of the Hamiltonian Monte Carlo, a popular Markov Chain Monte Carlo algorithm for avoiding random walks and sensitivity to between-parameter correlations.

Acknowledgments
The authors thank Karl Friston for valuable suggestions and discussions.

Funding
The study was supported by by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft), SFB 940/3 – Project A09 awarded to SS and SFB 940/3 – Project B7 awarded to AMFR. SS acknowledges further support by DFG TRR 265/1 (Project ID 402170461, B09) and Germany’s Excellence Strategy – EXC 2050/1 (Project ID 390696704) – Cluster of Excellence “Centre for Tactile Internet with Human-in-the-Loop” (CeTI) of Technische Universität Dresden. AMFR acknowledges further support by the German Research Foundation (DFG RE 4449/1-1, RTG 2660-B2) and by a 2020 BBRF NAR-SAD Young Investigator Grant.
from the Brain and Behavior Research Foundation. This study was partially funded by funding opportunities for young scientists (Anschubfinanzierung) from the Department of Psychology of the Technische Universität Dresden.

References

Attias H. Planning by probabilistic inference. In: AISTATS Citeseer; 2003.

Bermudez MA, Schultz W. Timing in reward and decision processes. Philosophical Transactions of the Royal Society B: Biological Sciences. 2014; 369(1637):20120468.

Blei DM, Kucukelbir A, McAuliffe JD. Variational inference: A review for statisticians. Journal of the American statistical Association. 2017; 112(518):859–877.

Borst JP, Anderson JR. The discovery of processing stages: Analyzing EEG data with hidden semi-Markov models. NeuroImage. 2015; 108:60–73.

Botvinick M, Toussaint M. Planning as inference. Trends in cognitive sciences. 2012; 16(10):485–488.

Bradtke S, Duff M. Reinforcement learning methods for continuous-time Markov decision problems. Advances in neural information processing systems. 1994; 7.

Buhusi CV, Meck WH. What makes us tick? Functional and neural mechanisms of interval timing. Nature reviews neuroscience. 2005; 6(10):755–765.

Buzsáki G, Llinás R. Space and time in the brain. Science. 2017; 358(6362):482–485.

Costa VD, Tran VL, Turchi J, Averbeck BB. Reversal learning and dopamine: a bayesian perspective. Journal of Neuroscience. 2015; 35(6):2407–2416.

Crockett MJ, Fehr E. Pharmacology of economic and social decision making. In: Neuroeconomics Elsevier; 2014.p. 259–279.

Daw N, Courville AC, Touretzky D. Timing and partial observability in the dopamine system. Advances in neural information processing systems. 2002; 15.

Duong TV, Bui HH, Phung DQ, Venkatesh S. Activity recognition and abnormality detection with the switching hidden semi-markov model. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), vol. 1 IEEE; 2005. p. 838–845.

Eagleman DM. Human time perception and its illusions. Current opinion in neurobiology. 2008; 18(2):131–136.

Eichenbaum H. Time cells in the hippocampus: a new dimension for mapping memories. Nature Reviews Neuroscience. 2014; 15(11):732–744.

Eichenbaum H. On the integration of space, time, and memory. Neuron. 2017; 95(5):1007–1018.

Friston K, Buzsáki G. The functional anatomy of time: what and when in the brain. Trends in cognitive sciences. 2016; 20(7):500–511.

Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, Pezzulo G. Active inference: a process theory. Neural computation. 2017; 29(1):1–49.

Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, Pezzulo G, et al. Active inference and learning. Neuroscience & Biobehavioral Reviews. 2016; 68:862–879.
Friston K, Rigoli F, Ognibene D, Mathys C, Fitzgerald T, Pezzulo G. Active inference and epistemic value. Cognitive neuroscience. 2015; 6(4):187–214.

Frölich S, Marković D, Kiebel SJ. Neuronal Sequence Models for Bayesian Online Inference. Frontiers in Artificial Intelligence. 2021; 4:50.

Fu Z, Beam D, Chung JM, Reed CM, Mamelak AN, Adolphs R, Rutishauser U. The geometry of domain-general performance monitoring in the human medial frontal cortex. Available at SSRN 3891069. 2021; .

