# Normative Circuit Dynamics Across Human Cortex During Evidence Accumulation in Changing Environments

Peter R Murphy[1,*], Niklas Wilming[1], Diana C Hernandez-Bocanegra[1,2], Genis Prat Ortega[3,4], Tobias H Donner[1,5,6,7,*]

[1]Section Computational Cognitive Neuroscience, Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Martinisrasse 52, Hamburg 20251, Germany

[2]Interactive Systems Group, Department of Computer Science and Applied Cognitive Science (INKO), University of Duisburg-Essen, Forsthausweg 2, Duisburg 47057, Germany

[3]Institut D'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Carrer del Rosselló 149, 08036 Barcelona, Spain

[4]Centre de Recerca Matematica, Campus UAB Edifici C, 08193 Bellaterra, Barcelona

[5]Bernstein Center for Computational Neuroscience, Charité Universitätsmedizin, Haus 6, Philippstraße 13, 10115, Berlin, Germany

[6]Department of Psychology, University of Amsterdam, Weesperplein 4, 1018 XA, Amsterdam, The Netherlands

[7]Amsterdam Brain and Cognition, University of Amsterdam, Nieuwe Achtergracht 129, 1018 WS, Amsterdam, The Netherlands

*Corresponding authors

## Abstract

Decision-making under uncertainty commonly entails the accumulation of decision-relevant 'evidence' over time. In natural environments, this accumulation process is complicated by the existence of hidden changes in the state of the environment. Optimal behavior in such contexts requires a rapid, non-linear tuning of the evidence accumulation. The neural basis of this adaptive computation has remained elusive. Here, we unraveled the underlying mechanisms across sensory, associative, and motor regions of the human cerebral cortex. We combined a visuo-motor choice task with hidden changes in the evidence source, monitoring of pupil-linked arousal, and model-based analysis of behavior and cortical population activity (assessed with magnetoencephalography). We found that normative, non-linear evidence accumulation results from the interplay of recurrent cortical dynamics and phasic arousal signals from ascending brainstem systems. Our insights forge a strong connection between normative computations for adaptive decision-making and the large-scale neural mechanisms for their implementation.

## Introduction

Many decisions under uncertainty entail the temporal accumulation of noisy information about the state of the world[1-4], a strategy that commonly increases choice accuracy. When the source of the evidence remains constant during decision formation, maximizing accuracy requires perfect, linear integration[4,5] (i.e., equal weight given to each sample of evidence). Progress in computational theory, behavioral modeling, and neurophysiological analysis has translated these ideas into an influential framework for decision-making wherein decision-relevant evidence is encoded by neurons in sensory cortical regions, and then accumulated over time along sensory-motor cortical pathways[2-4,6]. This computation produces the gradual build-up of choice-predictive activity that is commonly observed in association and (pre-)motor cortical regions during decision formation[7-17].

Despite its theoretical appeal and popularity, this 'canonical' framework is challenged by two insights (see also [18]). First, its development has mainly been informed by studying a single behavioral context: environments in which the source (i.e. stimulus category) generating evidence remains constant for single decisions. In this special case, the agent's uncertainty originates only from noise corrupting the evidence. However, natural environments can undergo unpredictable and hidden changes in their state, constituting an additional source of uncertainty[19]. Perfect evidence accumulation is normative for stable environments, but suboptimal for changing environments[20,21]. Second, the linear, feedforward integration of evidence in sensory-motor cortical pathways entailed in the above framework contrasts with several established features of the cerebral cortex: its recurrent organization, within and across areas of the cortical hierarchy[22,23]; non-linear dynamics within recurrent cortical circuits for evidence accumulation[24-26]; and, the dynamic modulation of these cortical circuits by ascending brainstem systems[27-30]. We reasoned that both limitations of the canonical framework may be closely connected, and set out to tackle them in an integrated approach that takes an ecological perspective on decision-making[31].

We leveraged the insight that normative decision-making in changing environments requires a non-linear evidence accumulation process, which balances stability of decision formation with sensitivity to change[20,32]. We unraveled the neural basis of this non-linearity through an integrated analysis of choice behavior in a changing environment, source-level magnetoencephalography (MEG) signals across the sensory-motor cortical pathway, and computational modelling of both behavior and cortical circuit dynamics. We find that normative evidence accumulation emerges naturally from the non-linear dynamics within parietal and (pre-)motor cortical circuits, combined with two separable effects on visual cortex: feedback of cortical decision variables from downstream, and modulation through ascending brainstem systems. These insights forge a strong link between decision computations and the large-scale neural mechanisms for their implementation[33], yielding a new framework for perceptual decision-making that is normative, ecological, and takes account of key features of cortical structure and function.

## Results

We developed a task to uncover mechanisms of adaptive perceptual evidence accumulation (Figure 1a). The task entailed a key feature of natural environments: hidden changes in the environmental state (i.e. evidence source). The presence of these state changes rendered perfect evidence accumulation a suboptimal strategy. The decision-relevant evidence was provided by a fast (inter-stimulus interval=0.4 s) sequence of small checkerboard patches ('samples') located at different polar angles along a semicircle in the lower visual hemifield

(see *Methods* for details). Each sample's position was generated from one of two noisy sources: the probability distributions shown in the upper row of Figure 1a. The 'active' source was chosen at random at the beginning of each trial and could change at any time during the trial, with low probability (hazard rate $H$=0.08). The agent's task was to judge which source was active *at the end* of each sequence. Human participants reported this judgment via left- or right-handed button presses, thus prompting translation of the sensory input on each trial into competing motor plans.

The Results section is organized as follows. In the first section, we establish that the non-linearity entailed in the normative (i.e. Bayesian) solution for this task[20,32] could be cast in terms of sensitivities of evidence accumulation to two intuitive computational quantities that might drive brainstem neuromodulatory systems[19,34,35]: change-point probability (a high-level form of surprise) and uncertainty. This insight set the stage for our characterization of behavioral and neural signatures of adaptive evidence accumulation in the next three sections. Sensitivities to the same two quantities were present in human behavior and choice-specific activity in multiple cortical regions; and, comparison of computational properties across cortical regions implicated selective feedback from decision-related to early sensory regions in the accumulation process. The final two sections inferred the circuit mechanisms underlying the above behavioral and neural signatures, showing that (i) these signatures emerge naturally from an influential circuit model of decision-making[26] and (ii) change-point-driven brainstem arousal responses (inferred via pupil dilation) contribute to the dynamic adaptation of evidence accumulation.

## Two quantities that govern normative evidence weighting

The normative model for our task (Figure 1b) entails the accumulation of evidence samples that carry information (in the form of log-likelihood ratios, *LLR*) about the two possible environmental states[20] – a process also known as 'belief updating'. The key difference between the normative model and previous accumulator models[1,2] is that it dynamically shapes the evidence accumulation (the combination of prior belief $\psi$ with new evidence *LLR* to yield posterior belief $L$) by the estimated rate of environmental state change ($H$)[20]. Specifically, the prior for the next sample ($\psi_{n+1}$) is determined by passing the updated belief ($L_n$) through a non-linear function that, depending on $H$, can saturate (slope≈0) for strong $L_n$ and entail more moderate information loss (0<<slope<1) for weak $L_n$ (Supplementary Figure 1). By this process, the normative model strikes an optimal balance between formation of strong beliefs in stable environments versus fast change detection in volatile environments.

We derived from this model two computational quantities that captured key features of the above non-linear transformation (see *Methods*). The first, change-point probability (denoted as *CPP*), quantified the likelihood, given an existing belief and new evidence sample, that a state change just occurred. *CPP* scaled with the inconsistency between the new sample and existing belief (Figure 1c), and thus captured a high-level form of surprise about the state of the environment. The second quantity, $-|\psi|$, quantified the uncertainty of the observer's belief about the environmental state before encountering a new sample (Figure 1d). Neither variable was explicitly used in the normative computation, but they helped pinpoint diagnostic signatures of adaptive evidence accumulation in behavior and neural activity (see *Discussion*).

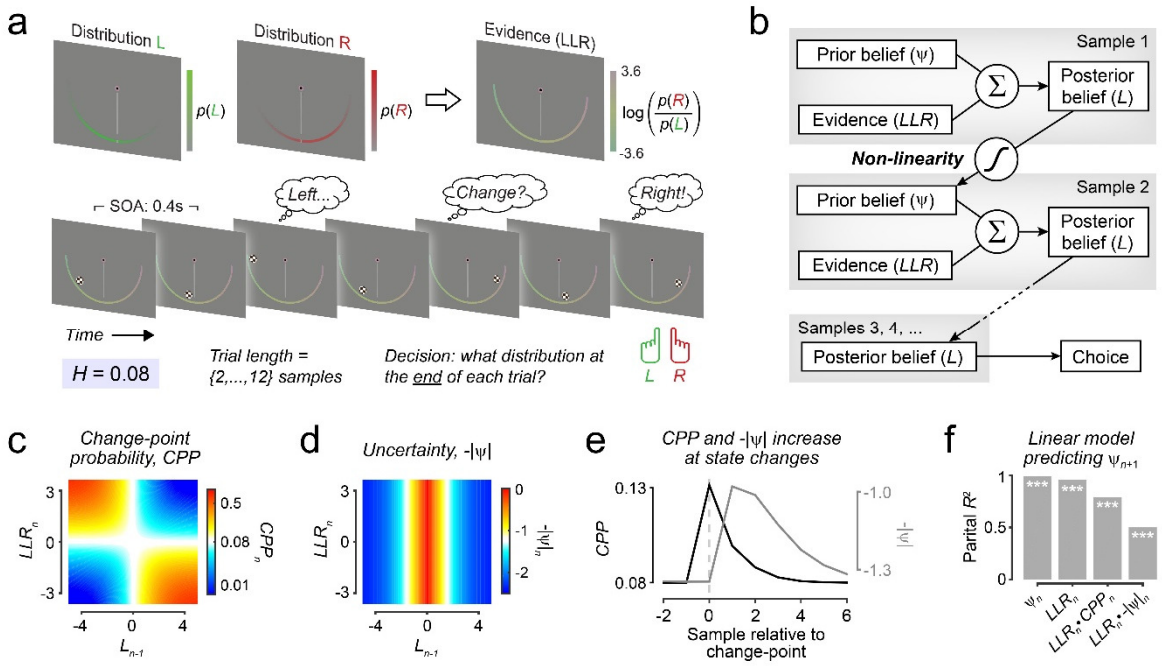



***Figure 1. Behavioral task and normative model with diagnostic signatures of non-linear evidence accumulation.*** *(**a**) Schematic of two-alternative perceptual choice task with hidden state changes. Upper row: probability densities for generative distributions and their combination into evidence (LLR) for different alternatives. Bottom row: example sample sequence. (**b**) Schematic of normative evidence accumulation process. (**c,d**) Change-point probability (CPP; c) and uncertainty (-|ψ|; d) as function of posterior belief after accumulating the previous sample ($L_{n-1}$) and evidence provided by the current sample ($LLR_n$). (**e**) CPP and –|ψ| dynamics derived from normative model and centered on change-points in generative task state. (**f**) Contribution of key computational variables to belief updating. Contributions are expressed as coefficients of partial determination for terms of linear model predicting the updated prior belief. ***$p<10^{-6}$, two-tailed t-test. Panels c-f based on generative statistics of behavioral task (see main text).*

In the normative model, *CPP* tended to increase transiently when a state change occurred, and this was followed by a period of heightened uncertainty (*-|ψ|*; Figure 1e). This was true across different levels of *H* and environmental noise (Supplementary Figure 1). A linear model in which both quantities modulated belief updating reliably predicted the extent to which an ideal observer integrated a new sample of evidence to form the prior belief for the next sample (**Model 1**, *Methods*; Figure 1f; Supplementary Figure 1; median $R^2$ across generative settings = 99.7%, range = [89.2% 99.98%]). Other forms of surprise only weakly modulated belief updating by comparison, and in a different direction (see Supplementary Figure 2 and *Discussion*). Thus, the non-linearity in the normative model gave rise to strong and specific sensitivities of evidence accumulation to *CPP* and *-|ψ|*.

