Stable choice coding during changes of mind

J Tyler Boyd-Meredith[1,*], Alex T Piet[1,2,*], Emily Jane Dennis[1], Ahmed El Hady[1,+,*], Carlos D Brody[1,3,4,+]

1. Princeton Neuroscience Institute, Princeton University, Princeton, United States.
2. Allen Institute, Seattle, Washington, United States.
3. Howard Hughes Medical Institute, Princeton University, Princeton, United States.
4. Department of Molecular Biology, Princeton University, Princeton, United States.

* equal contribution
+ equally contributing senior authors

correspondence should be addressed to:
Carlos D Brody (brody@princeton.edu)
Ahmed El Hady (ahady@princeton.edu)

Abstract

How do we choose the best action in a constantly-changing environment? Many natural decisions unfold in dynamic environments where newer observations carry better information about the present state of the world. Recent work has shown that rats can learn to optimally discount old evidence, updating their provisional decision when the environmental state changes. Provisional decisions are thought to be represented in the Frontal Orienting Fields (FOF), but this has only been tested in static environments where the provisional and final decisions are not easily dissociated. Here, we characterize the representation of accumulated evidence in rat FOF during decision-making in a dynamic environment. We find that FOF encodes evidence throughout decision formation with a temporal gain modulation that rises until the period when the animal may need to act. Using a behavioral model to predict the timing of changes of mind revealed that FOF neurons respond rapidly to these events, representing the new provisional decisions in their firing rates. Our results suggest that the FOF represents provisional decisions even in dynamic, uncertain environments, allowing for rapid motor execution when it is time to act.

Introduction

When making decisions, animals must weigh and combine the available evidence in favor of each alternative. With each new observation, evidence about the underlying state of the environment gradually accumulates until the animal is ready to act. This accumulation model successfully describes a wide array of decisions[1,2,3,4]. Neural correlates of this accumulation process are also present across many brain regions in animals performing perceptual categorization tasks[1,4]. Delivering perceptual evidence in streams of discrete pulses, like randomly-timed auditory clicks, provides additional power to estimate the evolution of each subject's latent accumulated evidence variable on individual trials[5]. Accurate prediction of the decision variable during each trial increases the resolution for estimating neural encoding across brain regions[6,7,8]. Hanks et al.[6] characterized the neural representation of accumulating evidence in rats performing accumulation of trains of auditory click evidence. In the task, two streams of randomly-timed auditory clicks were emitted from either side of a fixation location and rats were trained to orient toward the side that played a greater number of clicks. Experimenters recorded from the frontal orienting fields (FOF), a frontal cortical structure implicated in short term memory and preparation of orienting movements[9,10,11]. They proposed that FOF neurons used a single code throughout accumulation, which represented the accumulated evidence value categorically, providing a readout of the animal's provisional decision[6]. However, perturbations of this signal in FOF only impaired the animal's choice when they overlapped with the final time points of accumulation. These results, along with a two-node model of the FOF[12], indicate that the FOF is critical for maintaining the decision variable after it has been transformed into a categorical representation of the animal's choice.

While these experiments were conducted using stationary environments, many natural decisions unfold in dynamic environments. In stationary settings, all evidence samples in a trial reflect the same underlying environmental state. This means the best strategy is to equally weigh all samples of evidence throughout stimulus presentation[13]. In this regime it is difficult to dissociate the provisional from the final decision. In dynamic environments, the state of the world can change while the animal is deliberating. This means the animal should learn to discount old evidence via leaky integration, weighing more recently presented samples of information more heavily than older samples[14,15,16,17,18]. Unlike stationary environments, adopting the optimal strategy in a dynamic environment leads to frequent fluctuations in the animal's provisional decision, or changes of mind.

Recent work has shown that rats and humans can learn to adopt the optimal discounting rate in a dynamic environment[14,16]. However, it is unknown whether neural correlates of accumulated evidence observed in animals performing non-leaky integration in stationary environments persist in animals performing leaky integration in dynamic environments. Here, we recorded from FOF in rats during a dynamic accumulation of evidence task. We tested whether the stable code observed in the stationary environment persisted in the dynamic environment by applying and extending a method developed to characterize neural tuning to accumulated evidence[6]. Evidence tuning in FOF was described by a single sigmoidal tuning curve multiplied by a time varying gain modulation, which increased with time early in the trial and stabilized at the time of the earliest possible go cue. We reasoned that if FOF neurons track the accumulated evidence throughout the entire accumulation period, firing rates should respond rapidly to changes in the provisional decision. Using the behavioral model to predict the timing of changes of mind, we find that FOF neurons respond within 100ms, reflecting the new provisional decision in their activity. By recomputing the evidence tuning curves aligning to these change of mind events, we confirmed that FOF neurons encode evidence with a single tuning curve before and after changes of mind. These results suggest that FOF maintains a stable representation of accumulated evidence despite dynamic uncertainty in the environment. The time varying gain modulation may help ensure that the animal is ready when it is time to act. Our study opens up the opportunity for future work on the neural circuit level understanding of how animals integrate and decide in a volatile environment.

Results

The dynamic evidence accumulation task. We trained rats (n=5) to perform a previously developed dynamic evidence accumulation task[16]. This task requires the rat to report which of two hidden states the environment is in at the time of a go cue. At the beginning of each trial, the center port in an array of three nose ports is illuminated by an LED. This invites the rat to poke its nose into the center port, initiating presentation of an auditory stimulus. The stimulus is composed of two trains of auditory pulses (clicks) delivered in stereo from speakers positioned on either side of the center port. The left and right click trains are generated from different Poisson processes with rate parameters, r_R^i and r_L^i, that depend on the state i. When the environment is in state 1, the 'Go Right' state, the generative click rate is higher for the right speaker than the left ($r_R^1 = 38Hz$ and $r_L^1 = 2Hz$). In state 2, the 'Go Left' state, the click rates are reversed ($r_R^2 = 2Hz$ and $r_L^2 = 38Hz$). Trials begin in either state with equal probability and switch stochastically between states with a fixed hazard rate $h = 1Hz$. After a randomized duration, drawn from a uniform distribution between 500 and 2000 ms, the stimulus ends and the center LED turns off. This 'go' cue signals the rat to withdraw from the center port and poke its nose into one of two reward delivery ports on either side. The animal receives a drop of water (18 uL) if it chooses the side port corresponding to the final value of the hidden state. Incorrect choices were signaled with a white noise stimulus (Fig. 1A). In our dataset, roughly 33% of trials had no state changes, 33% had one, and 34% had more than one (Fig. 1C).

Behavioral model captures leaky integration strategy. We fit a previously-developed behavioral model[5,16] to rats' choices using an average of 108,126 trials per rat (63,494 to 185,091 trials each from 118 to 308 sessions). The model parameterizes the process by which the evidence available in each auditory click is integrated to produce a decision. At the start of each trial, time $t = 0$, this distribution has zero mean and initial variance σ_i^2. Each right and left click, $\delta_{R,i}$ and $\delta_{L,i}$, increments or decrements the accumulation value a, subject to sensory adaptation governed by parameters ϕ and τ_ϕ. Each click also introduces additional noise with variance σ_s^2. Memory noise with variance σ_a^2 is introduced at each time step. Evidence is discounted with rate λ which parameterizes the rate at which, in the absence of further input, a decays with time ($\lambda < 0$) or increases with time ($\lambda > 0$). When $\lambda < 0$, older pieces of evidence are discounted relative to newer evidence. While decision makers in stationary environments perform best when discounting is minimal ($\lambda = 0$), ideal observers in our task adopt a high-level of discounting of old evidence ($\lambda < 0$), reducing the impact of older clicks that may have been presented before a change in the hidden state[16,15,14,17]. As previously described[16], the optimal discounting rate in a dynamic environment depends on the quality of evidence, including the observer's per-click noise (σ_S^2). At the end of the trial, time $t = t_N$, the rat chooses to go right if the accumulation value a is greater than the decision boundary B, except on a fraction of "lapse" trials where, with some probability l, the rat chooses randomly (Fig. 1C) (see methods for mathematical details). For each rat, we found the best-fitting parameter set θ to describe this process using maximum likelihood estimation. These parameter sets were used these to estimate the probability distribution over accumulation values on each trial $P(a|t, \delta_R, \delta_L, \theta)$.

We present several assessments of task performance and model validation. Psychometric curves show a rat's choices as a function of the ideal observer log-odds favoring a rightward choice, as well as the correspondence with predictions from the behavioral model fit to an example rat (Fig. 1D) and all rats used in this study (Fig. S2). Final state chronometric curves show that performance was dependent on the final state duration, the elapsed time between the final state change and the 'go' cue (Fig. 1E and Fig. S3). Radillo et al.[17] demonstrated the

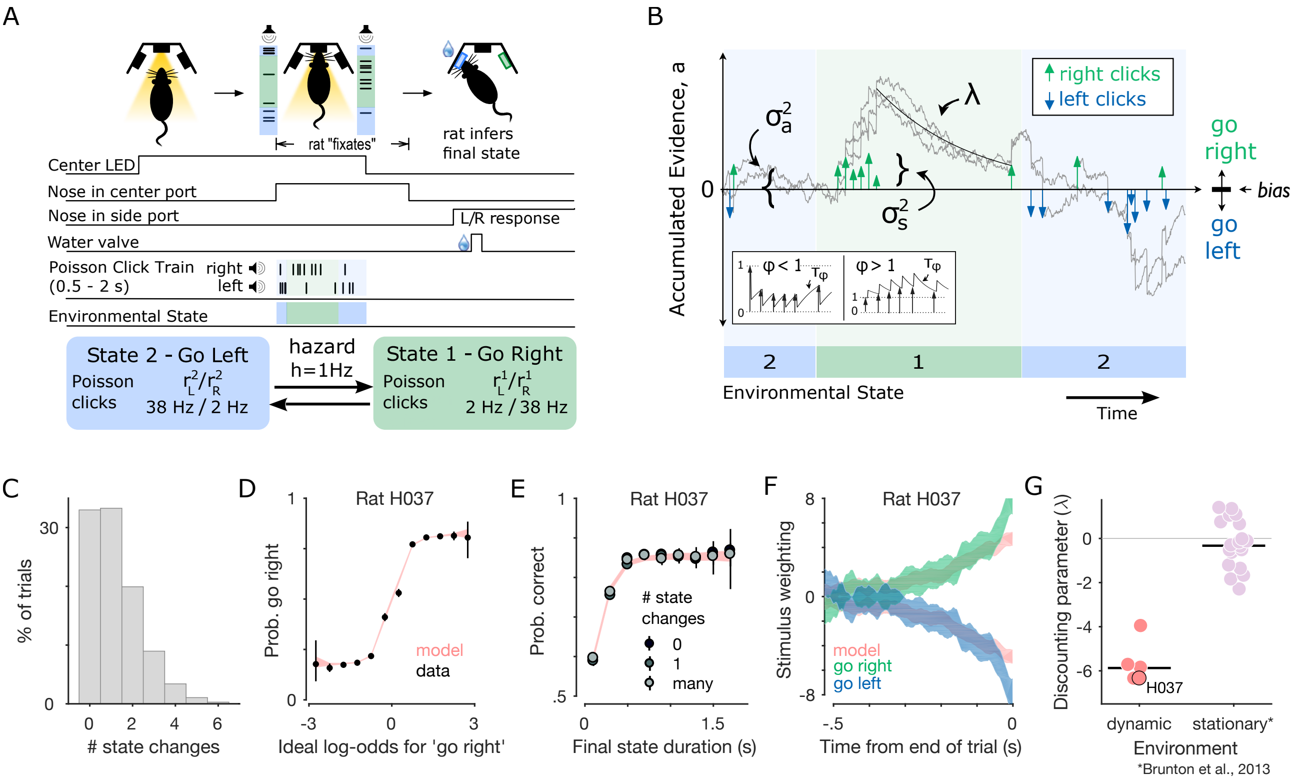

Figure 1: **Rats accumulate and discount evidence in a dynamic accumulation task** (A) Schematic showing task events and timing. The center port is illuminated by an LED. The rat pokes its nose into the port to initiate playback of randomly timed auditory clicks from speakers on either side. Clicks on each side are generated with different underlying Poisson rate parameters that depend on a hidden environmental state. The stimulus duration is drawn from a uniform distribution between 500 and 2000ms. During that time the hidden state switches stochastically at a fixed hazard rate, $h = 1Hz$. At the end of the stimulus presentation, the center LED turns off and reward is baited in the side port corresponding to the final state. (B) Schematic of the evolution of the accumulation model on an example trial. Three example accumulation traces are shown for different instantiations of the noise applied at each time point (σ_a) and the noise applied to each click (σ_s). Neighboring clicks can either depress or facilitate each other according to the adaptation parameters (ϕ and τ_ϕ). The evidence discounting rate (λ) determines how quickly the decision variable a decays back to zero. At the end of the trial, a choice is made by comparing the decision variable to the bias parameter. (C) Frequency of state changes per trial across all rats' datasets. (D) Example psychometric plot showing the probability that the rat chooses 'go right' as a function of the ideal observer log-odds supporting a 'go right' choice. Rat data (black points) is overlaid on predictions of the accumulation model with parameters fit to this rat (red traces). Errorbars for rat data represent 95% binomial confidence intervals around the mean. (E) Example final state chronometric plot showing accuracy (mean with 95% binomial confidence intervals) as a function of the duration of a trial's final state and the number of switches in a given trial. (F) Psychophysical reverse correlation kernel for an example rat. Green and blue patches indicate strength (mean ± s.d.) of evidence favoring rightward choice as a function of time until the trial ends for rightward and leftward choices, respectively. The red patches are corresponding predictions from the accumulation model. (G) Discounting parameters for each rat in this study (red points) compared to each rat in a previously published stationary environment (lilac points; Brunton et al.[5]). Group medians are plotted as black horizontal lines.

rate of increase and saturation level of the chronometric curve for an ideal observer depends only on the hazard rate and SNR of the click rates. Psychophysical reverse correlation kernels quantify the influence of clicks at each timepoint throughout the stimulus period, providing an assay of the rats' evidence discounting that is independent of the behavioral model. Reverse correlations for all rats in this study show heavier weighting of clicks presented at the end of the trial compared to the beginning (Fig. 1F and Fig. S4).