Gales M, Young S. The application of hidden Markov models in speech recognition. Now Publishers Inc; 2008.

Griffiths TL, Tenenbaum JB. Predicting the future as Bayesian inference: People combine prior knowledge with observations when estimating duration and extent. Journal of Experimental Psychology: General. 2011; 140(4):725.

Gupta N, Granmo OC, Agrawala A. Thompson sampling for dynamic multi-armed bandits. In: 2011 10th International Conference on Machine Learning and Applications and Workshops, vol. 1 IEEE; 2011. p. 484–489.

Hoffman MD, Gelman A, et al. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J Mach Learn Res. 2014; 15(1):1593–1623.

Itskov V, Curto C, Pastalkova E, Buzsáki G. Cell assembly sequences arising from spike threshold adaptation keep track of time in the hippocampus. Journal of Neuroscience. 2011; 31(8):2828–2834.

Janssen J, Limnios N. Semi-Markov models and applications. Springer Science & Business Media; 2013.

Jazayeri M, Shadlen MN. Temporal context calibrates interval timing. Nature neuroscience. 2010; 13(8):1020–1026.

Kaplan R, Friston KJ. Planning and navigation as active inference. Biological cybernetics. 2018; 112(4):323–343.

Kiebel SJ, Daunizeau J, Friston KJ. A hierarchy of time-scales and the brain. PLoS Comput Biol. 2008; 4(11):e1000209.

Liakoni V, Modirshanechi A, Gerstner W, Brea J. Learning in volatile environments with the bayes factor surprise. Neural Computation. 2021; 33(2):269–340.

MacDonald CJ, Fortin NJ, Sakata S, Meck WH. Retrospective and prospective views on the role of the hippocampus in interval timing and memory for elapsed time. Timing & Time Perception. 2014; 2(1):51–61.

Maheu M, Meyniel F, Dehaene S. Rational arbitration between statistics and rules in human sequence processing. Nature Human Behaviour. 2022; p. 1–17.

Marković D, Reiter AM, Kiebel SJ. Predicting change: Approximate inference under explicit representation of temporal structure in changing environments. PLoS computational biology. 2019; 15(1):e1006707.

Markovic D, Stojic H, Schwoebel S, Kiebel SJ. An empirical evaluation of active inference in multi-armed bandits. arXiv preprint arXiv:210108699. 2021; .
McGuire JT, Kable JW. Decision makers calibrate behavioral persistence on the basis of time-interval experience. Cognition. 2012; 124(2):216–226.

McGuire JT, Kable JW. Medial prefrontal cortical activity reflects dynamic re-evaluation during voluntary persistence. Nature neuroscience. 2015; 18(5):760–766.

Meck WH. Neuropharmacology of timing and time perception. Cognitive brain research. 1996; 3(3-4):227–242.

Mikhael JG, Gershman SJ. Adapting the flow of time with dopamine. Journal of neurophysiology. 2019; 121(5):1748–1760.

Murphy KP. Hidden semi-Markov models (HSMMs). unpublished notes. 2002; 2.

Niv Y, Chan S. On the value of information and other rewards. Nature neuroscience. 2011; 14(9):1095–1097.

Nobre AC, Van Ede F. Anticipated moments: temporal structure in attention. Nature Reviews Neuroscience. 2018; 19(1):34.

Parr T, Markovic D, Kiebel SJ, Friston KJ. Neuronal message passing using Mean-field, Bethe, and Marginal approximations. Scientific reports. 2019; 9(1):1–18.

Phan D, Pradhan N, Jankowiak M. Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro. arXiv preprint arXiv:191211554. 2019; .

Purcell BA, Kiani R. Hierarchical decision processes that operate over distinct timescales underlie choice and changes in strategy. Proceedings of the national academy of sciences. 2016; 113(31):E4531–E4540.

Parr T, Markovic D, Kiebel SJ, Friston KJ. Neuronal message passing using Mean-field, Bethe, and Marginal approximations. Scientific reports. 2019; 9(1):1–18.

Phan D, Pradhan N, Jankowiak M. Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro. arXiv preprint arXiv:191211554. 2019; .