## Signatures of adaptive evidence weighting in human evidence accumulation behavior

We asked 17 healthy human participants to perform the task from Figure 1a. Several observations indicate that their behavior was well explained by the adaptive accumulation process prescribed by the normative model, whereby this process operated on noisy and biased representations of evidence (*LLR)* and environmental stability (*H;* see Figure 2 and Supplementary Figure 3). For the stimulus sequences shown to our participants, an ideal observer with perfect knowledge of *H* and the generative distributions, and without internal noise or biases, achieved an accuracy of 81.8 % ± 0.9 % (mean and s.e.m. across stimulus sequences). Our participants' performance came close to this benchmark (77.5 % ± 1.9 %; mean and s.e.m.), albeit limited by internal noise, biases, and imperfect knowledge of the task

statistics. Critically, their performance was better than expected from two alternative strategies that are normative only in extreme contexts ($H$=0 or $H$=0.5, respectively; Figure 2a): (i) perfect, linear integration of all evidence samples (73.9 % ± 1.2 %; $t_{16}$=7.0, $p<10^{-5}$, two-tailed paired $t$-test); and (ii) deciding based only on the final sample without any evidence accumulation (71.8 % ± 0.6 %; $t_{16}$=11.7, $p<10^{-8}$, two-tailed paired $t$-test).

Our participants exhibited the same sensitivity to *CPP* and *-|ψ|* as the normative process. We quantified the modulation of evidence accumulation by these quantities through a logistic regression model (**Model 2**, *Online Methods*). The first set of regression weights (*LLR* at different sample positions) reflected the so-called psychophysical kernel that quantified the time course of the leverage of evidence on choice[15,36,37]. The evidence weights increased over time, a phenomenon known as recency ($t_{16}$=-7.8, $p<10^{-6}$, two-tailed paired $t$-test on weights for first 6 samples *vs*. last 6 samples; Figure 2b, *left*). Critically, the second and third sets of regression weights, which captured the modulation of evidence weighting by *CPP* and *-|ψ|*, each revealed strong positive modulations (*CPP*: $p$<0.0001, Figure 2b, *middle*; *-|ψ|:* $p$=0.0002, Figure 2b, *right;* two-tailed cluster-based permutation tests). The modulatory effect of *CPP* was also larger than that of *-|ψ|* ($t_{16}$=5.6, $p$<0.0001, two-tailed paired $t$-test averaging over sample position). In other words, both key computational quantities 'up-weighted' the impact of the associated evidence on choice and the effect of *CPP* was particularly strong, just as prescribed by the normative model.

The behavioral results enabled us to eliminate further alternative accumulation schemes, based on qualitative and/or quantitative criteria (Figure 2a-c; Supplementary Figures 3-4). The behavioral signatures from Figure 2b (black) qualitatively ruled out any linear (perfect or leaky) evidence accumulation. Specifically, leaky accumulation with exponential decay of accumulated evidence ('forgetting'; refs. [2,38]; purple, in Figure 2b-d) produced recency in psychophysical kernels (in contrast to flat kernels produced by perfect accumulation), but failed to capture the *CPP* or *-|ψ|* modulations of evidence weighting (Figure 2b). Consequently, all considered versions of a leaky accumulator model provided worse fits than the best-fitting version of the normative model, referred to as 'normative fit' and shown in red in Fig. 2 (mean Δ Bayes Information Criteria (BIC) = 71.1 ± 7.6 (s.e.m.); higher BIC for all 17 participants; Figure 2c; Supplementary Figure 4; see *Methods* for a detailed motivation of modelling choices). A different class of model that employed perfect accumulation combined with a non-linearity – non-absorbing bounds – did produce the qualitative behavioral signatures (Supplementary Figure 5), with only marginally worse fits to the data than the best-fitting version of the normative model (ΔBIC = 11.1 ± 2.9; higher BIC for 16 of 17 participants; Figure 2c). The improved performance of the bounded accumulator over the leaky accumulator (ΔBIC = -60.0 ± 7.6; lower BIC for all 17 participants) was informative: only the bounded accumulator is capable of approximating the non-linearity in the normative model applied to our task setting[20], indicating that this feature of normative accumulation (and not the leak present for weak belief states) was critical in accounting for participants' behavior.

In sum, a non-linear modulation of evidence weighting by *CPP* and *–|ψ|* was a key feature of human behavior in the current task. We next sought to pinpoint the modulatory effect of these quantities on the evidence accumulation process at the neural level, assessed with MEG and pupillometry.

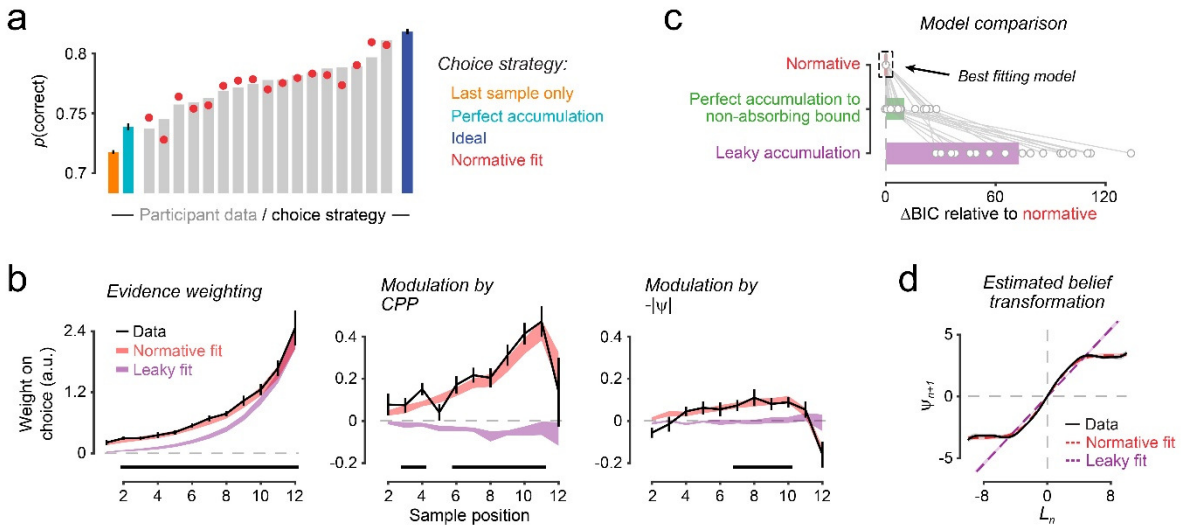



***Figure 2. Human task performance and model fits.*** *(**a**) Choice accuracies for 17 human participants (grey), normative model with perfect knowledge of task statistics (blue), and two suboptimal decision processes (perfect accumulation, cyan; and using only last evidence sample, orange). Red dots show accuracies of normative model fits. (**b**) Psychophysical kernels reflecting time-resolved weight of evidence on final choice (left), and modulation of this evidence weighting by sample-wise change-point probability (middle) and uncertainty (right). Black shows participants' data; associated significance bars highlight time points where weights differ from zero (p<0.05, two-tailed cluster-based permutation test). Red and purple shadings show fits of normative model and a leaky accumulator model, respectively. (**c**) Mean ΔBIC values for normative accumulation (red), perfect accumulation toward non-absorbing bound (green) and leaky accumulation (purple) models fit to the human choices, relative to the best-fitting model. Circles indicate individual participants. See Methods for additional model details. (**d**) Non-linearity in evidence accumulation estimated directly from choice data (black), and from fits of normative (red) and leaky (purple) accumulator models. All panels: error bars or shaded areas, s.e.m.*

## Selective motor preparatory cortical activity tracks adaptive evidence weighting

We first delineated motor preparatory activity in a separate task, in which participants prepared left- or right-handed button presses (instructed by an unambiguous cue) and executed that action after go signal (*Methods*). This yielded, as expected, a lateralized and sustained suppression of alpha- and beta-band (8-30 Hz) power overlying the motor cortex contralateral to the upcoming response (Figure 3a, left). We then used this preparatory activity to construct linear filters for isolating choice-specific motor preparatory activity during the decision-making task (Figure 3a, right; and *Methods*).

We found a characteristic build-up of motor preparatory activity during evidence accumulation (Figure 3b) also observed in previous work[8-10,15,16,30]. Critically, in our change-point task, this build-up activity closely mirrored the trajectories of the normative decision variable, but not the decision variable of a standard perfect accumulator model (Figure 3c). We quantified the relationship between the preparatory activity and inferred belief state by means of a linear model made up of the four key constituents of normative evidence accumulation identified earlier for each sample: prior belief ($\psi_n$), new evidence ($LLR_n$), and the interactions of new evidence with $CPP_n$ and $-|\psi|_n$ (*Methods*: **Model 3**). All four terms were encoded in the dynamics of motor preparatory activity following each sample onset ($\psi$: $p$=0.0006; $LLR$: $p$=0.0002; $LLR_n$*$CPP_n$: $p$=0.0009; $LLR_n$*$-|\psi|_n$: $p$=0.08; two-tailed cluster-based permutation tests; Figure 3d). The $CPP$ modulation was also stronger than the $-|\psi|$ modulation ($p$=0.002; two-tailed cluster-based permutation test). In sum, the dynamics of choice-specific motor preparatory activity exhibited the hallmark signatures of the normative decision variable in our change-point task.