The behavioral model parameter fits for each rat confirm that all rats used a leaky integration strategy ($\lambda < 0$). Best fit discounting parameters were significantly different from a previously reported[5] dataset of rats integrating in a stationary environment ($p < .01$; two-tailed Wilcoxon rank-sum test, n=5 in dynamic and n=19 stationary environments) (Fig. 1G). Consistent with previous work, rats adopted discounting rates that favor more recent evidence due to the environmental volatility[16].

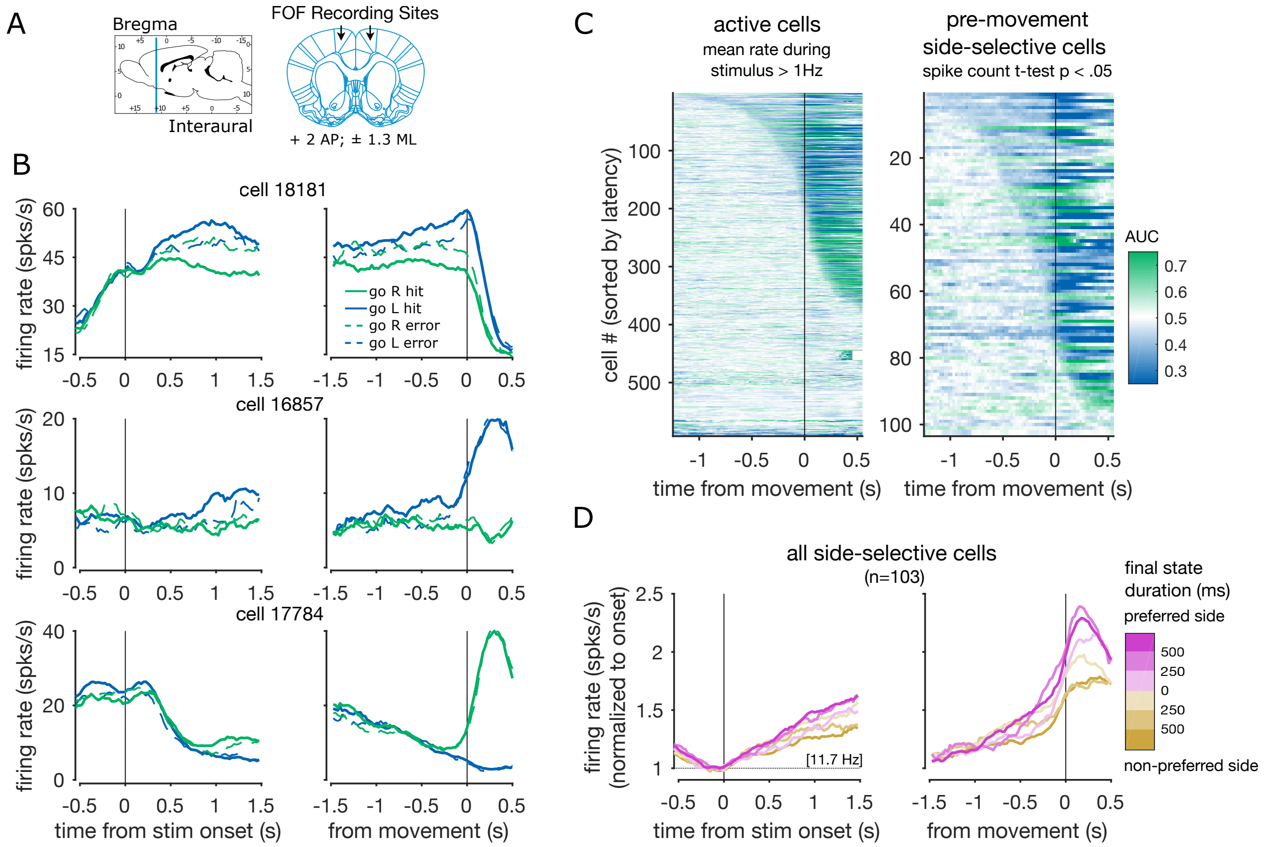

Figure 2: **FOF neurons encode the rat's upcoming choice** (A) Coordinates used for FOF recordings (+2 AP; ± 1.3 ML). (B) Average firing rates for three example FOF cells aligned to stimulus onset (left) and movement (right). Activity is conditioned on right (green) vs. left (blue) side choice, as well as hits (solid lines) vs. errors (dashed lines). (C) Side-selectivity at each time point relative to movement for all active cells (left; firing rate > 1 Hz) and for the subset of these cells that meet the spike count pre-movement side-selectivity criterion (right; 2-tailed t-test $p < .05$). AUC is computed on spike rates for right versus left choices. Plots are sorted by latency to 200ms (8 consecutive time bins) of significant AUC values relative to a distribution created by permuting choice labels across trials (2-tailed permutation test, 250 permutations, $p < .05$). (D) Average activity of all pre-movement side-selective cells conditioned on final state duration and cells' side preferences. Grand-average firing rate at stimulus onset (11.7Hz) is written in brackets.

FOF responses during dynamic accumulation. We recorded from the frontal orienting fields (FOF) of rats performing the dynamic evidence accumulation task. In 4 rats, we im-

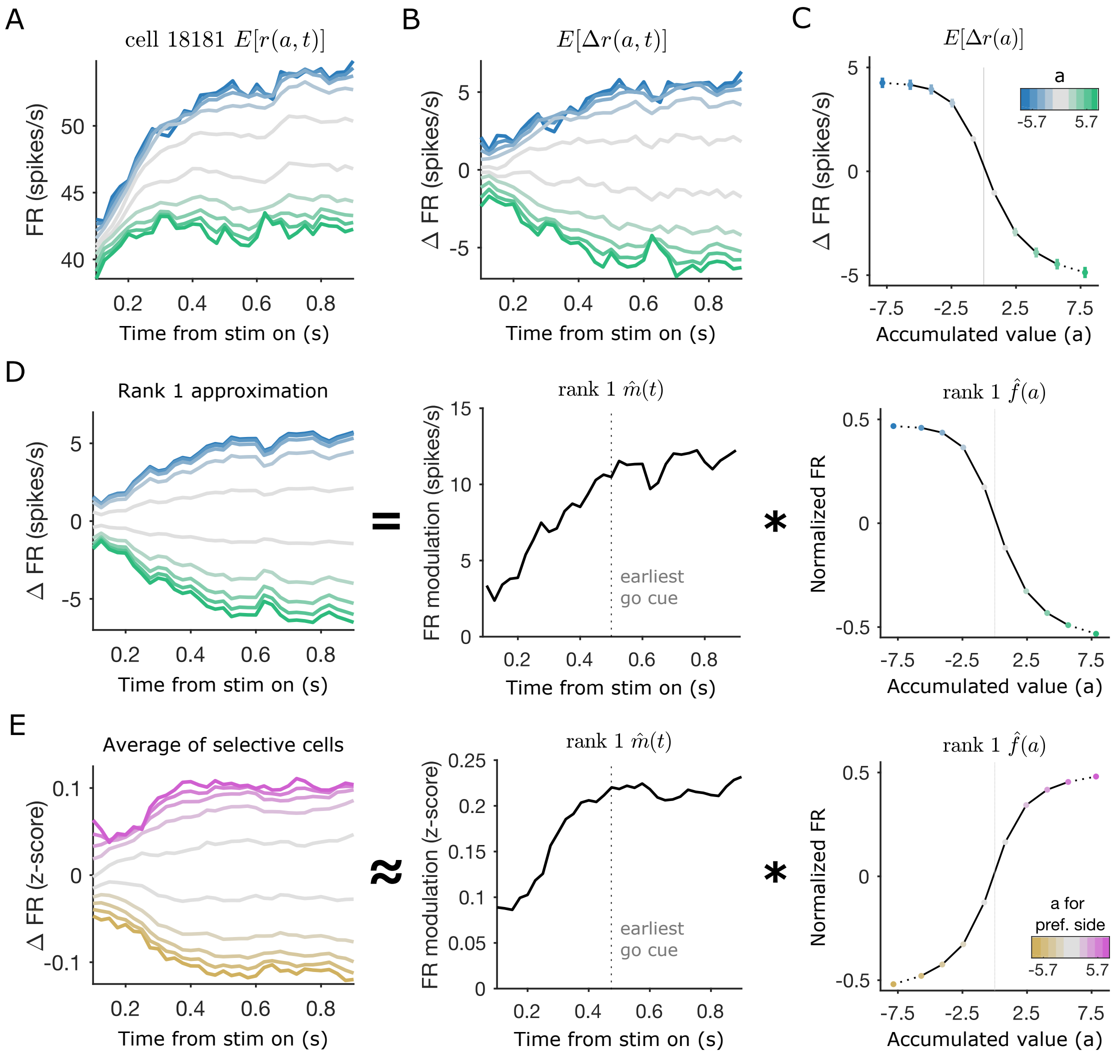

Figure 3: **FOF neurons encode the accumulated evidence throughout the trial** (A) Firing rate map as a function of accumulated evidence and time for an example neuron. Colors indicate accumulated evidence value with the same colors as in B and C. (B) Residual rate map in which the mean temporal trajectory is subtracted. (C) Average tuning over time. Points indicate mean (± s.e.m.) across time of the change in firing rate relative to temporal average as a function of accumulated evidence value a. (D) Rank 1 approximation of the residual rate map $E[\Delta r(a,t)]$ from B. The approximation (left) is equal to the outer product of a modulation over time $\hat{m}(t)$ (middle) and a tuning curve $\hat{r}(a)$ (right). (E) Average residual z-scored firing rate map (left). This plot is produced by averaging over the residual z-scored firing rate map of all pre-movement side-selective cells. This map is approximated by the outer product of a modulation curve (middle) and a tuning curve (right).

planted unilateral (n=2 left FOF, 2 right FOF) microwire arrays at coordinates (+2 AP; ± 1.3 ML) (Fig. 2A). In a 5th rat, we implanted a bilateral tetrode drive over the same coordinates. Recordings from 69 sessions yielded 738 units. See Supplementary Table 1 for a breakdown of data by rat (Method, location). Cells were considered active and included for further analysis if they had a mean firing rate of at least 1 Hz during the trial (n = 592 active cells).

Individual cells show stereotyped temporal dynamics aligned to both the onset of the trial (entering the center nose port), and the movement following the end of the stimulus (nose out of center port) (Fig. 2B). Many individual cells diverged throughout the trial based on the animal's upcoming choice. We tested the timecourse of selectivity for single neurons to right versus left choices by computing the area under the receiver operating characteristic curve (AUC) and comparing it to a permutation distribution computed by shuffling choice labels across trials. For purposes of visualization, cells are sorted by latency to 200ms (8 consecutive time bins) of significant AUC values (2-tailed permutation test, 250 permutations, $p < .05$). We present these plots for all active neurons and for a subset of pre-movement side-selective neurons (Fig. 2C). This subset made up 17% of the active population (n=103 selective) and was defined as cells with spike counts that differed significantly as a function of upcoming animal choice during the period between the start of the stimulus and the movement away from the fixation port (2-tailed t-test, $p < .05$). For each neuron, the side associated with the higher spike count is referred to as the cell's preferred side. Following Hanks et al.[6], we focus subsequent analyses on these pre-movement side-selective cells. Pre-movement side selectivity was slightly less common in this dataset than in previous studies of FOF in stationary environments[6]. This may be a consequence of frequent changes of mind, which create a dissociation between provisional and final choice throughout trials in the dynamic task. Across pre-movement side-selective neurons, we computed the average activity conditioned on final state duration and cell preference (Fig. 2D). We observe divergences at different latencies depending on the final state duration.

Stable accumulator tuning in dynamic environment. The choice-selectivity metrics presented above reveal coding of the final choice in average neural activity. However, during the trial, the hidden state can change multiple times (1.22 ± 1.20 state changes per trial). This creates dissociations during the trial between the animal's provisional choice and the final choice. To better describe tuning to the provisional decision throughout the trial, we applied and extended a method developed to quantify the tuning of single neurons to the accumulated value throughout the trial. Using this method, Hanks et al.[6] found that FOF neurons had a stable encoding of evidence throughout accumulation in a stationary environment. Here, we sought to test whether the FOF continues to encode the evidence throughout trials when the environment is dynamic. Further, we asked whether this encoding can still be captured by a single tuning curve in the dynamic environment.

We used the approach described by Hanks et al.[6], to produce an evidence tuning map of each neuron. First, we computed the joint distribution $P(r, a, t)$ of each cell's firing rate r, the instantaneous accumulation value a, and time in the trial t. The distribution over a, from the behavioral model described above, was further constrained using the animal's choice y on each trial, giving the posterior distribution $P(a|t, \delta_R, \delta_L, \theta, y)$. We improve on the method used by Hanks et al.[6] by using an analytical computation of the posterior distribution of accumulated evidence, allowing for more accurate estimation of $P(r, a, t)$ (see methods). Marginalizing the joint distribution with respect to firing rate provides an estimate of the expected firing rate as a function of accumulated evidence and time, for each cell $E[r(a, t)]$. We present this rate map for an example cell which is strongly tuned to the accumulator throughout the trial, firing more when accumulated evidence favors left choices (Fig. 3A). Because our neurons have stereotyped temporal dynamics aligned to stimulus onset, we subtract out the average temporal dynamics to get the expected firing rate modulation $E[\Delta r(a, t)] = E[r(a, t)] - E[r(t)]$ (Fig. 3B). Following Hanks et al.[6], a summary tuning curve was computed by averaging over time to get $E[\Delta r(a)]$ (Fig. 3C).