Purcell BA, Kiani R. Hierarchical decision processes that operate over distinct timescales underlie choice and changes in strategy. Proceedings of the national academy of sciences. 2016; 113(31):E4531–E4540.

Ray D, Bossaerts P. Positive temporal dependence of the biological clock implies hyperbolic discounting. Frontiers in neuroscience. 2011; 5:2.

Read D, Read NL. Time discounting over the lifespan. Organizational behavior and human decision processes. 2004; 94(1):22–32.

Reiter AM, Deserno L, Kallert T, Heinze HJ, Heinz A, Schlagenhauf F. Behavioral and neural signatures of reduced updating of alternative options in alcohol-dependent patients during flexible decision-making. Journal of Neuroscience. 2016; 36(43):10935–10948.

Reiter AM, Heinze HJ, Schlagenhauf F, Deserno L. Impaired flexible reward-based decision-making in binge eating disorder: evidence from computational modeling and functional neuroimaging. Neuropsychopharmacology. 2017; 42(3):628–637.

Retz Lucci C. Time, self, and intertemporal choice. Frontiers in neuroscience. 2013; 7:40.

Schwartenbeck P, FitzGerald T, Dolan R, Friston K. Exploration, novelty, surprise, and free energy minimization. Frontiers in psychology. 2013; 4:710.

Schwartenbeck P, Passecker J, Hauser TU, FitzGerald TH, Kronbichler M, Friston KJ. Computational mechanisms of curiosity and goal-directed exploration. Elife. 2019; 8:e41703.

Shappell H, Caffo BS, Pekar JJ, Lindquist MA. Improved state change estimation in dynamic functional connectivity using hidden semi-Markov models. NeuroImage. 2019; 191:243–257.

Shi Z, Church RM, Meck WH. Bayesian optimization of time perception. Trends in cognitive sciences. 2013; 17(11):556–564.
Smith R, Friston Kj, Whyte CJ. A step-by-step tutorial on active inference and its application to empirical data. Journal of Mathematical Psychology. 2022; 107:102632.

Story GW, Moutoussis M, Dolan RJ. A computational analysis of aberrant delay discounting in psychiatric disorders. Frontiers in Psychology. 2016; 6:1948.

Varmazyar M, Akhavan-Tabatabaei R, Salmasi N, Modarres M. Classification and properties of acyclic discrete phase-type distributions based on geometric and shifted geometric distributions. Journal of Industrial Engineering International. 2019; 15(4):651–665.

Vaseghi S. State duration modelling in hidden Markov models. Signal processing. 1995; 41(1):31–41.

Vilà-Balló A, Mas-Herrero E, Ripollés P, Simó M, Miró J, Cucurell D, López-Barroso D, Juncadella M, Marco-Pallarés J, Falip M, et al. Unraveling the role of the hippocampus in reversal learning. Journal of Neuroscience. 2017; 37(28):6686–6697.

Yu SZ. Hidden semi-Markov models. Artificial intelligence. 2010; 174(2):215–243.

Yu SZ. Hidden Semi-Markov models: theory, algorithms and applications. Morgan Kaufmann; 2015.

Yu SZ, Kobayashi H. An efficient forward-backward algorithm for an explicit-duration hidden Markov model. IEEE signal processing letters. 2003; 10(1):11–14.
**Figure 7—Figure supplement 1.** As in the main figure each line corresponds to an average over \( n = 50 \) simulated trajectories with \( v_{\text{max}} = 10 \) and \( \gamma = 5 \).

**Figure 8—Figure supplement 1.** Probability of assigning simulated behavioural responses to possible \( v_{\text{max}} \) values given the true \( v_{\text{max}} \in \{1, \ldots, 10\} \). The confusion matrix was estimated based on \( n = 100 \) simulated responses (50 simulated subjects in each condition) for each generative \( v_{\text{max}} \) value. In all simulation the remaining free model parameters were fixed to \( \gamma = 5 \), and \( P_o = (0.1, 0.6, 0.15, 0.15) \).
Figure 9–Figure supplement 1. Each line corresponds to a trajectory of: A performance, and B probing for individual subjects.