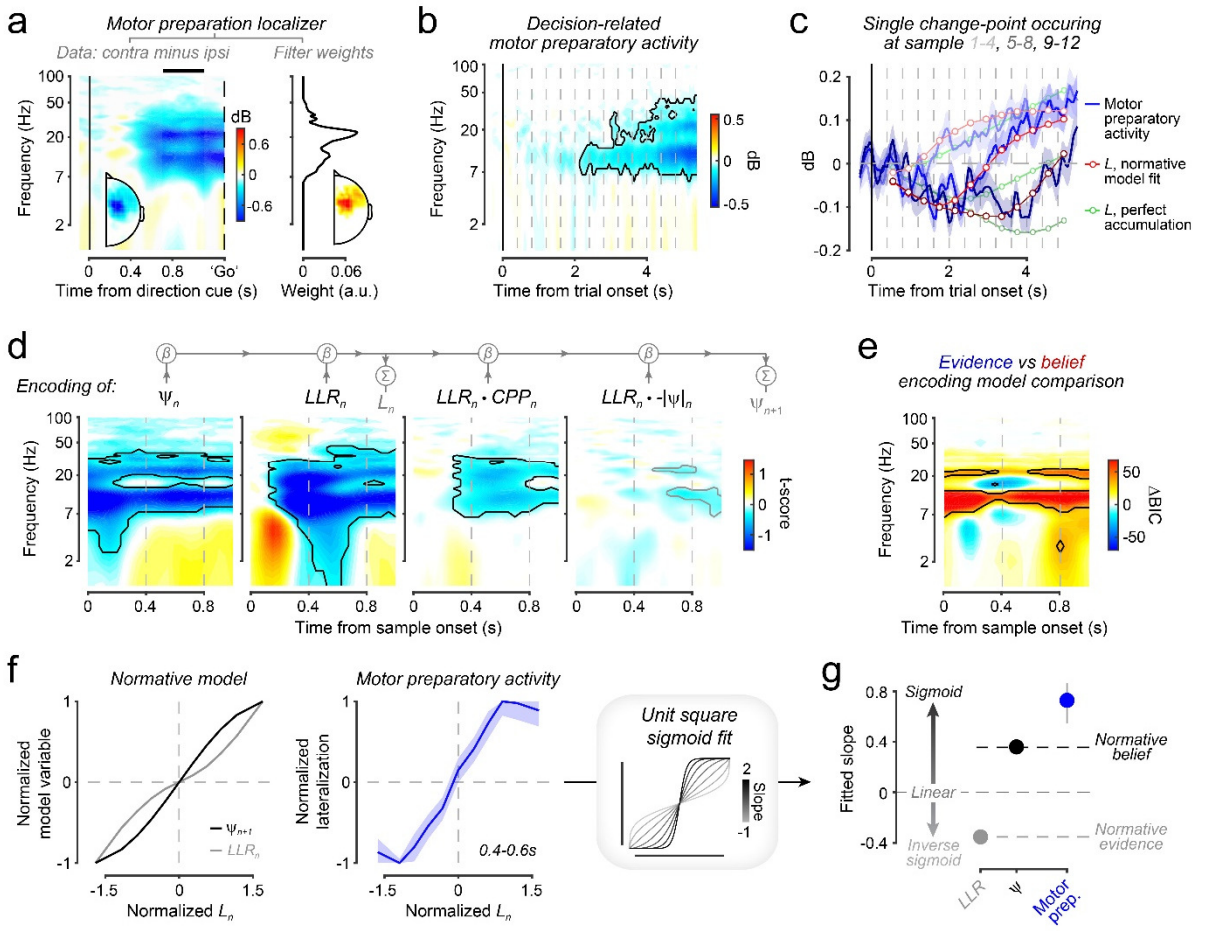



***Figure 3. Choice-specific motor preparation encodes all components of normative decision variable.*** *(**a**) Left: 3-dimensional cluster (time, frequency, space) of movement-specific (contra vs. ipsilateral to button press) MEG power lateralization during delayed hand movement task. Black bar, interval for derivation of spatial and spectral filters (shown on right). (**b**) Motor preparatory activity during decision-making task. Grey lines, sample onsets; contour, p<0.05 (two-tailed cluster-based permutation test). (**c**) Motor preparation signal integrated over frequencies for subsets of trials on which a single state change occurred, at different latencies (shaded areas, s.e.m.). Red traces, decision variable for same trials derived from normative model fits; green traces, decision variable derived from perfect linear accumulator model. Signs of motor signal and decision variables are here conditioned on generative state at end of trial. (**d**) Time-frequency maps of encoding of model variables in motor preparation signal that together approximate normative evidence accumulation. Black contours, p<0.05 (two-tailed cluster-based permutation test); gray contours, two largest sub-threshold clusters of uncertainty modulation (p<0.05, two-tailed t-test, uncorrected). (**e**) Group-level BIC difference scores reflecting relative goodness-of-fit of pure evidence ($LLR_n$) versus non-linearly transformed belief ($\psi_{n+1}$) encoding models fit to motor preparation signal. Contour, arbitrary threshold of ±35. (**f**) Relationships of normative decision variable ($L_n$) with other model variables ($\psi_{n+1}$ and $LLR_n$; left) and motor preparation signal (right; shaded area, bootstrap 95% CI). (**g**) Slopes of unit square sigmoids fit to relationships in panel f. Error bar, bootstrap 95% CI.*

This conclusion was corroborated by two further observations. First, formal comparison between 'pure' versions of the above regression model, containing only evidence encoding or only belief encoding (*Methods:* **Models 4-5**), showed that the dynamics of the motor preparatory activity was better explained by encoding of the non-linearly updated decision variable ($\psi_{n+1}$) than encoding of momentary evidence ($LLR_n$; Figure 3e, group-level BIC scores; $BIC_{dv}<BIC_{evidence}$ for 15 of 17 participants averaged over highlighted clusters). Second, the shape of the relationship between motor preparatory activity and the model-derived decision variable reflected another qualitative signature of normative evidence accumulation (Figure 3f). The normative model predicted that a neural signal tracking the decision variable after the non-linear transform, $\psi_{n+1}$ (i.e. prior for next updating step), should show a sigmoidal dependence on $L_n$ (for H=0.08; Figure 1b). By contrast, a neural signal only encoding

momentary evidence ($LLR_n$) should show an *inverse*-sigmoidal relationship to $L_n$ (Figure 3f, *left*). Fitting a unit square sigmoid (sigmoidal when slope>0, linear when slope≈0; inverse-sigmoidal when slope<0) to this 'belief encoding function' (i.e. dependence on $L_n$) of the motor preparatory activity (Figure 3f, right; *Methods*) yielded a positive slope (Figure 3g; $p<0.001$, two-tailed bootstrap hypothesis test against zero), establishing a sigmoidal relationship. Indeed, the encoding function was steeper for the motor preparatory activity than for the normative model fit prediction ($p<0.001$, two-tailed bootstrap hypothesis test against slope for normative model). Thus, motor preparatory activity in human cortex encoded the decision variable in a more categorical fashion than prescribed by the normative model.

## Encoding of normative computational variables in distinct frequency bands and cortical regions

The previous analyses focused on the evolution of the action plan during decision formation. Decision formation entails the selective signal flow across a large network of cortical regions, from sensory to association to motor cortex[39]. While these regions have often been conceived of as being specialized in encoding either the sensory evidence or decision variables (action plans), recent empirical observations[17,36,40,41] and theoretical insights[25] suggest that each cortical processing stage may encode mixtures of sensory evidence and decision variables, both of which are represented in a distributed fashion across cortex. To illuminate the encoding of computational variables across the cortical pathway transforming sensory input into action choice, we reconstructed the dynamics of spatially-specific (i.e., hemisphere-lateralized, exploiting the spatial arrangement of both sensory evidence and motor action) activity at the source level (see *Methods*). Based on previous functional MRI work, we focused on 13 topographically organized regions of interest (ROIs, *Methods* and Figure 4): nine clusters of retinotopic visual field maps in occipital, temporal, and intraparietal cortex[42] and four regions of motor and intraparietal cortex exhibiting hand movement-specific lateralization of activity[9,28,43]. We then fitted the above-described model (**Model 3**) comprising all four components of normative evidence accumulation identified earlier to the frequency-specific activity lateralization, separately for each ROI.

This revealed several differences between the functional properties of different cortical regions and different frequency bands. First, alpha (8-14 Hz) and beta (16-30 Hz) activity lateralization in several hand-movement selective regions (IPS/PCeS, PMd/v, M1, but not aIPS) encoded the decision-relevant computational variables in the same manner as the sensor-level motor preparatory signal from Figure 3: sustained encoding of *LLR* (Figure 4a) and of prior belief ($\psi$, Figure 4b), as well as modulation of *LLR* encoding by *CPP* and (weakly) uncertainty ($-|\psi|$, Figure 4b). The joint encoding of all four variables within overlapping frequency bands, time intervals, and cortical regions, was consistent with representation of the normative decision variable in the alpha-/beta-activity of these (pre-) motor cortical regions (compare Figure 1h). We show the encoding of $\psi$ and the two modulation terms pooled across IPS/PCeS, PMd/v, and M1 for compactness, but it was significant, with analogous spectral and temporal profiles, for each region individually (data not shown).

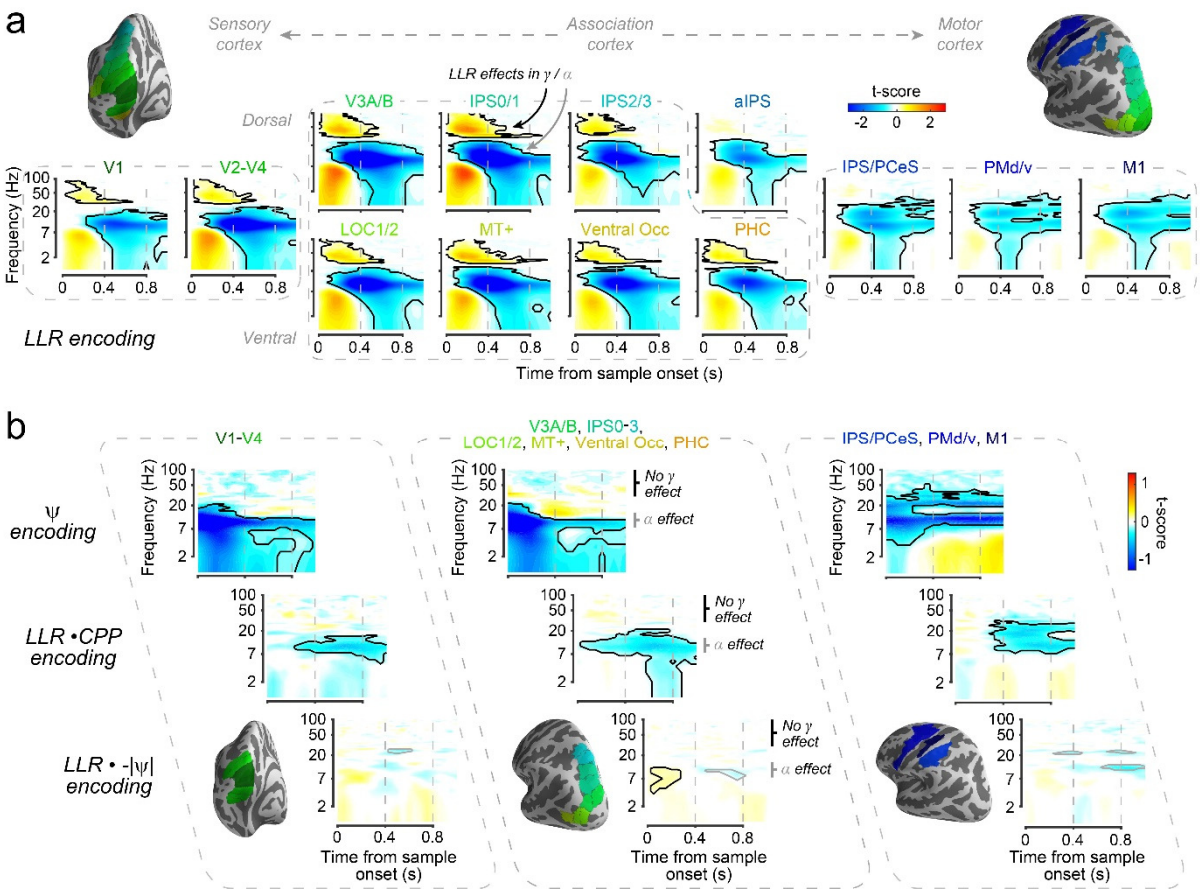



***Figure 4. Frequency-specific encoding of evidence and decision variable along visuo-motor cortical pathway.*** *(**a**) Time-frequency maps of evidence (LLR) encoding across anatomical ROIs in visual, temporal, posterior parietal and motor/premotor cortices (insets) known to exhibit either retinotopic organization or hand movement-selective activity lateralization. Encoding maps were derived through fits of a linear regression model (**Model 3**, see Methods) to hemisphere-lateralized, spectrally-resolved MEG activity in each ROI. Note robust positive LLR encoding in gamma-band (35-100 Hz; highlighted via black arrow for IPS0/1) in all visual field maps but absent in aIPS, IPS/PCeS, PMd/v, M1, and robust negative LLR encoding in alpha-band (8-14 Hz; highlighted via gray arrow for IPS0/1) in all ROIs. (**b**) Time-frequency maps reflecting encoding of other constituent components of adaptive evidence accumulation (same as Figure 3d), averaged across three sets of ROIs (insets) encompassing occipital visual field maps (V1-V4), temporal and parietal visual field maps (LOC1/2, MT+, Ventral Occ, PHC, V3A/B, IPS0-3), and anterior intraparietal / motor cortices (IPS/PCeS, PMd/v, M1). Note lack of any reliable encoding in gamma-band but negative encoding of all components in alpha-band in visual field maps (highlighted explicitly for temporal/parietal cluster). aIPS, which lacks both the transient gamma-band response of posterior cortical ROIs and the sustained beta-band (16-30 Hz) response of anterior ROIs (panel a), was omitted from the clustering; though panel b (middle) was qualitatively highly similar with aIPS included. Black contours, p<0.05 (two-tailed cluster-based permutation test); gray contours, largest sub-threshold clusters of LLR·-|ψ| effect, (p<0.05, two-tailed t-test).*

By contrast, the visual field maps, exhibited a feature not present in the decision-encoding regions: short-latency and transient encoding of the evidence (*LLR*) in the 'gamma' frequency band (35-100 Hz, Figure 4a), which was not accompanied by encoding of prior or modulations of *LLR* encoding in the same band (Figure 4b). The observation that the *LLR*-related gamma-band responses were robust (in particular in early IPS regions, Figure 4a) and insensitive to the latter three model terms (Figure 4b) is consistent with a 'pure' encoding of sensory evidence in visual cortical gamma-band activity.