We extend the method by computing the rank 1 approximation of the residual firing rate map $E[\Delta r(a, t)]$ using the singular value decomposition (Fig. 3D). For the example cell, this approximation captures 99.6% of the variance in the estimated residual firing rate map. The mean variance explained by this approximation for all pre-movement side-selective cells was 89.7% ± 9.8%. The approximation is equal to the outer product of the first left singular vector u_1 and the first right singular vector v_1, scaled by the first singular value s_1. These terms can be rearranged and interpreted as the outer product of a firing rate modulation, $\hat{m}(t) = u_1 s_1 \text{range}(v_1)$ and a tuning curve $\hat{f}(a) = v_1/\text{range}(v_1)$. Scaling by $\text{range}(v_1)$ gives $\hat{f}(a)$ unit scale and $\hat{m}(t)$

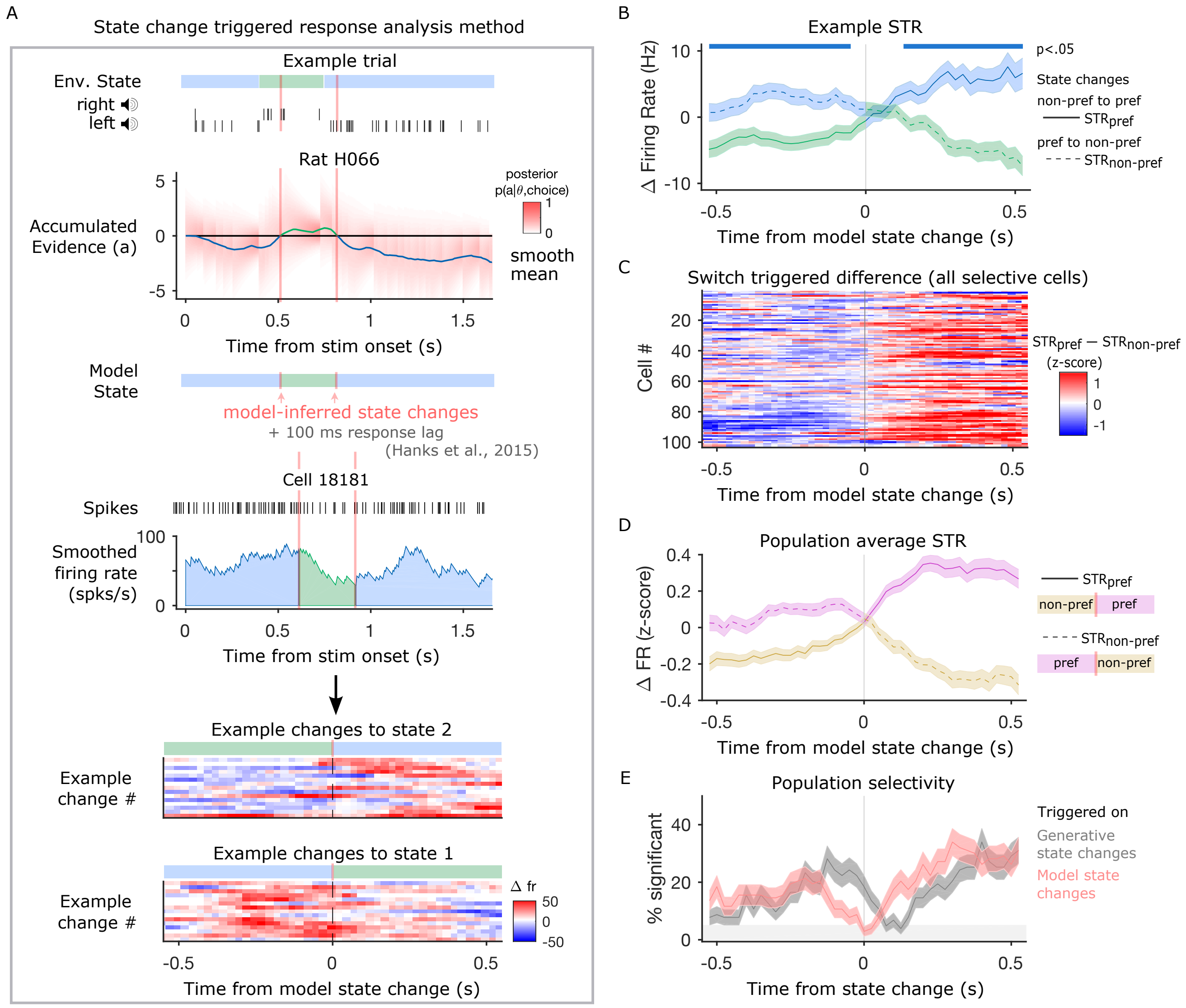

Figure 4: **FOF neurons track changes in the provisional decision** (A) Schematic explaining method used to compute state change triggered responses (STR). A given trial has a hidden environmental state (blue and green bar) used to generate click trains from each speaker. We compute the posterior distribution of accumulated evidence $P(a|\theta, \delta_R, \delta_L, y)$ (labeled $P(a|\theta, choice)$). We find time points where the smoothed posterior mean crosses the decision boundary and label these model-inferred state changes. We then select the residual smoothed firing rates from the 550ms before and after each state change and average together the residual responses for changes into state 1 and changes into state 2. (B) STR (mean ± s.e.m.) for the example cell used in panel A. Significance bars indicate time points when d' for discriminating model state is different from chance (2-tailed permutation test, 250 permutations, $p < .05$). The trace showing changes into the cell's preferred state (state 2 for this cell) is labeled $\mathrm{STR}_{\mathrm{pref}}$ (solid line) and the trace for changes into the cell's non-preferred state is labeled $\mathrm{STR}_{\mathrm{non\text{-}pref}}$ (dashed line). (C) Heat map showing difference between responses for changes into the preferred and non-preferred state ($\mathrm{STR}_{\mathrm{pref}} - \mathrm{STR}_{\mathrm{non\text{-}pref}}$) for each of the pre-movement side-selective cells. (D) Average z-scored STR (mean ± s.e.m.) across all pre-movement side-selective cells for state changes into cells' preferred states and non-preferred states. (E) Percentage of included cells (mean ± s.e.m.) with significant encoding across time relative to model predicted state changes (red trace) and generative state changes (gray trace).

units of spikes/s. Our complete tuning curve approximation becomes:

$$r(a,t) \approx E[r(t)] + \hat{m}(t) * \hat{f}(a). \tag{1}$$

We computed a population average residual rate map across all pre-movement side-selective cells by computing the residual firing rate map $E[\Delta r(a,t)]$ for each cell using z-scored firing rates. The accumulated value axis was inverted for left choice preferring cells and then the residual firing rate maps were averaged together (Fig. 3E). We computed the rank 1 approximation of this population residual rate map. This approximation explained 99.7% of the variance in the population residual rate map (Fig. 3E middle, right). The population firing rate modulation curve $\hat{m}(t)$ rises for the first 500 milliseconds and then plateaus at it's maximum value. Therefore, the population tuning can be described as a single tuning curve whose modulation increases during the period of the trial before a 'go' cue is possible. The modulation stabilizes at its maximum value during the period in which the trial may end and the animal may need to report its decision. Despite the dynamic environment, and changing provisional choice, we find FOF neurons continue to stably encode the evidence with a single tuning curve throughout evidence accumulation.

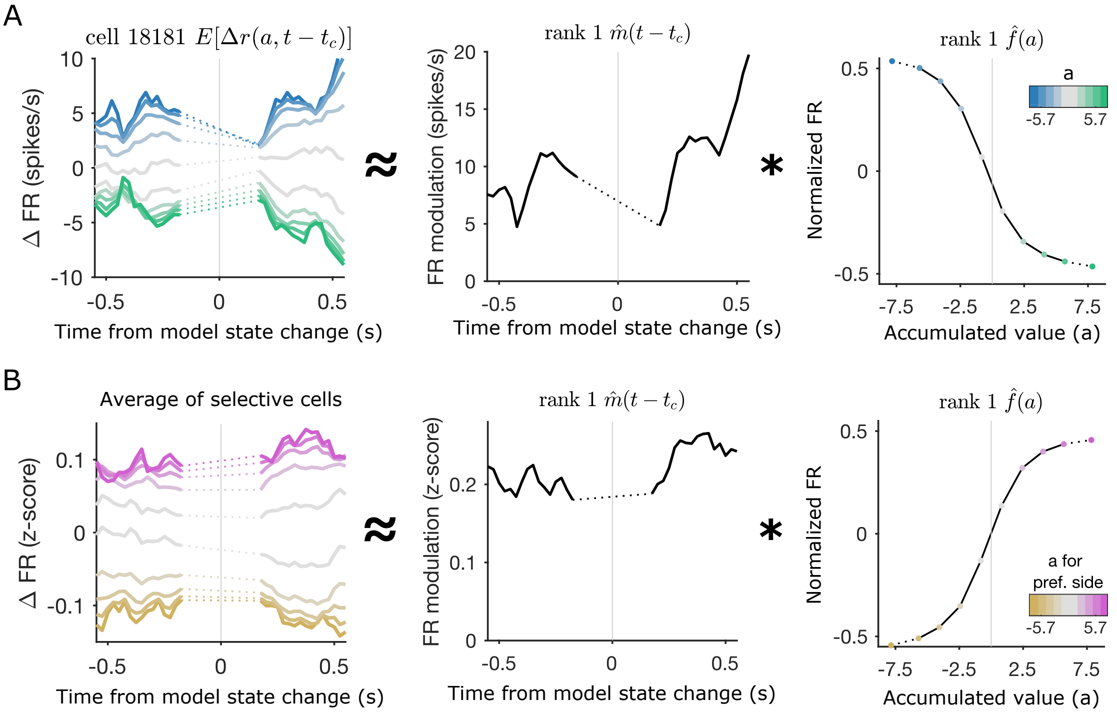

Figure 5: **State change triggered tuning curves show post state change modulation increase** (A) Example cell tuning map triggered on model-predicted state changes with rank 1 approximation derived temporal modulation and evidence tuning. Data is excluded from the 300ms around the state change where the accumulated value distribution is too narrow to estimate tuning (dotted lines). (B) Average of all pre-movement side-selective cells' tuning maps computed with z-scored firing rates and triggered on model-predicted state changes along with rank 1 approximation derived temporal modulation and evidence tuning for the average map.

Neurons track changes in provisional decision. If cells are stably tuned to the accumulated evidence throughout deliberation, we should be able to see rapid responses in their firing rates to changes in the animal's provisional decision. To look at responses to these changes of mind, we computed each cell's average deviation from its mean temporal trajectory aligned to time points when the behavioral model predicted a change in the animal's estimate of the environmental state (Fig. 4A). Following Hanks et al.[6], we introduced a 100ms response lag between model-predictions and FOF responses. For this analysis, changes of mind were selected at time points when a 100ms running average of the posterior mean crossed the decision boundary ($a = 0$). To avoid introducing noise into this analysis, changes of mind in the first and last 200ms of the trial were excluded, as were state changes that immediately reversed to the previous state (see methods). For each cell, this method produced two state-change triggered response curves describing responses to changes into states 1 (STR_1) and 2 (STR_2). STRs are also referred to as $\mathrm{STR}_{\mathrm{pref}}$ and $\mathrm{STR}_{\mathrm{non\text{-}pref}}$ according to cells' previously determined side-preference. STRs are shown for an example neuron (Fig. 4B). Discriminability before and after model-predicted changes of mind was measured using d' and tested for significance by permuting the state-change labels across trials (2-tailed permutation test, 250 permutations, p<.05).

To visualize the state change triggered response across the neural population, each cell's response is summarized by computing the difference between the z-scored STR for state changes into the preferred state and into the non-preferred state ($\mathrm{STR}_{\mathrm{pref}} - \mathrm{STR}_{\mathrm{non\text{-}pref}}$). We present these data as a heat map for all pre-movement side-selective cells (Fig. 4C). The z-scored STRs were averaged across these cells to give the average state-change triggered response across the population (Fig. 4D). We apply the permutation procedure described above to each cell and compute the fraction of the included cells that significantly encode state at each timepoint relative to the state change (Fig. 4E). If cells are encoding the animal's provisional decision, we should expect them to take intermediate firing rates during both the state change into and out of the preferred state. This would mean that cells are not significantly discriminating between states at the time of the state change. Consistent with this, we find that the population reaches its minimum fraction of cells differentiating between states at the time of the model-predicted state change. We recomputed the timecourse of discriminability across cells triggered on changes in the veridical environmental state, rather than the model-predicted changes. When we do this, we find the dip in the fraction of cells significantly discriminating the state is delayed relative to the dip produced by triggering on model-predicted changes. This is consistent with the FOF tracking changes in the sign of accumulated evidence rather than simply responding to the instantaneous stimulus. At the level of individual cells and across the population, we see rapid responses to changes of mind providing further evidence that neurons track the animal's provisional decision throughout the accumulation process.

Stable evidence tuning before and after changes of mind. To further characterize cell tuning to accumulated evidence during changes of mind, we recomputed the tuning maps aligning time to model-predicted state changes instead of the start of the trial. The computation and rank 1 decomposition of the tuning curves proceeded in the same manner as before except time in each trial was aligned to changes of mind:

$$r(a, t - t_c) \approx E[r(t)] + \hat{m}(t - t_c) * \hat{f}(a) \tag{2}$$

where t_c is the timing of changes of mind. Consistent with the state-change triggered responses and previous tuning curve analysis, we see that tuning in example neurons and the population is well described by a single evidence tuning curve multiplied by a temporal modulation before and after state changes (Fig. 5A,B). The rank 1 approximation for the example cell presented in Figure 5A explains 98.7% of the variance in the tuning map and the average variance explained for all selective cells is 84.9% ± 9.7%. The population average across z-scored tuning maps for all pre-movement side-selective cells is also well-described by the rank 1 approximation, which captures 95% of the variance (Fig. 5B). This demonstrates that neurons encode the accumulated evidence with a single tuning curve even at the times when the hidden state and provisional decision fluctuate.

Discussion

We recorded neural activity from the frontal orienting fields (FOF) of rats performing a dynamic decision-making task designed to induce frequent changes of mind. In our study, rats integrated sequential pieces of information, discounting older evidence, to track changes in a volatile hidden state. FOF responses have been characterized previously during a similar task in a stationary environment where rats learn to equally weigh all evidence and changes of mind are rare[6]. This previous work revealed categorical encoding of population activity to the accumulated evidence, characterized by a single tuning curve throughout the trial. This suggested that FOF encoded the provisional decision during decision-making. However, in a stationary environment, the provisional decision rarely differs from the final choice meaning that preparatory activity could begin without needing to be reversed. In a dynamic environment, where changes of mind are frequent, it might be advantageous to suppress choice coding until the final decision is reached. It was not clear whether FOF would play a similar role in representing evidence during decision-making in a constantly-changing environment and while the provisional decisions were still highly flexible.

We found that FOF responses to accumulation in a dynamic environment were similar to FOF responses during accumulation in a stationary environment. First, a subpopulation of about 17% of active neurons showed significant side-selectivity during the pre-movement stimulus period. This was a smaller fraction than previously reported, but was an expected result of a task with more frequent stimulus-induced changes of mind. Using a method developed by Hanks et al. [6], we measured the encoding of the decision variable in single neurons and across the population. We improved this method by using a rank 1 approximation to explain the evidence-encoding component of neural firing rates as the product of a temporal modulation and an evidence tuning curve. The rank 1 approximation supported the description of FOF neurons with a single evidence tuning curve that was modulated over the trial. Across the population, we found that the temporal modulation increased until the timing of the earliest possible 'go' cue and then plateaus at a maximum modulation strength during the rest of the trial.

The dynamic nature of the task allowed measurements that are not possible in stationary tasks, where evidence is drawn from a single distribution during each decision, and changes of mind are rare. We used our behavioral model to predict the timing of changes in the provisional decision during evidence accumulation and measured the response to these changes of mind events in neural activity. If the neurons use a single evidence tuning curve throughout accumulation, we expect the neural firing rates to encode the provisional decision before and after changes of mind. Computing state change triggered responses for each neuron, showed that FOF cells rapidly responded to changes of mind, reflecting the new provisional decision in their firing rates. Critically, neurons encoded provisional decisions both before and after changes of mind, which implies that provisional decisions are encoded even when they differ from the final choice. Neuronal responses were better aligned to changes of mind predicted by the behavioral model than to changes in the true environmental state, suggesting that these responses were not simply reflecting a change in sensory experience. Combining this approach with the method for computing accumulated evidence tuning maps, we found, as described above, that the product of a single evidence tuning curve and temporal modulation was still sufficient to explain the evidence response across changes of mind (rank 1 approximation). We observed that, after the moment when 'go' cues could arrive, the temporal modulation of evidence tuning was, on average, stable. Together, our results demonstrate that FOF neurons encode the animal's provisional decision and respond rapidly, updating this representation following changes of mind.

Changes of mind are not unique to dynamic environments and can also occur during evidence accumulation in stationary environments. These events can occur during stimulus presentation due to noise in the decision making process and can be predicted from neural activity[19]. Changes of mind may also occur after the subject begins to execute their choice due to post-processing delays[20] or constraints placed on action[21]. Our work differs from these studies, in that we use an environment designed to induce changes of mind and ask how neurons respond to these model-predicted events. To our knowledge, only one other study[22] has examined neural responses to behaviorally predicted changes of mind during evidence accumulation in a dynamic environment, and ours is the only such study in an animal model.

Previous inactivation studies suggest that while FOF is critical for performing actions and reporting decisions, it is not necessary for the integration of evidence[6,23]. This is consistent with the FOF representing the evidence after categorization into a provisional choice[12]. Work in mouse anterior lateral motor (ALM), a comparable cortical region, shows that categorical signals in this region recover quickly following photoinhibition, suggesting categorical input from other brain regions[24]. In a recent study, Finkelstein et al.[25] found that ALM choice signals were robust to distractors delivered during a delay period after the typical evidence presentation period, suggesting local circuitry maintained the choice signal. Our study considered a similar brain structure operating in a regime where, rather than ignoring distractors, it needed to flexibly update provisional decisions in response to new information. These studies, along with recent modeling work[12,26], suggest a common role for the FOF and the ALM in maintaining choice signals that are either robust to or responsive to new information according to task demands.

The dynamic decision-making task offers a complementary approach to typical studies of evidence accumulation in static environments. Here, we showed that in constantly-changing environments FOF neurons encode provisional choices and respond rapidly to changes of mind predicted from our behavioral model. Our quantitative methods and behavioral paradigm will be useful tools for investigation of the brain circuitry supporting evidence accumulation and the decision-making process more generally.

Methods

Subjects Animal use procedures were approved by the Princeton University Institutional Animal Care and Use Committee and carried out in accordance with NIH standards. All subjects were adult male Long Evans rats (Vendor: Taconic and Harlan, USA). Rats were pair-housed prior to implantation with recording electrodes and single-housed subsequently. Rats were placed on a water restriction schedule to motivate them to perform the task for water rewards.

Behavioral training We trained rats on the dynamic clicks task[16] (Figure 1). Rats went through several stages of an automated training protocol. In the final stage of training, each trial began with the illumination of a center nose port by an LED light inside the port. This LED indicated that the rat could initiate a trial by placed its nose into the center port. Rats were required to keep their nose in the center port (nose fixation) until the light turned off as a 'go' signal. During center fixation, auditory cues were played indicating the current hidden state. The duration of the stimulus period was drawn from a uniform distribution between 500 and 2000ms. After the 'go' signal, rats were rewarded for entering the side port corresponding to the final value of the hidden state. The hidden state did not change after the 'go' cue. Correct choices were rewarded with 18 microliters of water. Incorrect choices were signaled by a white noise stimulus (spectral noise of 1 kHz for a 0.7 seconds duration). The rats were put on a controlled water schedule where they receive at least 3% of their weight every day. Rats trained each day in training session of around 120 minutes. Training sessions were included for analysis if the overall accuracy rate exceeded 70%, the center-fixation violation rate was below 25%, and the rat performed more than 50 trials. In order to prevent the rats from developing biases towards particular side ports an anti-biasing algorithm detected biases and probabilistically generated trials with the correct answer on the non-favored side.

Psychometric and chronometric curves Task performance was assessed using psychometric curves, chronometric curves and psychophysical reverse correlations. For all task performance plots, rat data was overlaid on predictions from the accumulation model described below. These predictions were made by using the probability of a right or correct choice on each trial given by the acummulation model in place of the actual choice observed.

Psychometric plots show the probability that the rat chose to go right as a function of the ideal observer log-odds supporting a 'go right' choice. Final state chronometric plots show the probability of a correct choice as a function of the final state duration, the elapsed time between the final hidden state change (or the beginning of the stimulus, if there are no state changes) and the end of the stimulus. Data is plotted separately for trials with 0, 1, or more than 1 state changes.

Psychophysical reverse correlation The computation of the reverse correlation curves was similar to methods previously reported[5,6,23]. An additional step was included, as in Piet et al.[16], to deal with the changing hidden state. First,the click trains on each trial were smoothed with a causal Gaussian filter $k(t)$. The left click train was subtracted from the right, creating one smooth click rate for each trial. The filter had a standard deviation of 5 msec.

$$r_i(t) = \delta_{R,t} * k(t) - \delta_{L,t} * k(t) \tag{3}$$

Then, the smoothed click rate on each trial was normalized by the expected click rate for that time step, given the current state of the environment. This gives us the deviation (the excess click rate) from the expected click rate for each trial.

$$e_i(t) = r_i(t) - \langle r(t)|S_i(t)\rangle \tag{4}$$

Finally, we compute the choice triggered average of the excess click rate by averaging over trials based on the rat's choice.

$$\text{excess-rate}(t|\text{choice}) = \langle e(t)|\text{choice}\rangle \tag{5}$$

The excess rate curves were then normalized to integrate to one. This was done to remove distorting effects of a lapse rate, as well to make the curves more interpretable by putting the units into effective weight of each click on choice.

Accumulation Model The accumulation model characterizes the decision-making process as the evolution over time t of an accumulation value a in response to left and right click trains, δ_L and δ_R, with dynamics governed by a parameter set θ. Each rat's behavioral data is used to find the parameter set that maximizes the probability under the model of the rat's choices y. Evaluating this model with the best fit parameters produces a probability distribution over values of a at every timepoint in the trial. We refer to this as the forward model distribution $P(a|t, \delta_R, \delta_L, \theta)$. The forward model was described previously in Piet et al.[16] and will be reviewed in detail below. To characterize neural encoding of the accumulation value, we further constrained the accumulation value distribution on trials where we had simultaneous neural recordings by incorporating the rat's choice, y, to find the posterior distribution $P(a|t, \delta_R, \delta_L, \theta, y)$. To do this, we computed a distribution that we refer to as the backward model distribution, which we describe in the next section.

At each moment in the trial, the forward model $P(a|t, \delta_R, \delta_L, \theta)$ predicts a Gaussian distribution of accumulation values with mean $\mu(t)$ and variance $\sigma^2(t)$. At start of the trial, time $t = 0$, this distribution has zero mean and initial variance σ_i^2. At all timepoints, the mean and variance are given by:

$$\begin{aligned}\mu(t) &= \mu_0 e^{\lambda t} + \int\limits_0^t \left(\delta_{R,s} \cdot C\left(R(s)\right) - \delta_{L,s} \cdot C\left(L(s)\right)\right) ds \\ &= \sum_i^{\#R_t} e^{\lambda(t-R(i))} C(R(i)) - \sum_i^{\#L_t} e^{\lambda(t-L(i))} C(L(i)) \qquad (6) \\ \sigma^2(t) &= \sigma_i^2 e^{2\lambda t} + \frac{\sigma_a^2}{2\lambda}\left(e^{2\lambda t} - 1\right) + \int\limits_0^t \sigma_s^2 \left(\delta_{R,s} \cdot C\left(R(s)\right) - \delta_{L,s} \cdot C\left(L(s)\right)\right) e^{2\lambda t} ds \\ &= \sigma_i^2 e^{2\lambda t} + \frac{\sigma_a^2}{2\lambda}\left(e^{2\lambda t} - 1\right) + \sum_i^{\#R_t} \sigma_s^2 C(R(i)) e^{2\lambda(t-R(i))} + \sum_i^{\#L_t} \sigma_s^2 C(L(i)) e^{2\lambda(t-L(i))} \qquad (7)\end{aligned}$$

Where $\#R_t$ is the number of right clicks on this trial up to time t, $R(i)$ is the time of the i^{th} right click and $\delta_{R,i}$ is a delta function at time $R(i)$. $C(R(i))$ tells us the effective adaptation for that click.

The model parameters θ can be described in words as an initial noise variance σ_i^2, a per-click noise variance σ_s^2, a memory noise variance σ_a^2, a discounting rate λ, the strength and time constant of adaptation ϕ and τ_ϕ, a bias B and a lapse rate l.

To determine the probability of a right versus left choice, we first integrate the accumulation value distribution in the last timepoint of the trial $P(a|t = t_N, \delta_R, \delta_L, \theta) = \mathcal{N}\left(\mu(t_N), \sigma^2(t_N)\right)$ from the bias parameter B to $+\infty$

$$P(a > B|t = t_N, \delta_R, \delta_L, \theta) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{-\left(B - \mu(t_N)\right)}{\sigma(t_N)\sqrt{2}}\right)\right). \qquad (8)$$

On each trial, the rat makes a random choice with probability determined by lapse rate l. Then, the probability of a 'go right' choice is given by

$$\begin{aligned}P\left(y^i = 1|\theta\right) &= (1-l) P(a > B|t = t_N^i, \delta_R^i, \delta_L^i, \theta) + l/2 \qquad (9) \\ P(y^i = -1|\theta) &= (1-l)\left(1 - P\left(a > B|t = t_N^i, \delta_R^i, \delta_L^i, \theta\right)\right) + l/2 \qquad (10)\end{aligned}$$

Where

$$y^i = \begin{cases} 1, & \text{if rat chooses right on trial } i \\ -1, & \text{if rat chooses left on trial } i \end{cases} \qquad (11)$$

Parameters θ were fit to each rat individually by maximizing the likelihood function:

$$L = \prod_i^{\#\text{trials}} P(y^i|\theta). \qquad (12)$$

A half-gaussian prior was included on the initial noise σ_i and accumulation noise parameters σ_a. The priors were set to match the respective best fit values from Brunton et al.[5]. The numerical optimization was performed in MATLAB, using the function `fmincon`. To estimate the uncertainty on the parameter estimates, we used the inverse hessian matrix as a parameter covariance matrix[27]. To compute the hessian of the model, we performed automatic differentiation in *julia* to exactly compute the local curvature[28]. See the Supplemental Materials for parameter estimates and uncertainty values. Brunton et al.[5] extensively analyzed how well a similar model with an additional bound parameter recovers generative parameters, finding the model contains one maximum likelihood point in parameter space (See Section 2.3.3-6 of the supplement to Brunton et al.[5]). We compared parameter fits in this task to those reported in Brunton et al.[5], which developed the stationary version of this task.

Posterior model The forward model gives us a probability distribution over accumulation values at each time point in a trial as well as an estimated probability of the rat choosing to go right or left on that trial. Observing the rat's choice y at the end of each trial allows us to constrain the distribution of possible trajectories that the accumulation value could have taken. The resulting posterior distribution (previously called the backward pass distribution in Brunton et al.[5]) is useful for analyzing the neural encoding of accumulated evidence.

The posterior distribution can be computed by taking the product of the forward model distribution and a backward distribution. While the forward distribution is constrained to the start the trial with mean 0 and variance σ_i^2, the backward distribution is constrained to finish the trial with the probability density uniformly distributed on one side of the decision boundary B according to the rat's choice. To distinguish between the two distributions and to make the constraints on the initial and final distributions explicit, we will write the forward distribution

$$f(a) = P(a|t, \delta_R, \delta_L, \theta, a_0 = \mathcal{N}(0, \sigma_i^2)) \tag{13}$$

and the backward distribution

$$b(a) = \begin{cases} P(a|t, \delta_R, \delta_L, \theta, a_N = \mathcal{U}(B, \infty)), & \text{if } y_i = 1 \\ P(a|t, \delta_R, \delta_L, \theta, a_N = \mathcal{U}(-\infty, B)), & \text{if } y_i = -1. \end{cases} \tag{14}$$

Importantly, the forward and backward distributions are conditionally independent, conditioned on the final value of the accumulated evidence. Given that they are independent, the posterior distribution that combines the initial and final constraints is the product of the forward and backward distributions.

$$P\left(a|t, \delta_R, \delta_L, \theta, a_0 = \mathcal{N}\left(0, \sigma_i^2\right), y\right) = f(a)b(a) \tag{15}$$

We approximated the backward distribution as a mixture distribution over a grid of final accumulation values with spacing Δa. A unit of probability mass is initialized at each point in the grid and the solution is given by:

$$b(a) = \sum_{i=0}^{\pm\infty} w_i P(a|t, \delta_R, \delta_L, \theta, a_N = \mathcal{N}\left(B + i\Delta a, 0\right)) \tag{16}$$

The mixture weights w_i are all equal if the bin spacing is uniform. Each unit of probability mass evolves using the same solution as the forward model, but with time reversed. This solution is exact as $\Delta a \to 0$.

For tuning curve analyses we use the full posterior distribution, for the state change triggered response analyses we use the mean of the posterior. See the Supplementary Materials for a detailed discussion on the derivation and evaluation of the backward and posterior model.