Finally, the lateralization of alpha-band activity in the visual field maps exhibited a profile roughly analogous to the signature of the normative decision variable evident in the beta- and alpha-band lateralization of downstream decision-encoding regions. Visual cortical alpha-

band activity reliably encoded *LLR* (Figure 4a), prior belief ($\psi$, Figure 4b), and the interaction between *LLR* and *CPP* (Figure 4b). The similarity of encoding profiles for alpha-band activity across the cortical pathway indicates that signatures of the normative decision variable were widely distributed across cortex.

**Reverse hierarchy of decision signals across cortex**

To further dissect the nature of decision variable encoding in alpha-band activity across the visuo-motor cortical pathway, we focused on the hand movement-specific intraparietal and motor regions (aIPS, IPS/PCeS, PMd/v, M1) along with a hierarchically ordered set of visual field maps (V1, V2-4, V3A/B, IPS0/1, IPS2/3)[44,45]. For each ROI, we examined three functional properties of the alpha-band modulations induced by individual samples: (i) the slope of belief encoding functions (Figure 5a,b); (ii) the latency of evidence encoding (Figure 5d); and (iii) the timescale of evidence encoding (Figure 5e).

The dynamics of the decision-encoding regions (M1, PMd/v, IPS/PCeS) were consistent with robust encoding of the normative decision variable (Figure 5a-e). During evidence accumulation, belief encoding functions of all three ROIs had positive (slopes>0: all $p<0.002$, two-tailed bootstrap hypothesis test here and for all comparisons in this section) and steeper slopes than those of the visual field maps (all $p<0.001$ for comparisons with other ROIs excluding V1; all $p<0.04$ for comparisons with V1). They also had especially slow evidence encoding timescales ($p<0.005$ for all comparisons with other ROIs), and intermediate latencies ($p<0.001$ for comparisons with the ROI with the earliest latency, IPS2/3; no reliable differences with V3A/B, IPS0/1 or aIPS). Of the three regions with strong decision-variable encoding (Figure 4), IPS/PCeS had the fastest timescale ($p<0.001$ *vs*. PMd/v and M1) and was the only region with a belief encoding function that was indistinguishable from the normative model fits. Thus, encoding of the decision variable in action-planning cortical regions progresses from a veridical representation in parietal cortex (IPS/PCeS) to a categorical representation in frontal cortex (PMd/v, M1).

Strikingly, the above functional properties exhibited a reverse gradient across the hierarchy of visual cortical regions: V1 more strongly resembled the decision-encoding cortical regions than higher-tier, extrastriate visual (V2-V4, V3A/B) or intraparietal regions (IPS0/1 and IPS2/3; Figure 5a-e). Only IPS regions showed the inverse-sigmoidal encoding functions expected for veridical evidence encoding (slopes<0, Figure 5b; IPS0/1: $p=0.008$; IPS2/3: $p=0.004$; aIPS: $p=0.034$). The encoding function for V1, by contrast, lay between those expected of pure evidence and decision-variable encoders, and its slope was more positive than those of the IPS areas with inverse-sigmoid functions (all $p<0.03$, Figure 5b).

The time course of evidence encoding (Figure 5c) also had the longest latency (Figure 5d) and slowest timescale (Figure 5e) in V1 compared to all higher-tier visual cortical areas V2-4, V3A/B, IPS0/1, or IPS2/3 (all $p<0.001$). In fact, the V1 latency was even longer than the latencies measured in all downstream regions encoding the normative decision variable (IPS/PCeS, PMd/v, M1; all $p<0.001$). The decrease of evidence encoding timescales across the visual cortical hierarchy (from V1 to IPS3) stands in sharp contrast to the hierarchical increase of the relative contributions of slow vs. fast components of activity fluctuations in the pre-trial baseline interval (Figure 5f, Supplementary Figure 6; *Methods*). This latter result is in line with previous monkey work[46] and establishes that the 'reverse hierarchy' evident across panels a-e of Figure 5 was specific to the dynamics of task-related activity.

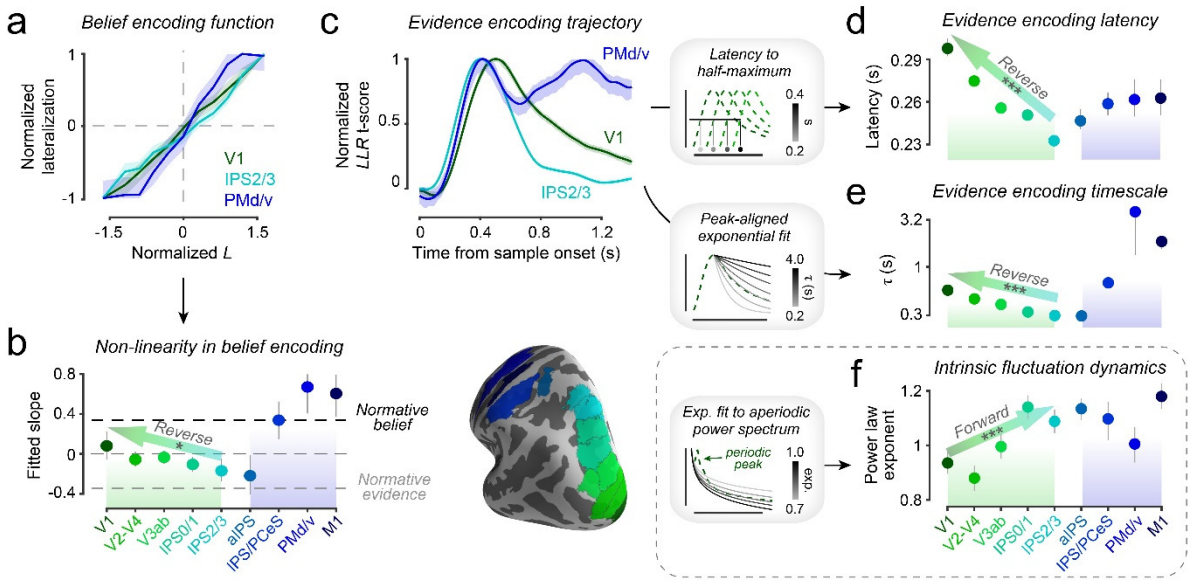



***Figure 5. Properties of decision signals across visual cortical hierarchy.*** *(**a**) Relationships of normative decision variable (L) with 8-14 Hz lateralization signal in three example ROIs. (**b**) Slopes of unit square sigmoids fit to these relationships for a dorsal set of visual cortical field maps, plus anterior intraparietal and (pre-)motor cortical regions showing robust encoding of the normative decision variable (inset at center). Regions ordered by position along the visuo-motor pathway. Lower and upper dashed lines, slopes of relationships with momentary evidence (LLR) and transformed belief (ψ) in normative model, respectively. (**c**) Time-courses of normalized LLR encoding in 8-14 Hz lateralization signal for three example ROIs. (**d**) Latencies to half maximum of LLR encoding traces for the same array of ROIs from panel b. (**e**) Timescales of decaying exponentials fit to the normalized, peak-aligned LLR encoding traces for same ROIs from panel b. (**f**) Power law scaling exponents of intrinsic activity fluctuations during baseline interval for same selection of ROIs. All panels: error bars or shaded areas, 95% confidence intervals. Arrows and asterisks in panels b and d-f, linear trend across dorsal visual cortical hierarchy (*p<0.05, ***p<0.001, two-tailed bootstrap hypothesis test for panels b, d and e, two-tailed permutation test for panel f, Methods); 'forward' and 'reverse' labels indicate direction of hierarchy inferred from measured slopes of parameters across ROIs.*

Monkey[36] and human[47] physiology has inferred decision-related feedback signaling from the co-variation between fluctuations of neural activity in visual cortex and behavioral choice. We computed these fluctuations as the residuals over and above activity explained by decision-relevant computational variables (**Model 6**, *Methods*). In downstream regions (IPS/PCeS, PMd/v, M1), fluctuations toward the end of the trial in both the beta- and alpha-bands were robustly predictive of choice. In visual cortex, by contrast, only alpha-band fluctuations in V1 (and, weakly, in V2-4) were choice-predictive (Figure 6).

In sum, we found stronger decision encoding when progressing backwards across the visual cortical hierarchy from IPS to V1. All effects reported in this section can be explained by a task-dependent propagation of choice-specific feedback signals from (pre-)motor cortical regions encoding the normative decision variable to early visual cortex: Such feedback signals rendered evidence encoding in V1 similarly sustained (Figure 5e), with even longer latency (Figure 5d) than in (pre-)motor regions; and they rendered fluctuations of V1 activity similarly choice-predictive as in (pre-)motor regions (Figure 6).

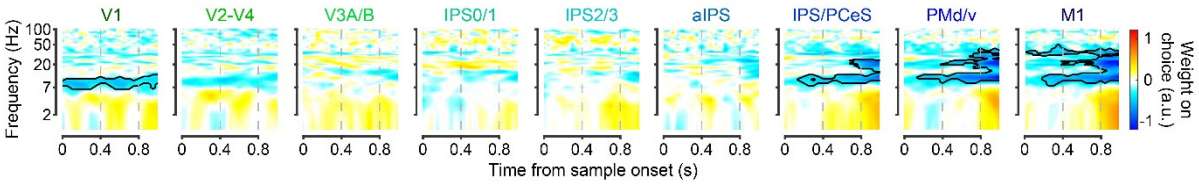



***Figure 6. Dynamics of choice-predictive fluctuations in activity across cortical regions.*** *Weight of residual fluctuations in the lateralization signal on final choice, over and above the weight on choice exerted by key model variables identified in previous analyses. Effects shown are averaged over final 3 evidence samples in the trial. Contours, p<0.05 (two-tailed cluster-based permutation test).*

## Adaptive evidence accumulation emerges from cortical attractor dynamics

Which circuit mechanisms gave rise to the behavioral and cortical signatures of normative evidence accumulation we uncovered here? We simulated an influential biophysical model of a single cortical decision circuit[26] (e.g. lateral intraparietal area of the macaque; Figure 7a) to illuminate these mechanisms. In the circuit model, strong recurrent excitation within two choice-selective populations of excitatory neurons, coupled with feedback inhibition, instantiate winner-take-all attractor dynamics (*Methods*). These dynamics are inherently non-linear: once the system reaches an attractor state, there is a saturation of accumulated evidence in the face of consecutive consistent evidence. They can also give rise to changes of mind (i.e. sign flip of model decision variable, computed as the difference in activity between both populations), particularly in response to new evidence that strongly conflicts with the current belief state (Figure 7b).

Our simulations showed that the sensitivity to surprising evidence observed in our behavioral and neural measurements is a natural feature of this model (Figure 7c-f). The model's choice behavior exhibited all three observed behavioral signatures of adaptive evidence weighting, including the strong sensitivity to *CPP* (Figure 7c; compare Figure 2b; see Supplementary Figure 7 for an assessment of the boundary conditions for this behavior). The dynamics of the model's decision variable also closely resembled the measured dynamics of cortical motor preparatory activity (shown for M1 in Figure 7d), including the strong sensitivity to change-points. The belief encoding function for the biophysical model decision variable (Figure 7e) resembled those of several cortical regions, with a slope that fell between IPS/PCeS (matching normative predictions) and M1 (Figure 7f). In sum, key features of normative and human evidence accumulation in a changing environment may be a natural product of the properties of cortical decision circuits.