Microwire array recordings Microwire array implant surgery: Four rats were implanted with microwire arrays in their left or right FOF (n= 2 in lFOF,n = 2 in rFOF) The target region was accessed by craniotomy, using standard stereotaxic techniques (centred 2 mm anterior to the bregma and 1.3mm lateral to the midline). Dura mater was removed over the entire craniotomy with a small syringe needle. The remaining pia mater, even if not usually considered to be resistant to penetration, nevertheless presents a barrier to the entry of the microelectrode arrays because of the high-density arrangement of electrodes in the multi-channel electrode arrays. This dimpling phenomenon, when the electrodes are pushing the brain cortex down without penetrating, is more pronounced for arrays with larger numbers of electrodes. In addition to potentially injuring the brain tissue, dimpling is a source of error in the determination of depth measurements. Ideally, if dimpling could be eliminated, the electrodes would move in relation to the pial surface, allowing for effective and accurate electrode placement. To overcome the dimpling problem, we implemented the following procedure. After the craniotomy was made, and the dura was carefully removed over the entire craniotomy, a petroleum-based ointment (such as bacitracin ointment or sterile petroleum jelly (Puralube Vet Ointment)) was applied to the exact site of electrode implantation. The cyanoacrylate adhesive (Vetbond Tissue Adhesive) was then applied to the zone of the pia surrounding the penetration area. This procedure fastens the pia mater to the overlying bone and the resulting surface tension prevents the brain from compressing under the advancing electrodes. Once the polymerization of cyanoacrylate adhesive was complete, over a period of few minutes, the petroleum ointment at the target site was removed, and the 32-electrode microwire array (Tucker-Davis Technologies) was inserted by slowly advancing a Narishige hydraulic micromanipulator. After inserting the array(s), the remaining exposed cortex was covered with biocompatible silicone (kwik-sil), and the microwire array was secured to the skull with C&B Metabond and dental acrylic.

During a ten-day recovery period, rats had unlimited access to water and food. Recording sessions in the apparatus began thereafter, using Neuralynx acquisition systems. Once rats had recovered from surgery, recording sessions were performed in a behavioral chamber outfitted with a 32 channel recording system (Neuralynx). Spiking data was acquired using a bandpass filter between 600 and 6000 Hz and a spike detection threshold of 30 microV.

For array recordings, clusters were manually cut (Spikesort 3D, Neuralynx), and both single- and multi-units were considered.

Tetrode recordings Tetrode drives were 3d printed from custom designs (design files available upon request) on a Form2 3d printer in tough resin. Each drive consists of a drive body, a cone and cap to protect the drive body, and four bundles of 8 tetrodes in glass tubes. Each bundle was glued together and to a cannula. Each cannula was attached to a screw using dental cement, and cured with UV light. Each wire from each tetrode was fed through a unique channel in a 128 channel Electrode Interface Board (SpikeGadgets) and pinned with a gold pin. After loading all tetrodes, trimming, and building of the drive, the day-of or night before the surgery, we electroplated the drive in gold using a nanoZ impedence tester (White Matter LLC) and measured impedences.

Tetrode drive implant surgery proceeded as described for microwire arrays, except we did not need to vetbond the brain surface because each tetrode bundle produced very little dimpling. A silver wire and skull screw were used to ground the drive. Drives were secured with metabond and acrylic until secure. Tetrodes were advanced 0.1mm into the brain.

During a seven-day recovery period, animals had unlimited access to water and food. Animals were then returned to training and water restriction. To acclimate animals to the weight of the wireless apparatus, every three days, we replaced the cap on the implant with a new cap 3g heavier than the previous cap. If animals' behavioral performance or weight dropped, or if we noticed any excess tilting of the head from the weight, we returned the animal to the previous weight and waited an additional 2 days before moving to the next weight. This process was repeated until the animals were behaving well with caps weighing 27g.

Once animals were acclimated to the weight, recordings could begin. Tetrodes were advanced 0.25mm at a time, at least 20 hours before recording. For each recording session, the animal's cap was replaced with a 500mAh lithium battery, 128Gb Sandisk extreme plus SD card, a 160-pin Amphenol Lynx connector, and datalogger (SpikeGadgets). At the end of each session, the datalogger, SD card, and battery were removed and the 27g cap replaced.

The tetrode recordings were automatically clustered using Kilosort2[29]. Automatically determined clusters were manually curated using the Phy GUI (`https://github.com/kwikteam/phy`).

Electrophysiological analysis We computed the firing rates for all neurons aligned to the time of stimulus onset (when the rat first broke the center port IR beam triggering playback of the stimulus) and to movement (when the rat first stopped breaking the center port IR beam to make its choice). Firing rates were computed by binning spikes into 25ms bins and smoothing them with a casual gaussian filter with a standard deviation of 100ms. Stimulus onset aligned firing rates were masked on each trial after the movement and movement aligned firing rates were masked prior to stimlus onset. Firing rates for example cells were averaged over trials conditioned on choice and outcome.

Cells were considered active if their average stimulus onset aligned firing rate was greater than 1 Hz during the time from 1 second prior to the stimulus onset to the time of movement onset. Cells were considered pre-movement side-selective if the spike counts during the period between stimulus onset and movement were different on trials that resulted in a left versus a right choice (2-tailed t-test, $p < .05$). The side with the higher firing rate is referred to as the cell's preferred side.

A population-average PSTH was computed by averaging over all trials from all pre-movement side-selective cells conditioned on final state duration and whether the trial ended in a choice to the cell's preferred side.

We analyzed the timecourse of choice selectivity by computing the area under the receiver operating characteristic curve (AUC) at each 25ms time bin in for the smoothed firing rates in left choice versus right choice trials. To compute significance, we performed a permutation test where the left/right choice labels were permuted relative to the firing rates across trials. For visualization purposes, we sorted cells by latency to reach 8 significant 25ms time bins in a row (2-tailed permutation test, 250 permutations, $p < .05$).

Evidence tuning curves We compute evidence tuning curves using a method based on the one used in Hanks et al.[6]. First, the posterior accumulation distribution

$$P\left(a|t, \delta_R, \delta_L, \theta, a_0 = \mathcal{N}\left(0, \sigma_i^2\right), y\right) \tag{17}$$

for each trial is computed. To simplify notation, we will refer to this distribution as $P(a|\theta, y)$. The joint distribution of $P(a|\theta, y)$, the firing rate r, and time t, which we will call $P(r, a, t)$ is computed by binning time, accumulation values, and firing rates. For each trial and each timepoint, the probability mass in each accumulation value bin in $P(a|\theta, y)$ is added to the bin in $P(r, a, t)$ associated with that timepoint and that firing rate. Because the shortest trials are 500ms, not all trials contribute to each time point, so each time bin is normalized according to the number of trials that contribute to it. Each neuron's firing rate was divided into 100 bins spanning the minimum to the maximum firing rate of the neuron. Time was binned into 25ms bins. Accumulation value bin size was divided into 10 bins with width set to 1.625 except the last bin which was larger to capture the tails of the distribution.

The expected firing rate as a function of accumulation value a and time t is computed by marginalizing with respect to the firing rate $E[r(a,t)] = \Sigma_i r_i P(r_i, a, t)$. The expected difference from the average firing rate is computed by computing the average firing rate as a function of time $E[r|t] = \Sigma_j r_i \Sigma_j P(r_i, a_j, t)$ and subtracting this from average expected firing rate at each timepoint $E[\Delta r|a, t] = E[r|a, t] - E[r|t]$. Average accumulation value tuning over the trial is computed by averaging over time $E[\Delta r|a] = \frac{1}{N}\Sigma_i E[\Delta r|a, t_i]$ where N is the number of time bins. The same procedure is used to compute a map of z-scored firing rates. And these maps are averaged across pre-movement side-selective cells to produce a population average.

Rank 1 approximations of the $E[\Delta r|a,t]$ are computing using the singular value decomposition. The approximation is equal to the outer product of the first left singular vector u_1 and the first right singular vector v_1, scaled by the first singular value s_1. These terms are rearranged to give the outer product of a firing rate modulation, $\hat{m}(t) = u_1 s_1 \text{range}(v_1)$ and a tuning curve $\hat{f}(a) = v_1/\text{range}(v_1)$. Scaling by range$(v_1)$ gives $\hat{f}(a)$ unit scale and $\hat{m}(t)$ units of spikes/s. Our complete tuning curve approximation becomes:

$$r(a,t) \approx E[r(t)] + \hat{m}(t) * \hat{f}(a). \tag{18}$$

The variance explained by this approximation is given by the ratio between the first singular value and the sum of all singular values:

$$\frac{s_1}{\sum_{i=1}^{N} s_i}. \tag{19}$$

State change triggered response On each trial, we computed a residual firing rate by subtracting the average firing rate at each time point during the trial. We then aligned these residual firing rates to either the model-predicted state changes or the generative environmental state changes. We masked firing rates before the preceding state change and after the following state change when applicable. We computed the mean of these residual firing rates for visualization. To test for significant discrimination of state, we compared d' in the real data to a permutation distribution created by permuting state labels across state changes (2-tailed permutation test, 250 permutations, $p < .05$).

We defined model-predicted state changes as time points where the running average of the mean of the posterior accumulator value crossed 0. The running average was computed over 100 bins of 1ms. To avoid introducing noisy state changes, we excluded state changes from the first and last 200ms of the trial. We also excluded state changes that did not meet two change strength criteria designed to identify state changes that were immediately reversed. The first, was based on the average value of the posterior mean in the 100ms before the change compared to the 100ms after the change. State changes were excluded if these strength values were inconsistent with the direction of the identified state change. The second state change strength was based on the slope of the running average of the posterior mean at the time of the change. If the sign of the slope was inconsistent with the sign of the state following the state change, this means that the accumulation value immediately returned back to the previous state. We excluded these state changes. Our results were robust to variations in the state change inclusion criteria.

State change triggered tuning State change triggered tuning maps $E[\Delta r(a, t - t_c)]$ were computed using the tuning curve methods described above, but using time relative to state changes instead of stimulus onset. Firing rates were masked before and after the preceding and following state changes as described above. Data was also masked in the 300ms around the state change where the accumulated value distribution is too narrow to estimate tuning. Rank 1 approximations and population tuning maps were computed as above.

Author contributions

AP, and AEH designed the study. AEH managed rat training and care. AEH, and EJD recorded the neural data. AP, and TB analyzed the neural data. AP, AEH, and TB wrote the manuscript. AEH, and CB oversaw all aspects of the project.

Acknowledgements

We thank members of the Brody lab and Zachary Kilpatrick for useful conversations and feedback. AEH acknowledges support by NIH grant 1R21MH121889-01. EJD is supported by an HHMI Hanna H. Gray Fellowship and an HHMI-Helen Hay Whitney Postdoctoral Fellowship.

References

[1] Joshua I. Gold and Michael N. Shadlen. The neural basis of decision making. *Annual Review of Neuroscience*, 30(1):535–574, 2007. doi: 10.1146/annurev.neuro.29.051605.113038.

[2] Ian Krajbich, Todd Hare, Bjrn Bartling, Yosuke Morishima, and Ernst Fehr. A common mechanism underlying food choice and social decisions. *PLoS Computational Biology*, 11:e1004371, 2015. doi: 10.1371/journal.pcbi.1004371.

[3] Roger Ratcliff. A theory of memory retrieval. *Psychological Review*, 85(2):59–108, 1978.

[4] Carlos D Brody and Timothy D Hanks. Neural underpinnings of the evidence accumulator. *Current Opinion in Neurobiology*, 37:149 – 157, 2016. ISSN 0959-4388. doi: https://doi.org/10.1016/j.conb.2016.01.003.

[5] Bingni W. Brunton, Matthew M. Botvinick, and Carlos D. Brody. Rats and humans can optimally accumulate evidence for decision-making. *Science*, 340(6128):95–98, 2013. ISSN 0036-8075. doi: 10.1126/science.1233912.

[6] Timothy D. Hanks, Charles D. Kopec, Bingni W. Brunton, Chunyu A. Duan, Jeffrey C. Erlich, and Carlos D. Brody. Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature*, 520(7546): 220–223, Apr 2015. ISSN 0028-0836.

[7] Benjamin B. Scott, Christine M. Constantinople, Athena Akrami, Timothy D. Hanks, Carlos D. Brody, and David W. Tank. Fronto-parietal cortical circuits encode accumulated evidence with a diversity of timescales. *Neuron*, 95(2):385–398.e5, 2017. ISSN 0896-6273. doi: 10.1016/j.neuron.2017.06.013.

[8] Michael M Yartsev, Timothy D Hanks, Alice Misun Yoon, and Carlos D Brody. Causal contribution and dynamical encoding in the striatum during evidence accumulation. *eLife*, 7:e34929, aug 2018. ISSN 2050-084X. doi: 10.7554/eLife.34929. URL `https://doi.org/10.7554/eLife.34929`.

[9] Charles D. Kopec, Jeffrey C. Erlich, Bingni W. Brunton, Karl Deisseroth, and Carlos D. Brody. Cortical and subcortical contributions to short-term memory for orienting movements. *Neuron*, 88(2):367 – 377, 2015. ISSN 0896-6273. doi: https://doi.org/10.1016/j.neuron.2015.08.033.

[10] Jeffrey C Erlich, Max Bialek, and Carlos D Brody. A cortical substrate for memory-guided orienting in the rat. *Neuron*, 72(2):330–343, 2017/07/13 2011. ISSN 0896-6273. doi: 10.1016/j.neuron.2011.07.010.

[11] Christian L. Ebbesen, Michele N. Insanally, Charles D. Kopec, Masayoshi Murakami, Akiko Saiki, and Jeffrey C. Erlich. More than just a "motor": Recent surprises from the frontal cortex. *Journal of Neuroscience*, 38(44):9402–9413, 2018. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.1671-18.2018. URL `https://www.jneurosci.org/content/38/44/9402`.

[12] Alex Piet, Jeff Erlich, Charles Kopec, and Carlos D. Brody. Rat prefrontal cortex inactivations are explained by bistable attractor dynamics. *Neural Computation*, 2017.

[13] Rafal Bogacz, Eric Brown, Jeff Moehlis, Philip Holmes, and Jonathan D Cohen. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol Rev*, 113(4):700–765, October 2006. doi: 10.1037/0033-295X.113.4.700.

[14] Christopher M Glaze, Joseph W Kable, and Joshua I Gold. Normative evidence accumulation in unpredictable environments. *eLife*, 4:e08825, aug 2015. ISSN 2050-084X. doi: 10.7554/eLife.08825.

[15] Alan Veliz-Cuba, Zachary P. Kilpatrick, and Kresimir Josic. Stochastic models of evidence accumulation in changing environments. *SIAM Review*, 2016.