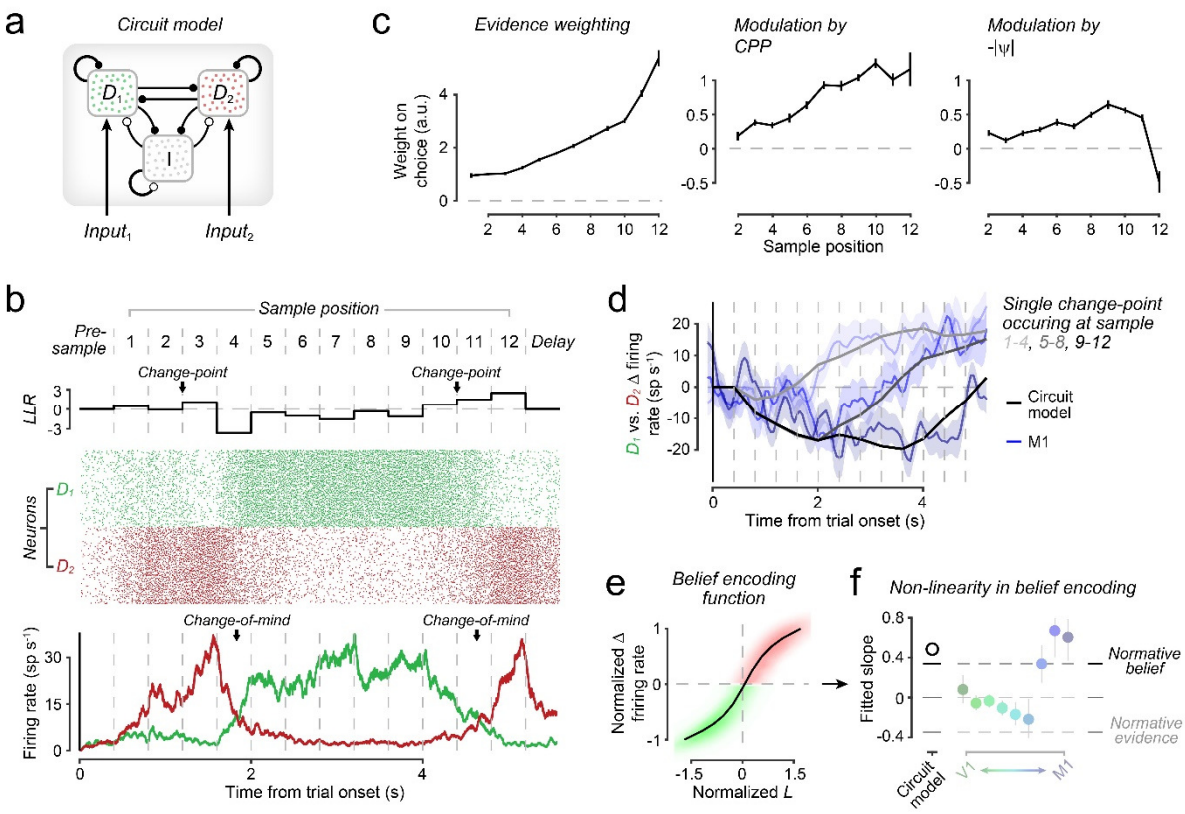



***Figure 7. Cortical circuit model reproduces signatures of normative evidence accumulation.*** *(**a**) Schematic of circuit model. Circuit includes choice-selective populations of excitatory neurons $D_1$ (left choice, green) and $D_2$ (right choice, red), and a shared population of inhibitory interneurons (I, grey). Choice-selective populations receive input currents determined by the sample-wise LLR (Methods). (**b**) Response of choice-selective populations to an example trial. Top, LLRs of each sample; middle, raster plot of spiking activity from all neurons in $D_1$ and $D_2$; bottom, instantaneous population firing rates. Arrows in top and bottom panels highlight changes in environmental state and corresponding changes-of-mind, respectively. (**c**) Psychophysical kernels produced by circuit model, reflecting the time-resolved weight of evidence on the model's final choice and the modulation of evidence weighting by sample-wise change-point probability and uncertainty. (**d**) Difference in firing rates between choice-selective populations in circuit model for subsets of trials on which only a single state change occurred, at different latencies. Blue traces, hemisphere-lateralized 8-14 Hz M1 activity for same trials. (**e**) Shape of the relationship between decision variables from normative fits and circuit model (black, see main text). Red and green shading, Gaussian kernel density estimation of the distribution of sample-wise values. (**f**) Left (black): Slope of unit square sigmoid fit to relationship in panel e. Right (pale colors): fitted slopes for 8-14 Hz MEG signal from cortical regions spanning the visuo-motor pathway (reproduced from Figure 5b for comparison).*

## Phasic arousal modulates evidence accumulation by boosting evidence encoding

The dynamical regimes of cortical circuits are continuously shaped by brainstem arousal systems[29,48-51], which may, in turn, modulate inference and decision-making. Specifically, arousal signals may dynamically adjust the weight that new evidence exerts on existing belief states, particularly in situations of high surprise or *CPP*[19,34,35,52,53]. We used dilations of the pupil to assess the involvement of the brain's arousal systems in adaptive evidence accumulation and the underlying cortical dynamics. Non-luminance-mediated pupil responses have been linked to the activity of noradrenergic[28,54-57], cholinergic[28,57] and dopaminergic[28] systems. We quantified rapid (phasic) arousal responses to individual samples in terms of the first derivative (i.e. rate of change) of the sample-evoked pupil responses (Figure 8a), because (i) this measure has higher temporal precision than the overall response, and (ii) it is strongly

correlated with noradrenaline release[57], a neuromodulator which has been implicated in surprise-driven modulation of belief updating[34,35].

Samples strongly indicative of a change in environmental state evoked strong pupil responses (Figure 7b-c). This effect was apparent when binning trials based on the timing of single change-points and comparing against trials without change-points (Figure 8b). We quantified this effect by fitting a linear model consisting of *CPP*, *-|ψ|*, and absolute evidence strength (|*LLR*|) associated with each sample to the corresponding pupil response (first derivative; **Model 7**, *Methods*). In the current context, absolute evidence strength scales negatively with the probability of encountering a given stimulus independent of the current belief, thus capturing a low-level form of surprise. We found a robust positive contribution of *CPP* ($p<0.0001$, two-tailed cluster-based permutation test; Figure 8c) but neither -|ψ| ($p=0.27$) nor absolute evidence strength ($p=0.40$). The sensitivity to *CPP* was also greater than the sensitivity to a different measure of surprise that is determined by both low-level and high-level components (termed 'Shannon surprise' in Supplementary Figure 2). Thus, in our task, phasic arousal was specifically recruited by a high-level form of surprise reflecting the inconsistency between prior belief and new evidence (see *Discussion*).

The sample-to-sample fluctuations in the evoked arousal response predicted an 'up-weighting' of the impact of the associated evidence sample on choice (Figure 8d-f). Variations in pupil responses beyond those explained by the above computational variables (**Model 8**, *Methods*) exhibited a positive, multiplicative interaction with *LLR* in its impact on choice (Figure 8d; $p=0.04$, two-tailed cluster-based permutation test). The magnitude of this effect also correlated with a participant-specific gain parameter, estimated in our normative model fits, which quantified the weighting of belief-inconsistent evidence beyond the upweighting entailed in the normative model (Figure 8e; peak Pearson $r_{17}=0.59$, uncorrected $p=0.012$, two-tailed).

How was the up-weighting of the impact of evidence on choice by phasic arousal instantiated in the sensory-motor cortical pathway? We expanded the previous model-based decomposition of cortical dynamics (**Model 3**; Figures 3, 4) by an additional term that reflected the modulation of evidence encoding by the associated pupil response (**Model 9**, *Methods*). This indicated a selective enhancement of *LLR* encoding that was restricted to the alpha band within visual cortical regions: V1, V2-4, V3A/B, IPS0/1, and IPS2/3 (all $p\leq0.01$, two-tailed cluster-based permutation tests; Figure 8f). This modulatory effect was not present in any of the downstream decision-encoding regions (all $p>0.4$; Figure 8f). Thus, phasic arousal dynamically modulated evidence accumulation, and ultimately choice, through strengthening the neural encoding of surprising evidence in visual and intraparietal cortex.

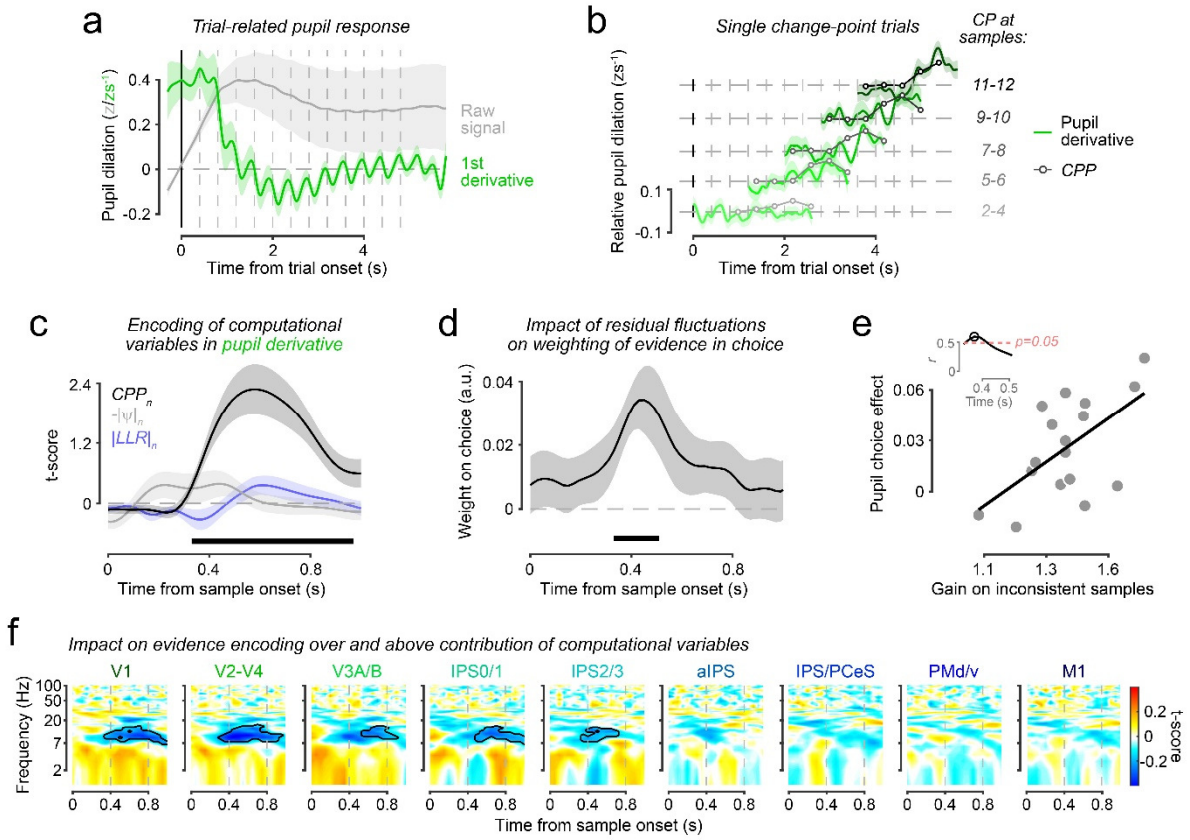



***Figure 8. Arousal responses to perceived state changes up-weight the associated evidence in brain and behavior.*** *(**a**) Average trial-related pupil response (grey) and its first derivative (green). Dashed grey vertical lines, sample onsets. (**b**) Pupil derivative for subsets of trials where only single state changes occurred at different latencies, relative to pupil derivative on trials without any state change. Greyscale traces, relative change-point probability for same trials from normative model fits. For clarity, each trace truncated from 3 samples before to 1 sample after average time of state change. (**c**) Encoding of sample-wise change-point probability, uncertainty and absolute evidence strength in pupil derivative. (**d**) Multiplicative effect of intrinsic pupil derivative fluctuations on weighting of evidence samples in final choice, over and above the weight on choice exerted by key model variables. (**e**) Between-participant correlation of pupil choice effect (panel d) with fitted model parameter reflecting additional weight ascribed to new evidence that was inconsistent with belief. Inset, time-resolved Pearson correlation coefficients spanning significant temporal cluster from panel d; highlighted time-point used to generate scatterplot. (**f**) Multiplicative effect of sample-wise pupil response on encoding of new evidence in the lateralized, source-localized MEG signal. Contours, p<0.05 (two-tailed cluster-based permutation test). Shaded areas in panels a-d indicate s.e.m. Significance bars in panels c and d, p<0.05 (two-tailed cluster-based permutation test).*

## Discussion

Higher brain functions such as perceptual decision-making have commonly been assessed at different levels of analysis[33]. Work on the computational and/or algorithmic levels has developed normative models like the one we used here[20] (Figure 1). Work on the implementation level has dissected the neural circuit interactions that realize cognitive behavior[24-26]. Here, we developed a mechanistically-detailed, neurobiological account of how normative computations for perceptual decision-making in a changing world are approximated by neural circuit interactions. Our account starts from the realization that all natural environments are characterized by the occurrence of hidden changes. Normative models for changing environments prescribe non-linear accumulation of evidence, with an ongoing, adaptive modulation of evidence weighting. Precisely such modulations of evidence weighting are evident in human behavior, and in neural signals in multiple cortical areas. These

diagnostic signatures of normative evidence accumulation emerge naturally from the dynamics of a detailed (spiking neuron) model of cortical circuits, and they are shaped by pupil-linked arousal. Our computational analysis of neural signals across the visual cortical hierarchy uncovered a distributed encoding of the normative decision variable, along with the feedback of decision signals from downstream regions to primary visual cortex.

### Nature and functional role of change-point probability and uncertainty

In the normative model, the strength of the non-linear modulations of evidence accumulation that we have illuminated varies across environments (Supplementary Figure 1). They are negligible in contexts dominated by high external or internal noise and/or a high expected rate of change (i.e. contexts that preclude the formation of strong belief states). Indeed, a linear accumulator model provided good fits to the behavior of rats with high levels of internal noise performing a perceptual choice task in a changing environment[38]. However, our simulations (Supplementary Figure 1) showed that the non-linear modulations of the accumulation entailed in the normative model play a key role in optimizing behavior across a wide range of contexts characteristic of natural environments.

We reiterate that sensitivities to change-point probability and uncertainty are *implicit* properties of the normative evidence accumulation algorithm – neither variable plays an explicit role in the computation[20]. By contrast, other models of probabilistic inference make explicit use of surprise and uncertainty measures[35,58,59]. Our rationale for the use of both variables was that twofold: (i) they yielded mechanistically-interpretable and diagnostic signatures in behavior (psychophysical kernels) and modulations of selective cortical population signals; and (ii) change-point probability was robustly and specifically encoded in pupil-linked arousal responses measured in our participants. Both aspects were central to the development of our implementation-level account of adaptive evidence accumulation.

Change-point probability differs from the more general measure of Shannon surprise (i.e., negative log probability of an observation) which, in some models, serves as the objective function to be minimized by the inference algorithm[59]. We found that, for our task context, Shannon surprise could be factorized into two components, one of which was change-point probability (Supplementary Figure 2a). Of these components, only change-point probability exerted a strong, positive modulatory effect on belief updating in the normative model (Supplementary Figure 2c,d) as well as on measured pupil responses of our participants (Supplementary Figure 2e). Change-point probability also behaved very similarly to another surprise measure (Supplementary Figure 2b), which has been postulated to drive phasic noradrenaline release during inference[34] and exhibited a similar effect on sample-evoked pupil responses (data not shown).

### Relationship to accounts of statistical learning

Dissecting perceptual evidence accumulation in the above manner revealed striking parallels with insights from previous studies of statistical learning, in which observations and associated actions unfolded at timescales at least an order of magnitude slower than in our task[19,58]. This previous work has shown that humans adaptively tune their learning rate as a function of two key variables for probabilistic inference in a changing world: surprise (or change-point probability) and uncertainty[58,60,61]. This adjustment of learning rate is analogous to the 'up-weighting' effect of these variables on perceptual evidence accumulation shown here and raises the intriguing possibility of shared neural mechanisms. Indeed, a similar role for pupil-

linked arousal systems to what that we describe (see below) has previously been reported during slow probabilistic inference[35,52]. These computational and apparent mechanistic analogies encourage the conceptualization of perceptual evidence accumulation as a process of rapidly learning about the state of the world.

Critically, the previous work on learning has focused on neural substrates encoding the above statistical variables, which are unsigned with respect to the available choice alternatives[53,58,60,61]. This work has not illuminated how surprise or uncertainty modulate the selective neural signals encoding the (signed) variables prior, evidence, and decision variable, which underlie the transformation of sensory input to motor action within the cortical decision pathways studied here. Our approach sets the stage for unravelling similar large-scale cortical dynamics underlying probabilistic reasoning in the domain of learning. Of particular interest will be the question of whether there are limits to the timescales of the cortical circuit operations characterized here for perceptual evidence accumulation, that prevent generalization to long-timescale learning processes.

### Frequency-specific encoding of evidence and decision variables

Our model-based analyses revealed frequency specific encoding of several computational variables. The motivation behind this frequency-specific analysis approach was two-fold. First, we assumed that neural representations of evidence, decision variable and their modulations by change-point probability and uncertainty emerged from intra-cortical, recurrent network interactions[26] that produce timing jitter[62]. Their signatures in cortical population dynamics should therefore not be precisely phase-locked to the onset of sensory evidence samples. Spectral analysis was well-suited for detecting non-phase-locked activity components. Second, different frequency bands are differentially involved in the encoding of computational quantities (sometimes with opposite sign)[62], and mounting evidence suggests that the gamma- and alpha-bands provide separate channels for feedforward vs. feedback signal flow across the cortical hierarchy[45,63,64]. We therefore reasoned that a frequency-resolved encoding analysis should increase sensitivity.

Our results corroborate this reasoning. The encoding of all computational variables considered here occurred in confined frequency bands (Figures 3-6). In visual cortical areas, we found (i) robust, short-latency, and pure (i.e., without encoding of other variables) evidence encoding in the gamma-band, most strongly in IPS regions (Figure 4); and (ii) opposite-polarity encoding of the normative decision variable in the alpha-band, most strongly and sustained in early visual cortex (Figures 4, 5). This separation is consistent with the multiplexing of feedforward vs. feedback signaling in distinct frequency channels found in previous analyses of inter-regional interactions[45,63,64]. Our findings link these physiological observations to dynamic belief updating.

### Distributed cortical attractor circuits for normative evidence accumulation

Our findings shed new light on the current quest for the substrate and mechanism of the evidence accumulator[6]. We identified clear signatures of decision variable encoding in multiple stages of the cortical sensory-motor pathway: (pre-)motor cortex, posterior parietal cortex (IPS/PCeS), but most remarkably, also V1, which has so far been conceived of as a mere input stage to perceptual decision computations. These results indicate that the cortical dynamics underlying evidence accumulation are distributed across several, recurrently connected cortical areas[24,25]. This idea may explain why behavior can be robust to inactivation

of single nodes of the large-scale network[65,66] (but see ref. [14]). The particularly pronounced decision feedback signatures in V1 alpha-band activity, more evident there than in any other visual cortical area (Figures 5, 6), may point to a special role for V1 in the large-scale evidence accumulation process.

Several insights indicate that attractor dynamics in a fronto-parietal network including intraparietal[24] and (pre-)motor[67,68] cortices provide a basis for the formation of decision states and their maintenance in working memory[26,69,70]. We found that a cortical circuit model of decision-making with such dynamics between choice-selective populations[26] reproduced the key characteristics of both behavior and preparatory activity in intraparietal (IPS/PCeS), pre-motor (PMd/v) and primary motor (M1) cortices. Previous work on cortical attractor dynamics has reproduced observed neural data under biologically plausible circuit configurations. Our current work uncovered the normative function of the non-linearity inherent in such decision circuits in natural (i.e. changing) environments: sensitivity to change-point probability and uncertainty. These features are a natural consequence of the in-built circuit dynamics. The network is maximally sensitive to input when at the 'saddle' point between attractor states (i.e. when uncertain). It also prevents runaway excitation in favor of a given choice alternative during environmental stability, producing insensitivity to consistent evidence but retaining sensitivity to strongly inconsistent evidence (i.e. to samples with high change-point probability). Together, these properties adapt the evidence accumulation to possible changes in environmental state. These insights forge a strong link between computational- and implementation-level perspectives on decision-making.

### Interplay of arousal systems and cortical circuit dynamics

The observation that pupil-linked, phasic arousal mediated (part of) the non-linearity in the evidence accumulation process is broadly consistent with prominent computational frameworks of neuromodulatory brainstem systems, in particular the locus coeruleus (LC) noradrenaline system. These systems can modulate cortical network dynamics on the fly[27,48]. An influential account holds that surprise-related phasic LC responses cause a shift toward more 'bottom-up' relative to 'top-down' signaling across the cortical hierarchy, and thus a greater impact of new evidence on belief[19,34]. In line with this idea, we found that pupil dilations driven by change-point probability predicted an up-weighting of the impact of the corresponding evidence sample on choice. Previous work has observed similar effects during slow, feedback-driven predictive inference from one choice to the next[35,52]. Here, we show that analogous effects are at play at a more rapid timescale, within the formation of individual decisions.

Critically, we also uncovered the underlying modulations of neural dynamics in the cortical visuo-motor pathway. Although arousal is often associated with widespread changes in cortical state[71], we found that phasic, pupil-linked arousal predicted a spatially-specific modulation of evidence encoding across the visual cortical hierarchy (most pronounced in early visual cortex), but not in downstream regions engaged in motor preparation. These selective modulations matched signatures of decision-related feedback (restricted to alpha-band and occurring late after evidence onset), argueing against a simple strengthening of evidence encoding that was indexed by visual cortical gamma-band activity (Figure 4). One possibility is that phasic arousal stabilizes the new network state induced by the surprising evidence. This would be consistent with the timing of surprise-related brainstem responses, which might be driven by a preceding surprise detection mechanism in cortex[27,60,72] and take time to impact cortical networks.

How can the role of phasic arousal be reconciled with the identification of cortical circuit dynamics as the mechanism for normative evidence accumulation? We propose that the core function of brainstem arousal systems is to maintain appropriate *adaptivity* in the cortical inference machinery across different environmental contexts, which is well-established for human behavior[58,73] but mechanistically not well understood. Here, we probed a single level of environmental volatility, and hence, putatively a single point in the space of neuromodulatory interactions with cortex. Contextual variability would require appropriate adjustment of cortical attractor dynamics (i.e. energy landscape) if accurate decision-making is to be maintained, and altered neuromodulation at different timescales may play a key role in implementing this adjustment[27,29,74]. Neuromodulatory signals may alter the excitation-inhibition balance in cortical circuits[50,51], thus translating into different dynamical regimes. These ideas can be explored through manipulations of environmental statistics and neuromodulatory systems in future work.

## Acknowledgements

We thank Klaus Wimmer for discussion and advice on cortical circuit model and Florent Meyniel for detailed comments on the manuscript.

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – DO 1240/3-1, DO 1240/4-1, and SFB 936 - Projekt-Nr. A7.

## Author contributions

PRM: Conceptualization, Methodology, Investigation, Software, Formal analysis, Visualization, Writing – original draft, Writing – review and editing; NW: Conceptualization, Methodology, Software, Writing – review and editing; DCHB: Investigation, Formal analysis; GPO: Software, Formal analysis, Writing – review and editing; THD: Conceptualization, Methodology, Resources, Writing—original draft, Writing—review and editing, Supervision.

## Data availability

Raw behavioral data and preprocessed MEG/pupil data will be made available on Figshare upon acceptance.