[16] Alex T. Piet, Ahmed El Hady, and Carlos D. Brody. Rats adopt the optimal timescale for evidence integration in a dynamic environment. *Nature Communications*, 9(1):4265, 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-06561-y. URL `https://doi.org/10.1038/s41467-018-06561-y`.

[17] Adrian E. Radillo, Alan Veliz-Cuba, Krešimir Josić, and Zachary P. Kilpatrick. Performance of normative and approximate evidence accumulation on the dynamic clicks task. *Neurons, Behavior, Data Analysis and Theory*, 3(1), 2019. doi: https://arxiv.org/pdf/1902.01535v3.pdf.

[18] Zachary P. Kilpatrick, William R. Holmes, Tahra L. Eissa, and Krešimir Josić. Optimal models of decision-making in dynamic environments. *Current Opinion in Neurobiology*, 58:54–60, 2019. ISSN 0959-4388. doi: https://doi.org/10.1016/j.conb.2019.06.006. URL `https://www.sciencedirect.com/science/article/pii/S095943881830223X`. Computational Neuroscience.

[19] Diogo Peixoto, J. R. Verhein, Roozbeh Kiani, Jonathan C. Kao, Paul Nuyujukian, Chandramouli Chandrasekaran, Julian Brown, Sania Fong, Stephen I. Ryu, Krishna V. Shenoy, and William T. Newsome. Decoding and perturbing decision states in real time. *Nature*, 591:604–609, 2021. doi: https://doi.org/10.1038/s41586-020-03181-9. URL `https://www.nature.com/articles/s41586-020-03181-9#citeas`.

[20] Arbora Resulaj, Roozbeh Kiani, Daniel M. Wolpert, and Michael N. Shadlen. Changes of mind in decision-making. *Nature*, 461(7261):263–266, 2009.

[21] Ignasi Cos, Giovanni Pezzulo, and Paul Cisek. Changes of mind after movement onset: a motor-state dependent decision-making process. *bioRxiv*, 2021. doi: 10.1101/2021.02.15.431196. URL `https://www.biorxiv.org/content/early/2021/02/16/2021.02.15.431196`.

[22] Peter R. Murphy, Niklas Wilming, Diana C. Hernandez-Bocanegra, Genis Prat-Ortega, and Tobias H. Donner. Adaptive circuit dynamics across human cortex during evidence accumulation in changing environments. *Nature Neuroscience*, 2021.

[23] Jeffrey C Erlich, Bingni W Brunton, Chunyu A Duan, Timothy D Hanks, and Carlos D Brody. Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat. *eLife*, 4:e05457, apr 2015. ISSN 2050-084X. doi: 10.7554/eLife.05457.

[24] Nuo Li, Kayvon Daie, Karel Svoboda, and Shaul Druckmann. Robust neuronal dynamics in premotor cortex during motor planning. *Nature*, 532:459–464, 2016. doi: doi:10.1038/nature17643.

[25] Arseny Finkelstein, Lorenzo Fontolan, Michael N. Economo, Nuo Li, Sandro Romani, and Karel Svoboda. Attractor dynamics gate cortical information flow during decision-making. *Nature Neuroscience*, Apr 2021. ISSN 1546-1726. doi: 10.1038/s41593-021-00840-6. URL `https://doi.org/10.1038/s41593-021-00840-6`.

[26] Genís Prat-Ortega, Klaus Wimmer, Alex Roxin, and Jaime de la Rocha. Flexible categorization in perceptual decision making. *Nature Communications*, 12(1):1283, 2021.

[27] Nathaniel D. Daw. Trial-by-trial data analysis using computational models. In *Decision Making, Affect, and Learning: Attention and Performance XXIII*. Oxford University Press, 5 2011. ISBN 9780191725623. doi: 10.1093/acprof:oso/9780199600434.003.0001.

[28] Jarrett Revels, Miles Lubin, and Theodore Papamarkou. Forward-mode automatic differentiation in julia. *arXiv:1607.07892 [cs.MS]*, 2016.

[29] Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Charu Bai Reddy, Matteo Carandini, and Kenneth D. Harris. Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, 364(6437), 2019. ISSN 0036-8075. doi: 10.1126/science.aav7893. URL `https://science.sciencemag.org/content/364/6437/eaav7893`.

Supplementary Materials

The supplementary materials contains extended figures, control analyses, and in-depth method discussions.

Supplementary Materials - Individual Rat Behavior

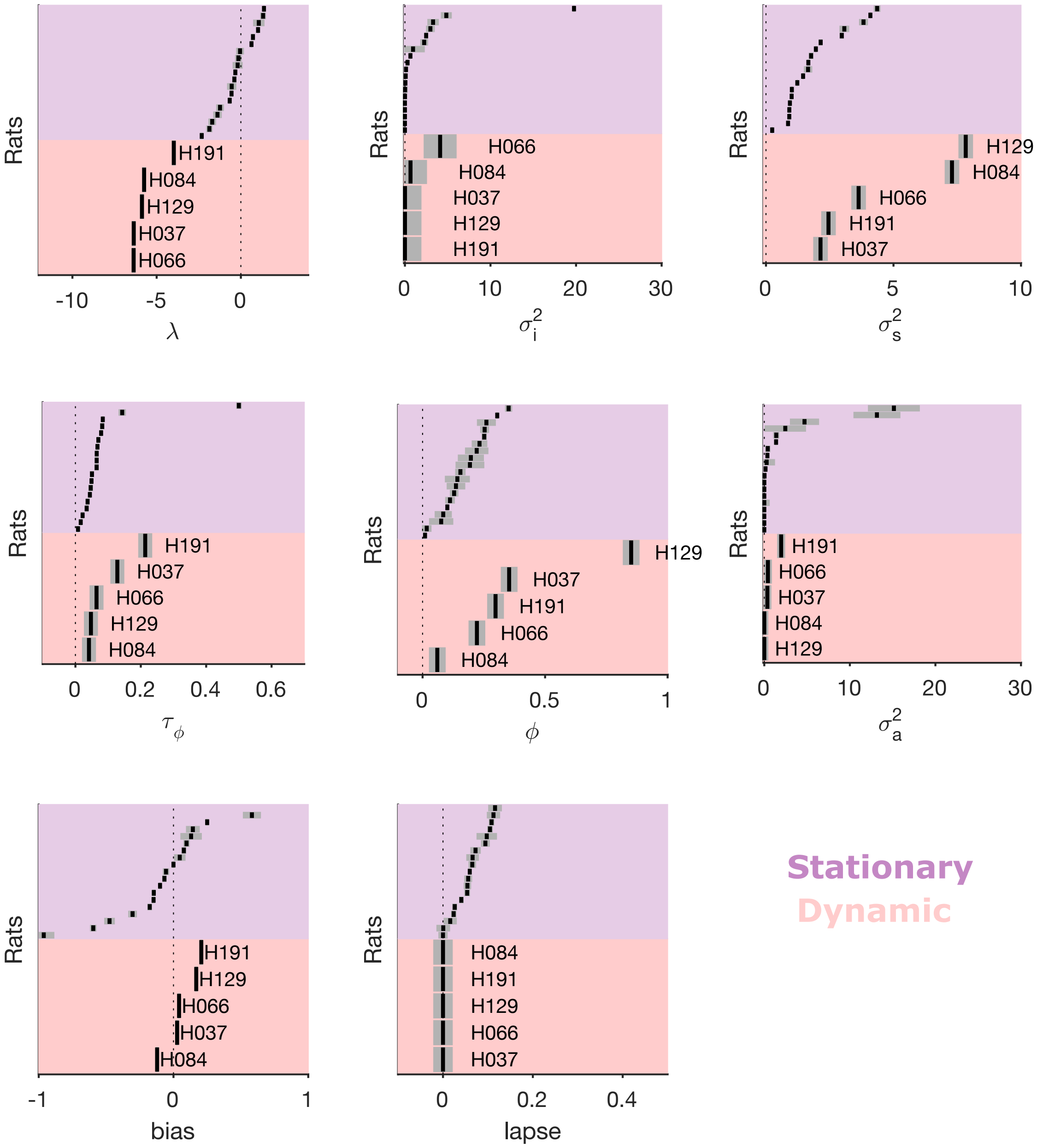

Figure S1: **Accumulation Model Parameters.** Best fit model parameters and 95% confidence intervals for each rat in this study. In addition the model parameter fits reported in Brunton et al.[5] for 19 rats in a stationary environment are included for comparison.

687

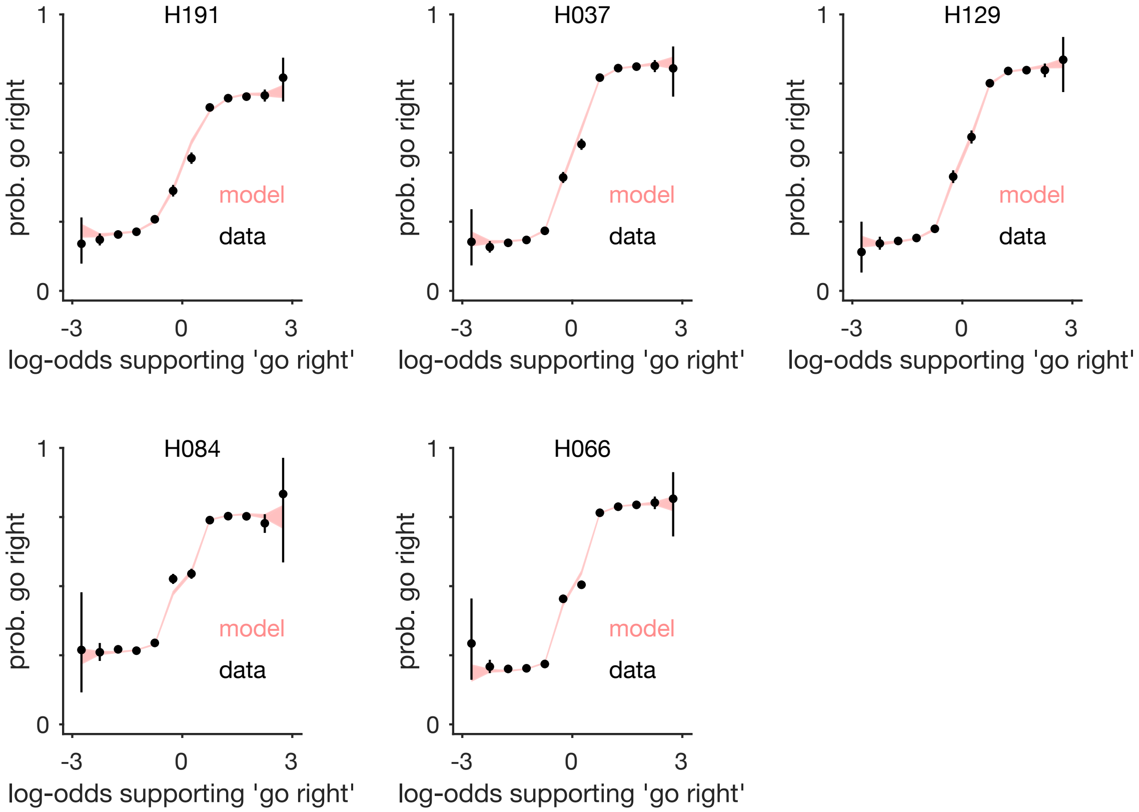

Figure S2: **All rats' psychometric curves.** Each plot shows the probability of a right choice given an ideal observer's log-odds supporting a go-right choice.

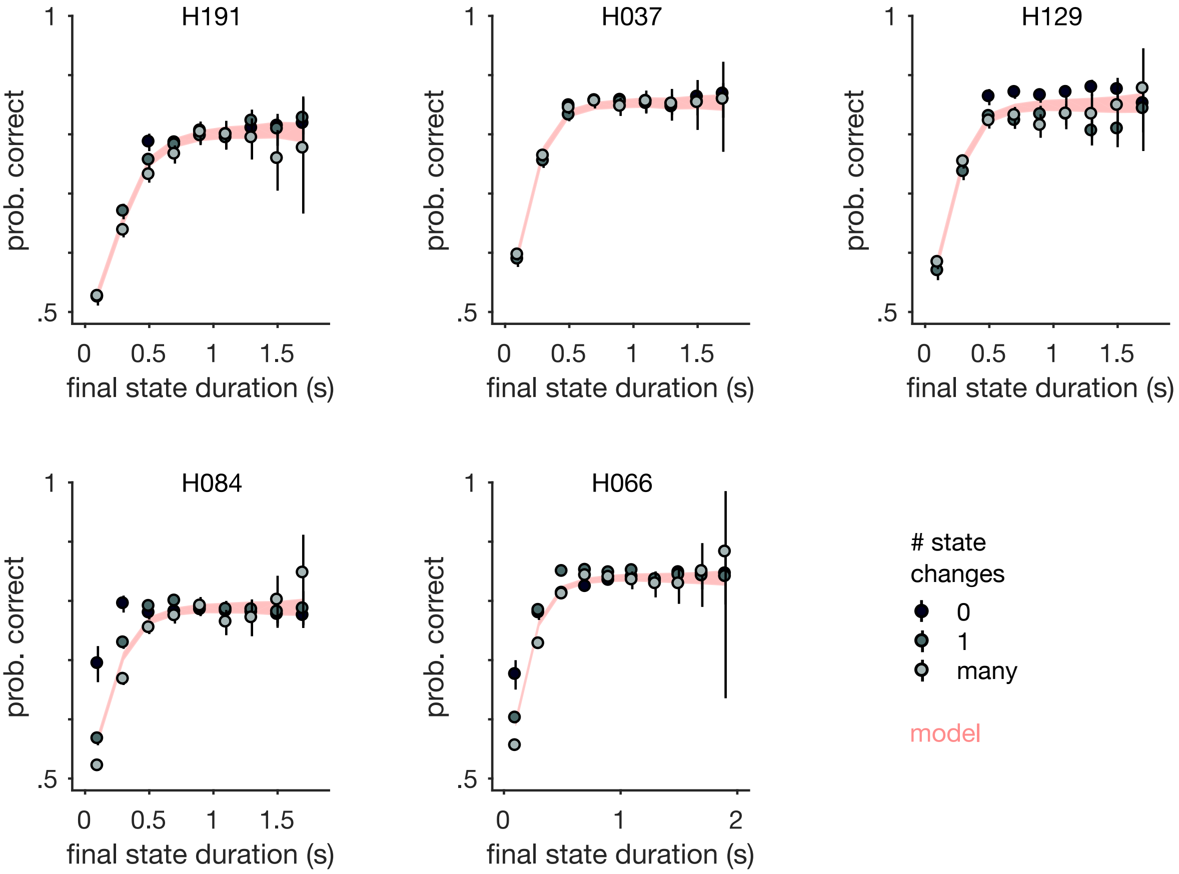

Figure S3: **All rats' final state chronometric curves.** Each plot shows the probability the rat made the correct choice as a function of the final state duration, and the total number of state changes in the trial. The final state duration was defined as the time from the last environmental state change to the end of the stimulus period. The best fit model prediction averaged over the total number of state switches is shown in pink.