## Code availability

All analysis code to reproduce the reported results and figures from the shared data will be made available on GitHub (https://github.com/DonnerLab) upon acceptance.

## Methods

### Participants

Seventeen human participants (mean ± s.d. age of 28.22 ± 3.89 years, range 23–36; 11 females), including the first author of this manuscript, took part in the study after providing written informed consent. All had normal or corrected-to-normal vision and no history of

psychiatric or neurological diagnosis. The experiment was approved by the ethics committee of the Hamburg Medical Association and conducted in accordance with the Declaration of Helsinki. Each participant completed one training session (120 min), three (16 participants) or four (1 participant, the first author) sessions of the main experiment (about 150 min each), and one session to obtain a structural MRI (30 min). Participants received the following remuneration for their participation: 9.50 €/h of testing plus a study completion bonus of 40 € and a performance-dependent bonus ranging between 0-40 €. The performance-dependent bonus was determined upon study completion by drawing four of their task blocks at random. The participant received €10 for each block in which their choice accuracy exceeded a criterion of 75%.

## Main behavioral task

### *Task design*

The main task was a two-alternative forced choice discrimination task, in which the generative task state $S$={*left*, *right*} change unpredictably over time (Figure 1a). On each trial of the task, participants viewed a sequence of evidence samples consisting of small checkerboard patches (see below: *Stimuli*). Samples were presented for 300 ms each, with sample-onset asynchrony of 400 ms. Samples were centered on spatial locations (specifically, polar angles) $x_1,\ldots,x_n$ drawn from one of two probability distributions $p(x|S)$. These distributions $p(x|S)$ were truncated Gaussians, with matched variance ($\sigma_{left}=\sigma_{right}=29°$), means symmetric with respect to the vertical midline ($\mu_{left}=-17°$, $\mu_{right}=+17°$), and truncated at -90° (+90°). In instances where a drawn $x_i$ was <-90° (>+90°), it was replaced with -90° (+90°). The generative state at the start of each trial (i.e., the distribution generating the first sample), was chosen at random. After the presentation of each sample, $S$ could change with a fixed probability, or hazard rate: $H=p(S_n=right|S_{n-1}=left)=p(S_n=left|S_{n-1}=right)=0.08$. Participants were asked to maintain fixation at a centrally presented mark (see below: *Stimuli*) throughout the sample sequence, monitor all samples, and report $S$ at the *end* of the sequence. That is, participants needed to infer which of the two probability distributions generated the position of the final sample.

The majority (75%) of sequences in each block of trials contained 12 samples. The lengths of the remaining sequences (25%, randomly distributed throughout each block) were uniformly distributed between 2 and 11 samples. Thus, the sequence durations ranged between 0.8 s (2 samples) and 4.8 s (12 samples). The shorter sequences were introduced in order to discourage participants from ignoring the early samples in the sequence and encourage them to accumulate evidence over time. For 12-sample sequences the hazard rate of $H$=0.08 yielded 39.9% of trials with no state change, 38.0% with one state change, and 22.1% with >1 state change.

Before the onset of each trial, there was a preparatory period of variable duration (uniform between 0.5 and 2.0 s) during which participants were instructed to maintain fixation and a stationary checkerboard patch was presented at a central location (0° polar angle). Trial onset was signaled when this checkerboard began to flicker, and 400 ms later the first evidence sample was presented (see *Stimuli*). Immediately after the sample sequence, the stationary patch was again presented at 0° and remained there until the start of the following trial. After a variable interval following sequence completion (uniform between 1.0 and 1.5 s) a ‘Go’ cue instructed participants to report their choice via button press with the left or right thumb (indicating state *left* or *right*, respectively). Auditory feedback was provided 0.1 s post-response informing the participant about the accuracy of their choice relative to the true generative state at the end of the sequence (see *Stimuli*). A rest period of 2 s immediately

followed feedback, during which participants were instructed to blink if necessary, and this was followed by the preparatory period for the next trial.

### *Stimuli*

All visual stimuli described below were presented against a grey background. Two placeholder stimuli were on screen throughout each block of trials: a light-grey vertical line extending downward from fixation to 7.4 degrees of visual angle (d.v.a.) eccentricity; and a colored half-ring in the lower visual hemifield (polar angle: from -90 to +90°; eccentricity: 8.8 d.v.a.), which depicted the log-likelihood ratio associated with each possible sample location (see below). The colors comprising this half-ring, along with those of the fixation point, were selected from the Teufel colors[75].

Each evidence sample consisted of a black and white, flickering checkerboard patch (temporal frequency: 10 Hz; spatial frequency: 2 d.v.a.) within a circular aperture (diameter = 0.8 d.v.a.). The patches varied in position from sample to sample, whereby their eccentricity was held constant (8.1 d.v.a.) and polar angle varied across the lower visual hemifield (i.e., from -90 to +90°.). The colored half-ring described above was presented at a larger eccentricity than the checkerboard patches to avoid overlap. Patches marking trial onset and in the evidence sequence itself were presented for 300 ms, followed by a 100 ms blank interval. This blank interval ensured that each patch was differentiable from a temporally adjacent patch even in rare cases when they overlapped in space.

The fixation mark, which was presented in the center of the screen, was a black disc of 0.18 d.v.a. diameter superimposed onto a second disk of 0.36 d.v.a. diameter and with varying color. The fixation mark was present throughout each task block, with the color of the second disk informing participants about task-relevant events. During the preparatory period, sample sequence presentation, and subsequent delay, the color of the second disk was light red; the 'Go' cue took the form of the second circle becoming light green; and, the inter-trial rest period was indicated by the second circle becoming light blue.

Auditory feedback consisted of a 0.25 s ascending 350→950Hz tone if the choice was correct and a descending 950→350Hz tone if incorrect.

## Task training protocol

Each participant completed a task training protocol in a separate session that took place in a behavioral psychophysics room one week or less prior to the first main experimental session. The protocol started with a visual illustration of each generative distribution, explained through analogy with a deck of cards. The participant then completed 12 trials of a static version of the task ($H$=0), after which they received feedback on their mean choice accuracy and were given the option to complete another 12 trials. The experimenter enforced repetition in cases where accuracy was less than 90%. Next the participant was informed that, during the real task, the deck that was being used to draw the dot locations could change at any time within a sequence of dots. This was explained through analogy with a biased coin flip between each dot presentation, where a 'heads' would be the outcome on "about 11 out of 12 flips" and would not change the deck, while a 'tails' would be the outcome on "about 1 out of 12 flips" and would change the deck.

The participant then performed 20 trials of the full task ($H$=0.08) in which, whenever a change in generative state occurred, the text string 'CHANGE!' (22-point white Helvetica font) appeared 1.45 d.v.a. above fixation for the duration of the following dot. Participants were

informed that they could use this change-point signaling to help them make decisions during the training exercise but that it would not appear during the real task. They again received feedback on their mean choice accuracy and were given the option to complete another 20 trials, with the experimenter enforcing repetition in cases where accuracy was less than 80%. Lastly, participants completed a mode of 5 (range = 2–6) blocks of the full task, without change-point signaling. Each block consisted of 76 trials, after which participants received feedback on their mean choice accuracy in that block and took a short, self-timed break.

### Localizer tasks

Within each session of the main experiment, participants also completed a single block of each of two 'localizer' tasks for (i) decoding the position (polar angle) of the checkerboard patterns from visual cortical responses and (ii) measuring motor preparatory activity for the hand movement used to report the choice, without participants performing the decision-making task described above. We obtained unreliable decoding of polar angle, so the data from this task were not used for the current study.

We used a delayed hand movement task to measure motor preparation in the absence of evidence accumulation and decision-making. The task was analogous to the delayed saccade task used to identify neurons encoding saccade plans in occulomotor structures of the monkey brain in previous studies of visual decision-making (e.g. refs. [76]). Participants fixated the central fixation mark while a sequence of lexical cues and subsequent 'Go' cues for responding were presented. A trial began with the presentation of one of two lexical cues ('LEFT' or 'RIGHT'; 15-point white Trebuchet font; 0.3s duration) that appeared 1.25° above the central fixation mark (of same construction as above; color of second disk light red). The participants' task was to prepare the associated motor response (left thumb button press for 'LEFT'; right thumb for 'RIGHT') during a following 1 s blank delay period within which only the fixation mark was presented, and to execute that response as quickly as possible when the delay was over (marked by the second disk turning light green). The second fixation disk became light blue for 2 s post-response, indicating a rest period during which participants were instructed to blink if necessary. The fixation mark color returned to light red after the rest period and, after a variable interval (uniform distribution with bounds of 0.75–1.5 s), the next trial began. A block of the motor preparation task was comprised of 60 trials (30 left and 30 right responses, randomly distributed), and was administered after the sixth block of the decision-making task in each session.

### Procedure and Apparatus

The experimental sessions for each participant took place on consecutive days. They comprised between 7 and 9 blocks depending on time constraints, and included 2-minute breaks between blocks and one longer break of approximately 10 minutes in the middle of the session. Like during training, each block was comprised of 76 trials and followed by feedback about choice accuracy, now including the mean for that block and a running mean for that session. In total participants completed a mode of 26 blocks (range=22–33), corresponding to a median of 1976 trials (range=1628–2508).

All stimuli were generated using Psychtoolbox 3 for Matlab[77,78]. Visual stimuli were back-projected on a transparent screen using a Sanyo PCL-XP51 projector with a resolution of 1920x1080 at 60 Hz (presented on a VIEWPixx monitor during training with the same

resolution and refresh rate). Subjects were seated 61 cm from the screen in the MEG room, or with their head in a chinrest 60 cm from the monitor (training) in an otherwise unlit room.

### Normative model and derivation of change-point probability and uncertainty

The normative model for solving the inference problem in the above decision-making task prescribes that incoming evidence about the generative task state is accumulated over time in a manner that balances stable belief formation with fast change detection[20,32]:

$$L_n = \psi_n + LLR_n \tag{1}$$

$$\psi_n = L_{n-1} + \log\left[\frac{1-H}{H} + \exp\left(-L_{n-1}\right)\right] - \log\left[\frac{1-H}{H} + \exp\left(L_{n-1}\right)\right] \tag{2}$$

Here, $L_n$ was the belief of the observer after having observed the $n^{th}$ sample of evidence, and expressed in log-posterior odds of the alternative task states; $LLR_n$ was the relative evidence for each alternative carried by the $n^{th}$ sample, expressed as a log-likelihood ratio ($LLR_n$ = $\log(p(x_n|right)/p(x_n|left))$); and $\psi_n$ was the prior expectation of the observer before encountering the $n^{th}$ sample. The key feature that sets this model apart from traditional evidence accumulation models[1] is the transformation of $L_{n-1}$ into $\psi_n$ (Equation 2) and how this takes into account the hazard rate (i.e., probability of state change), $H$. In the special case that $H$=0 (no state changes), the two rightmost terms in Equation 2 cancel and Equation 1 reduces to perfect evidence accumulation, as in drift diffusion[1,79]. When $H$=0.5 (no state stability), $\psi_n$ cancels and so no evidence is accumulated (posterior belief depends only on current evidence). For intermediate values of $H$, the model strikes a balance between these extremes, and thus between stability and sensitivity to change (i.e. flexibility; Supplementary Figure 1).

We used this previously established model to derive two computational quantities: *CPP* and -|$\psi$|. This exercise was motivated by similar treatments for a different form of change-point task (continuous belief updating[35]). The goal was to recast the non-linearity in eq. 2 in a way that would directly illuminate the underlying neural computations and hence, decision-related MEG and pupil signals.

#### *Change-point probability*

We derived a formal expression for the probability that a change in generative task state has just occurred given the expected $H$, the evidence for each state carried by a newly encountered sample $x_n$, and the observer's belief before encountering that sample $L_{n-1}$. We refer to this quantity as *CPP*. In what follows we denote all encountered samples up to and including the new sample as $x_{j\in N}$, refer to *right* and *left* states as $S_1$ and $S_2$ for convenience, and note that the log-posterior odds $L$ can be re-expressed as the probability of each generative state: $p(S_1)=1-p(S_2)=1/(e^L+1)$.