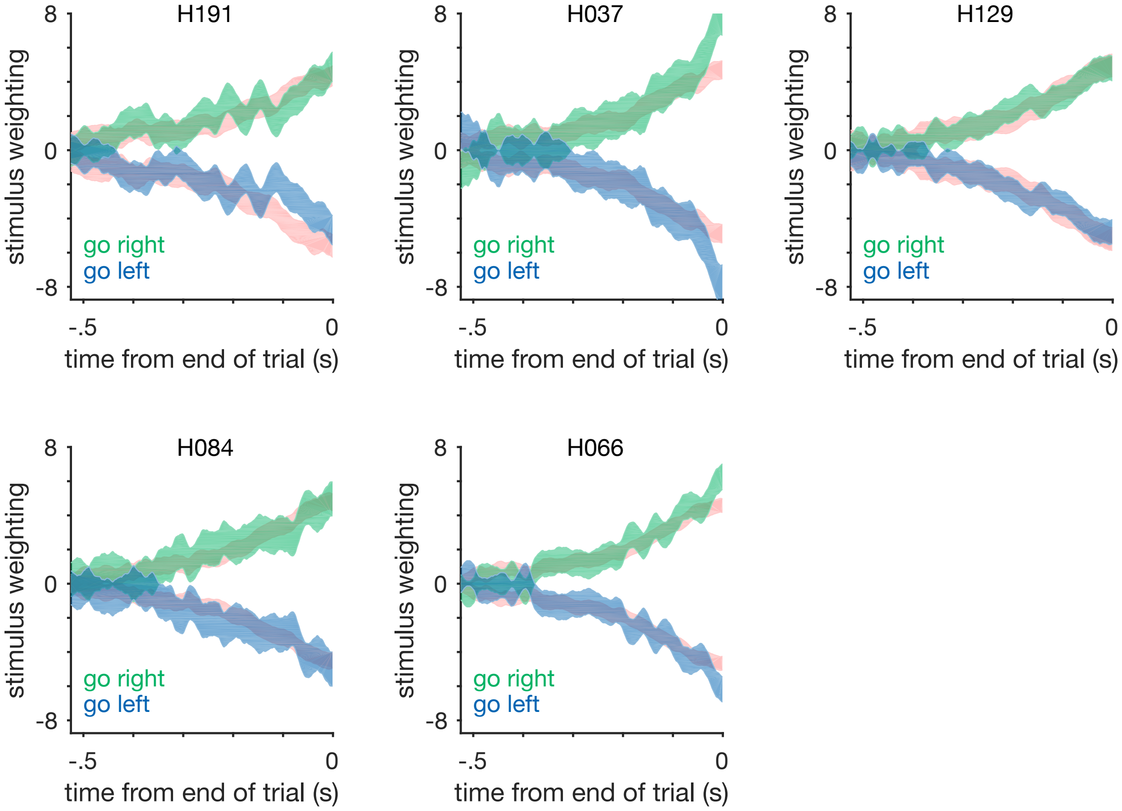

Figure S4: **All rats' psychophysical reverse correlation kernels.** Each plot shows the reverse correlation kernel for the rat (blue/green) and the best fit model (pink).

Supplementary Materials - Table of Electrophysiology Details

Rat	Implant Location	Recording Type
H191	blFOF	tetrode
H037	rFOF	array
H129	lFOF	array
H084	lFOF	array
H066	rFOF	array

Table 1. List of rats used in the study, and the recording location

Supplementary Materials - Posterior Accumulation Model

The accumulation model has two components, the forward model and the backward model. When these components are combined it creates the posterior model used for analyzing neural data. The forward model is used to estimate a set of model parameters θ. It assumes the initial distribution of accumulation values is gaussian distributed with mean zero and variance $\sigma_i^2 \in \theta$,

$$P(a|t=0) = \mathcal{N}\left(\mu_0 = 0, \sigma_i^2\right). \tag{20}$$

So we can write the full forward model as:

$$f(a) = P(a|t, \delta_R, \delta_L, \theta, a_0 = \mathcal{N}\left(\mu_0 = 0, \sigma_i^2\right)) \tag{21}$$

At each moment in the trial, the forward model distribution of accumulation values $f(a)$ is Gaussian distributed with mean μ and variance σ^2 given by:

$$\begin{aligned}
\mu(t) &= \mu_0 e^{\lambda t} + \int\limits_0^t \left(\delta_{R,s} \cdot C\left(R(s)\right) - \delta_{L,s} \cdot C\left(L(s)\right)\right) ds & (22)\\
&= \sum_i^{\#R_t} e^{\lambda(t-R(i))} C(R(i)) - \sum_i^{\#L_t} e^{\lambda(t-L(i))} C(L(i)) & (23)\\
\sigma^2(t) &= \sigma_i^2 e^{\lambda t} + \frac{\sigma_a^2}{2\lambda}\left(e^{2\lambda t} - 1\right) + \int\limits_0^t \sigma_s^2 \left(\delta_{R,s} \cdot C\left(R(s)\right) - \delta_{L,s} \cdot C\left(L(s)\right)\right) e^{2\lambda t} ds & (24)\\
&= \sigma_i^2 e^{\lambda t} + \frac{\sigma_a^2}{2\lambda}\left(e^{2\lambda t} - 1\right) + \sum_i^{\#R_t} \sigma_s^2 C(R(i)) e^{2\lambda(t-R(i))} + \sum_i^{\#L_t} \sigma_s^2 C(L(i)) e^{2\lambda(t-L(i))} & (25)
\end{aligned}$$

Where $\#R_t$ is the number of right clicks on this trial up to time t, and $R(i)$ is the time of the i^{th} right click. $C(R(i))$ tells us the effective adaptation for that click.

Computing the posterior distribution is more complicated than the forward model. First, we find the parameter set that best explains the choices y on each trial by maximizing $P(y|\theta)$. Then we use the best fit parameter set to evaluate the forward model for each trial, producing a probability distribution over accumulated evidence value at each time point consistent with the initial conditions. Next, we compute the backward model $b(a)$. Note this is not the "backward pass distribution" discussed by Brunton et al.[5]. The backward model here ignores the forward model, and instead computes the probability distribution over observing accumulated evidence values at each time point consistent with the stimulus and the choice at the end of the trial t_N.

$$b(a) = \begin{cases} P(a|t, \delta_R, \delta_L, \theta, a_N = \mathcal{U}(B, \infty)), & \text{if } y = 1 \\ P(a|t, \delta_R, \delta_L, \theta, a_N = \mathcal{U}(-\infty, B)), & \text{if } y = -1. \end{cases} \tag{26}$$

Importantly, the forward and backward distributions are conditionally independent, conditioned on the final value of the accumulated evidence. Given that they are independent, the posterior distribution that combines both observations is the product of the forward and backward distributions.

$$P\left(a|t, \delta_R, \delta_L, \theta, a_0 = \mathcal{N}\left(0, \sigma_i^2\right), y\right) = f(a)b(a) \tag{27}$$

One technical wrinkle is that our analytical solution for solving the model (forward or backward) relies on initial conditions that are gaussian. The forward model assumes an initial distribution $\mathcal{N}\left(0, \sigma_i^2\right)$. However, for the backward model the choice data only constrains the sign of a at the end of the trial, meaning our initial conditions for the backward model is a uniform distribution over $(B, \pm\infty)$. Therefore, we constructed a solution by discretizing the a-value axis into small bins of width Δa, and solved the backward model for each bin assuming a delta function of initial probability mass in each bin. We refer to the backward distribution from each bin i as the delta-backward solution $b_i(a)$. Our entire backward distribution is the mixture distribution over all the individual delta-backward solutions.

$$b(a) = \sum_{i=0}^{\pm\infty} w_i b_i(a) \tag{28}$$

$$b_i(a) = P(a|t, \delta_R, \delta_L, \theta, a_N = \mathcal{N}(B + i\Delta a, 0)) \tag{29}$$

To solve each delta-backward solution $b_i(a)$, we use the time-reversed solution to the forward model. The mixture weights w_i are all equal if the bin spacing is uniform. Note that it might be tempting to think that we need to weight each individual delta-backward solution by the forward model's probability mass in each bin; however, this is not correct. Given that the backward model is independent of the forward model, we want the complete backward distribution to reflect all possible states consistent with the choice observation, which is the uniform distribution over the correct sign of a. With a set of $b_i(a)$ solutions, we can now combine them into the posterior distribution, $p(a)$. The exact solution as $\Delta a \to 0$ is given by:

$$p(a|t, \delta_R, \delta_L, \theta, a_0 = \mathcal{N}\left(0, \sigma^2_{initial}\right), y) \propto f(a) \sum_{i=0}^{\pm\infty} w_i b_i(a). \tag{30}$$

To clarify notation, we now write the initial variance of the distribution as $\sigma^2_{initial}$ to distinquish this parameter from the series of grid solutions indexed by "i". In practice we truncate the infinite series at a suitable extrema value of a, and use a finite bin spacing $\Delta a = 1$. On each trial, the extent of the grid of delta solutions was determined by finding the accumulation value where less than $1e-4$ probability mass of the posterior model lay beyond that point at the end of the trial. Given that $f(a)$, and $b_i(a)$ will be gaussian, let $p_i(a) = f(a)b_i(a)$, which is also gaussian. This lets us write the posterior distribution as the sum of many delta-posterior modes.

$$p(a|t, \delta_R, \delta_L, \theta, a_0 = \mathcal{N}\left(0, \sigma^2_{initial}\right), y) \propto \sum_{i=0}^{\pm\infty} w_i p_i(a). \tag{31}$$

Toy Model Example

To illustrate the posterior model, consider a simple random walk process. On each time step there is a 1/3 probability of staying in place, 1/3 probability of taking a step of size $+1$, and a 1/3 probability of taking a step of size -1. We start at $a(t = 0) = 0$, and then observe the process ten time steps later, $t = 10$. We now want to compute the forward model, and the posterior distribution for this data.

Using a binomial distribution we can analytically compute the forward model, which describes the probability of observing the process given the initial conditions and an elapsed duration (Fig. S5A). To compute a backward model, we must define the final value. For example, if we let the final value of the random walk be $a(t = 10) = 2$, then we can again use the binomial distribution to compute the distribution of possible values at earlier time steps (Fig. S5B). We can alternatively let the final value be the sign of the random walk process and compute the full backwards distribution. Combining the independent forward and backwards distributions, we can predict the posterior distribution, which is all possible states the random walk could be in given these two observations (Fig. S5E). We can check this against a particle simulation by sampling $2,000,000$ trajectories from the random walk process to get the forward distribution (Fig. S5C). To get the posterior distribution we filter our samples for trajectories that ended

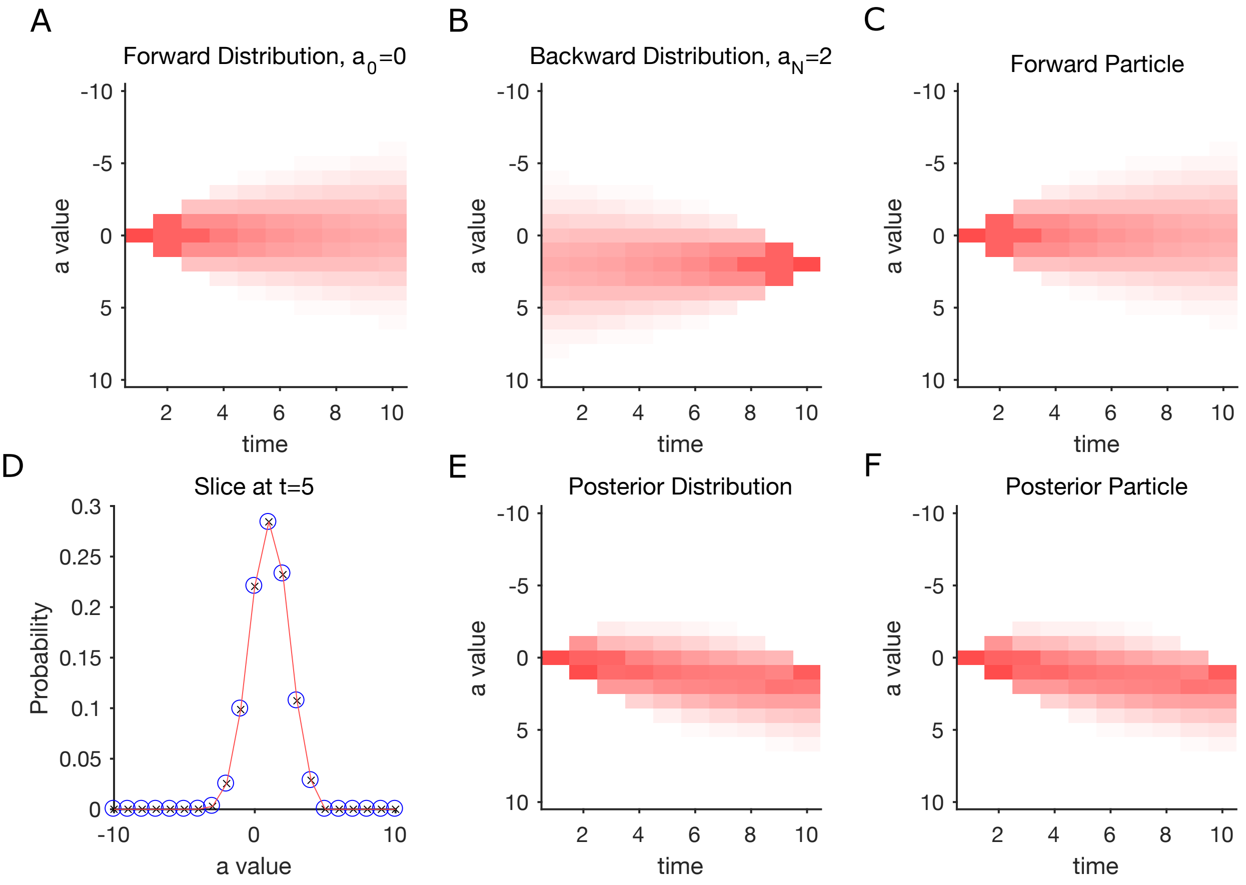

Figure S5: **Toy Model Demonstration** (A) Forward model analytical solution for the toy model. (B) Delta-Backward model analytical solution for final accumulation value of $a_N = 2$. (C) Distribution of the forward model from sampling. (D) Comparing the distributions from E (red) and F (blue) at $t = 5$. (E) Posterior model analytical solution for initial conditions $a_0 = 0$ and final "choice" $a > 0$. (F) Posterior model solution from sampling.

up with positive value (Fig. S5F). We can compare these two distributions by taking a slice in time (Fig. S5D). We can now move on to the accumulation model, which has the same basic random walk structure, but with a few more bells and whistles.

Verification of Accumulation Model Posterior Solution

To illustrate the posterior model solution on the full accumulation model we can again compare the analytical solution to sampled trajectories, this time for an example trial. Each trajectory has a unique noise realization.

For both the model and sample trajectories, we have four distributions (Fig. S6). The forward distribution showing the predicted trajectories given the initial conditions. The backwards-delta distribution showing the possible trajectories that result in a single final accumulation value. The posterior-delta distribution showing the possible trajectories that start at the initial condition and result in a single final accumulation value. Finally, the full posterior distribution that shows the possible accumulation values that start at the initial condition and result in the appropriate choice or sign of the accumulation value. In this example trial we consider a trial where there is change in state at 750ms, and we evaluate a left choice ($a < B$) at $t = 1$s. The entire solution was computed using 51 backward-delta solutions on a grid from $(-50, B)$ with $\Delta i = 1$. We can also examine slices through the posterior distribution at various time points to confirm agreement between the trajectories and the analytical solution (Fig. S7).

A few notes on the advantages of the analytical solution. First, the analytical model offers a large increase in accuracy of the model over previous numerical approaches. Second, the analytical model is much faster to fit and evaluate. Second, we can compute the posterior distribution for only a subset of all time points without computing the solution for all time points. This fact allows for very rapid computation of the posterior distribution.

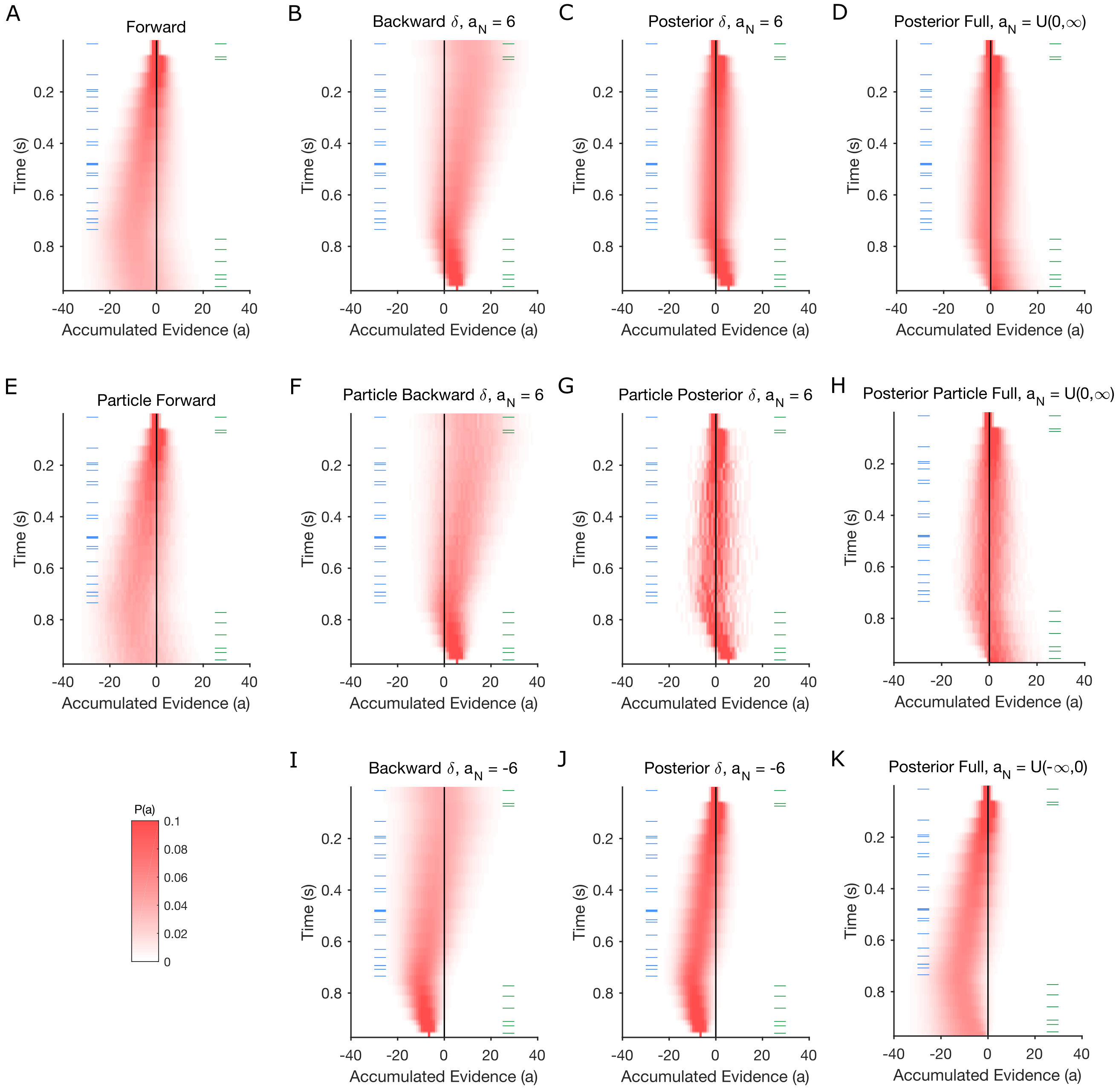

Figure S6: **Posterior Model Validation.** Comparison of the model distributions computed from the analytical solution (top and bottom) and sampled trajectories (middle) for an example trial. Green ticks mark times of right clicks, blue ticks mark times of left clicks. (A,E) Forward distribution assuming $a_0 = 0$. (B,F) Backwards-delta distribution assuming a final accumulation value of $a_N = 6$. (C,G) Posterior-delta distribution assuming $a_N = 6$ and $a_0 = 0$. (D,H) Full posterior distribution assuming $a_0 = 0$ and a right choice $(a > 0)$. (I) Backwards-delta distribution assuming a final accumulation value of $a_N = -6$. (J) Posterior-delta distribution assuming $a_N = 6$ and $a_0 = 0$. (K) Full posterior distribution assuming $a_0 = 0$ and a left choice $(a < 0)$.

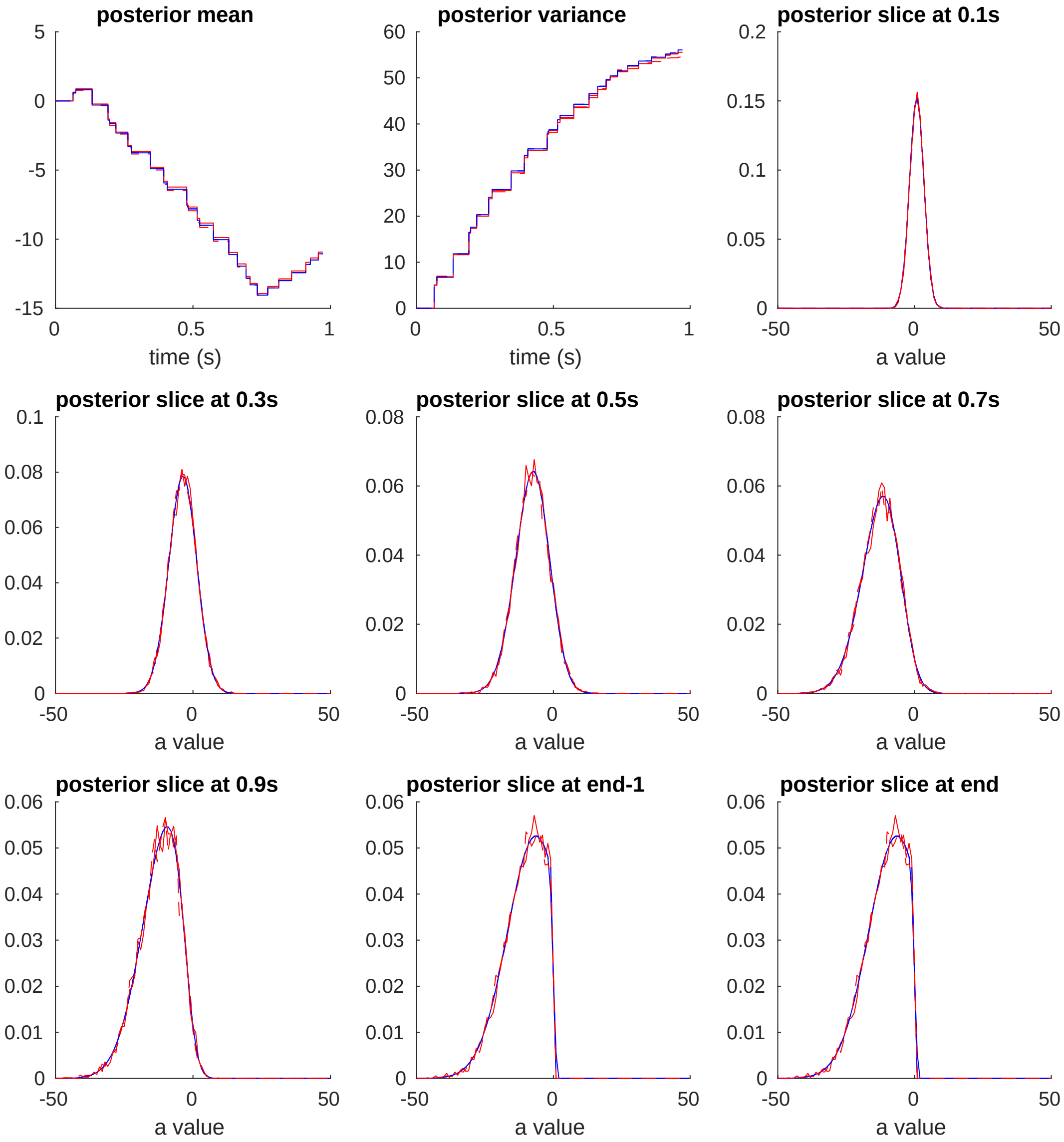

Figure S7: **Posterior Model Validation.** Comparison of the posterior model distribution computed from the analytical solution (blue) and sampled trajectories (red) for an example trial. This figure shows the same distribution as Fig. S6K (Red). (A) Mean of posterior over time. (B) Variance of posterior over time. (C-I) Posterior comparisons at specific time points.

Supplementary Materials - Switch Triggered Averages

To determine the times of changes of mind, we used the mean of the posterior distribution $P(a|t, \delta_R, \delta_L, \theta, y)$ on each trial. These trajectories have several sources of noise that complicated our analysis. First, they have sharp discontinuities at the times of each stimulus click. This sometimes resulted in the mean trajectory repeatedly crossing the decision boundary in a short period of time. It is unlikely that each of these crossing represented a true change of mind, since the subject was likely in a general state of indecision (Fig S8A). To resolve this issue we smoothed the mean trajectories with a 100 ms running average. This smoothing resolved the flickering changes of mind from individual stimuli (Fig S8B). However, this reveals a second issue. There are changes of mind that briefly cross the decision boundary, or still oscillate around the decision boundary. To detect and remove these time points we estimated the local slope of the smoothed trajectory, and filtered out changes of mind where the local slope had an inconsistent sign with the direction of the change of mind (Fig S8C). We excluded any changes of mind that occurred in the first or last 200ms of the trial.

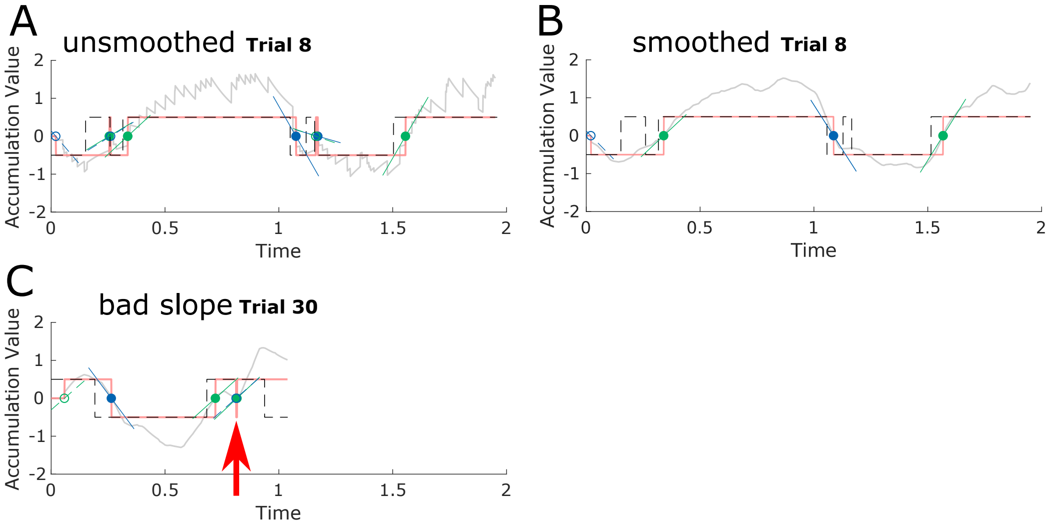

Figure S8: **Analysis of smoothing model mean and weak and strong state switches** (A) The mean trajectory of the posterior model is shown in gray, along with the categorical decision taken from the sign of the model accumulation value in pink. The generative state (dashed black traces) is overlaid on the model state indicator. Each change of mind is marked with a circle. Changes of mind to the 'go right' state are in green, and changes of mind to the 'go left' state in blue. The local slope is marked with a colored line. The initial decision at the start of the trial was excluded (empty circle). (B) The mean trajectory was smoothed, removing many spurious changes of mind when the trajectory flickered near the decision boundary. (C) In this example trial around 750ms, the mean trajectory is generally moving to the 'go right' (positive accumulation value) state, but briefly returns to the go-left state. This creates a situation where there are multiple changes of mind at the same time, with an incongruent slope (red arrow).