The task structure enabled two state transitions that define a change-point: $S_1 \rightarrow S_2$ and $S_2 \rightarrow S_1$; and two transitions that are not a change-point: $S_1 \rightarrow S_1$ and $S_2 \rightarrow S_2$. Thus, we could write the probability of a change-point transition as follows:

$$CPP_n = \frac{p\left(S_{1,n}, S_{2,n-1} \mid x_{j\in N}\right) + p\left(S_{2,n}, S_{1,n-1} \mid x_{j\in N}\right)}{p\left(S_{1,n}, S_{2,n-1} \mid x_{j\in N}\right) + p\left(S_{2,n}, S_{1,n-1} \mid x_{j\in N}\right) + p\left(S_{1,n}, S_{1,n-1} \mid x_{j\in N}\right) + p\left(S_{2,n}, S_{2,n-1} \mid x_{j\in N}\right)} \tag{3}$$

where the denominator accounted for all possible state transitions between consecutive samples $n$-1 and $n$. Using the definition of conditional probabilities and separating the set of all observed samples into the new sample $x_n$ and all previous samples $x_{1...n-1}$, we decomposed each of the four possible state transition probabilities in the following way:

$$p\left(S_{1,n}, S_{2,n-1} \mid x_n, x_{1...n-1}\right) = \frac{p\left(S_{1,n}, S_{2,n-1}, x_n, x_{1...n-1}\right)}{p\left(x_n, x_{1...n-1}\right)}$$

This expression could be expanded via the product rule:

$$\begin{aligned}
&= \frac{p\left(x_n \mid S_{1,n}, S_{2,n-1}, x_{1...n-1}\right) p\left(S_{1,n}, S_{2,n-1}, x_{1...n-1}\right)}{p\left(x_n, x_{1...n-1}\right)} \\
&= \frac{p\left(x_n \mid S_{1,n}, S_{2,n-1}, x_{1...n-1}\right) p\left(S_{1,n}, S_{2,n-1} \mid x_{1...n-1}\right) p\left(x_{1...n-1}\right)}{p\left(x_n \mid x_{1...n-1}\right) p\left(x_{1...n-1}\right)} \\
&= \frac{p\left(x_n \mid S_{1,n}, S_{2,n-1}, x_{1...n-1}\right) p\left(S_{1,n}, S_{2,n-1} \mid x_{1...n-1}\right)}{p\left(x_n \mid x_{1...n-1}\right)} \\
&= \frac{1}{Z} p\left(x_n \mid S_{1,n}, S_{2,n-1}, x_{1...n-1}\right) p\left(S_{1,n}, S_{2,n-1} \mid x_{1...n-1}\right)
\end{aligned}$$

where $Z = p(x_n|x_{1...n-1})$. Because previous observations and beliefs are irrelevant for determining the probability of a new sample or state transition when the current generative state is known, we could simplify the above expression and expand via the product rule:

$$\begin{aligned}
&= \frac{1}{Z} p\left(x_n \mid S_{1,n}\right) p\left(S_{1,n}, S_{2,n-1} \mid x_{1...n-1}\right) \\
&= \frac{1}{Z} p\left(x_n \mid S_{1,n}\right) p\left(S_{1,n} \mid S_{2,n-1}, x_{1...n-1}\right) p\left(S_{2,n-1} \mid x_{1...n-1}\right) \\
&= \frac{1}{Z} p\left(x_n \mid S_{1,n}\right) p\left(S_{1,n} \mid S_{2,n-1}\right) p\left(S_{2,n-1} \mid x_{1...n-1}\right)
\end{aligned}$$

As established previously, $p(S_{1,n}|S_{2,n-1})=p(S_{2,n}|S_{1,n-1})=H$, thus reducing the above to:

$$= \frac{H}{Z} p\left(x_n \mid S_{1,n}\right) p\left(S_{2,n-1} \mid x_{1...n-1}\right)$$

Finally, in our case the generative stimulus distribution $p(x|S_1)$ is Gaussian such that:

$$= \frac{H}{Z} \mathrm{N}\left(x_n \mid S_1\right) p\left(S_{2,n-1} \mid x_{1...n-1}\right) \tag{4}$$

where $\mathrm{N}(x|S)$ denoted the probability of sample $x$ given a normal distribution with mean $\mu_S$ and s.d. $\sigma_S$. As described above, the final term in eq. 4, which we abbreviated to $p(S_{2,n-1})$ below, could be computed directly from $L_{n-1}$.

The remaining transition probabilities in eq. 3 could be derived analogously to eq. 4. Replacing each term in Equation 3 with the expressions derived by doing so, the $Z$ terms cancelled to yield the following expression for $p(CP)$ that could be readily computed:

$$CPP_n = \frac{H\left(\mathrm{N}\left(x_n \mid S_1\right) p\left(S_{2,n-1}\right) + \mathrm{N}\left(x_n \mid S_2\right) p\left(S_{1,n-1}\right)\right)}{H\left(\mathrm{N}\left(x_n \mid S_1\right) p\left(S_{2,n-1}\right) + \mathrm{N}\left(x_n \mid S_2\right) p\left(S_{1,n-1}\right)\right) + \left(1-H\right)\left(\mathrm{N}\left(x_n \mid S_1\right) p\left(S_{1,n-1}\right) + \mathrm{N}\left(x_n \mid S_2\right) p\left(S_{2,n-1}\right)\right)} \tag{5}$$

This quantity has intuitive characteristics (Figure 1c). First, the numerator computes a weighted sum of the likelihood of the new sample under both generative states assuming that a change has occurred, with each weight determined by the strength of the observer's existing belief in the *opposing* state. This means that a new sample of evidence that is inconsistent with the observer's belief (i.e. sign($LLR_n$)≠sign($L_{n-1}$)) will yield a larger *CPP* than a sample that is consistent. Second, if the new sample carries no information about the current generative state (i.e. $LLR_n$=0), eq. 5 evaluates to *H*. In other words, when a new sample is ambiguous, the observer must rely more on their base expected rate of state change as an estimate for *CPP*. Similarly, if the observer is completely agnostic as to the task state (i.e. $L_{n-1}$=0), eq. 5 again evaluates to *H*. That is, a belief about a change-point having occurred over and above the base expected rate of change can only form if the observer has some level of belief in the task state before encountering the new sample. These characteristics mean that *CPP* can differentiate instances – caused by asymptotes in the non-linearity of the normative model – when belief states either plateau upon encountering consecutive consistent samples (low *CPP*), or change significantly upon encountering a strongly inconsistent sample (high *CPP*).

***Uncertainty***

We also defined belief uncertainty prior to observing a new evidence sample $x_n$ as $-|\psi_n|$. Thus, the closer the observer's prior belief to the category boundary of zero, the higher their uncertainty (Figure 1d). This measure thus identifies instances when the observer is at the steepest part of the non-linearity of the normative model, where belief updating is generally strong.

***Influence of change-point probability and uncertainty on evidence accumulation***

We used simulations to understand the influence of *CPP* and $-|\psi|$ on evidence accumulation in the normative model. We evaluated this influence as a function *H* and signal-to-noise ratio (SNR) of the generative distributions (defined as the difference between distribution means divided by their matched s.d.). For each point on a 5x5 grid (*H* ={0.01, 0.03, *0.08*, 0.20, 0.40}, SNR={0.4, 0.7, *1.2*, 2.0, 5.0}; italicized values match the generative statistics of our task), we simulated a sequence of 10,000,000 observations, passed these through the normative accumulation rule described by eqs. 1 and 2, and calculated the per-sample computational variables described above. We then assessed the influence of different variables on belief updating by fitting the following linear model (**Model 1**):

$$\psi_{n+1} = \beta_0 + \beta_1 \cdot \psi_n + \beta_2 \cdot LLR_n + \beta_3 \cdot LLR_n \cdot CPP_n + \beta_4 \cdot LLR_n \cdot -|\psi|_n \tag{6}$$

where $CPP_n$ was log-transformed to reduce skew, and both log($CPP_n$) and $-|\psi|_n$ were z-scored to reduce co-linearity between the interaction terms and $LLR_n$. In this regression model, *CPP* and $-|\psi|$ modulated the gain with which new evidence influenced the observer's existing belief. Note that $\beta_1$ and $\beta_2$ fully captured the simple summation part of the normative accumulation rule (eq. 1). $\beta_3$ and $\beta_4$ thus approximated the non-linear component of the accumulation dynamics introduced by eq. 2. We assessed the contribution of each of the four terms to sample-wise belief updating by calculating their coefficients of partial determination (Figure 1f; Supplementary Figure 1):

$$R^2_{partial} = \frac{SSR_{reduced} - SSR_{full}}{SSR_{reduced}} \quad (7)$$

where $SSR_{full}$ was the sum of squared residuals of the full model in eq. 6 and $SSR_{reduced}$ was the sum of squared residuals of an otherwise identical model that excluded the term of interest.

We repeated the above analysis with two alternative surprise metrics: 'unconditional' Shannon surprise calculated using only knowledge of the generative distributions (Spearman's $\rho$ with log(*CPP*)=0.00); and 'conditional' Shannon surprise calculated using both knowledge of the generative distributions and the observer's existing belief state (Spearman's $\rho$ with log(*CPP*)=0.36). Definitions of these metrics and their modulatory effects on normative belief updating are reported in Supplementary Figure 2.

**Modelling human choice behavior**

The similarity between our human participants' choices and those of various candidate decision processes was evaluated in three ways. First, we computed the accuracy of the humans' choices with respect to the true generative state at the end of each trial, and compared this to the accuracy yielded by three idealized decision processes presented with the same stimulus sequences as the humans (Figure 2a): the normative accumulation process for our task described above (with $H$=0.08), perfect accumulation of all *LLR*s (which is normative only for the special case of $H$=0), and basing one's decision on the sign of the final evidence sample (normative for the special case of $H$=0.5). For each strategy and trial, choice $r$ (*left* = -1, *right* = +1) was determined by the sign of the log-posterior odds after observing all samples: $r_{trl} = sign\left(L_{n,trl}\right)$ for the normative rule, $r_{trl} = sign\left(\sum_{i=1}^{n} LLR_{i,trl}\right)$ for perfect accumulation, and $r_{trl} = sign\left(LLR_{n,trl}\right)$ for last sample only, where $n$ indicated the number of samples presented on trial *trl*.

Second, for each strategy and human participant, we computed choice accuracy as a function of the duration of the final environmental state on each trial (i.e. the number of samples presented to the participant after the final change-point occurred; ranging from 1 on trials where a change-point occurred immediately before the final sample, to 12 on full-length trials where no change-points occurred; Supplementary Figure 3).

Third, we assessed the consistency between the humans' choices and the choices generated by each idealized strategy by computing the slope of a psychometric function relating the strategy-specific log-posterior odds to human choice (Supplementary Figure 3). For each strategy and participant, we normalized log-posterior odds across trials and described the probability $\hat{r}$ of a making a *right* choice on trial *trl* as:

$$\hat{r}_{trl} = \gamma + \frac{1-\gamma-\lambda}{1+e^{-(\delta+\alpha DV_{trl})}} \quad (8)$$

where $\gamma$ and $\lambda$ were lapse parameters, $\delta$ was a bias term, $DV_{trl}$ was the z-scored log-posterior odds on trial *trl*, and $\alpha$ was the slope parameter, which reflected the consistency between the choices produced by a given strategy and those of a human participant. We estimated $\gamma$, $\lambda$, $\delta$ and $\alpha$ by minimizing the negative log-likelihood of the data using the Nelder-Mead simplex search routine, under the constraint that the lapse rate was not dependent on choice (i.e. $\gamma$

and $\lambda$ were equal). Differences in $\alpha$ between candidate strategies were tested for via paired *t*-test.

***Normative model fits***

We also fitted variants of the normative model to the participants' behavior. We assumed that choices were based on the subjective log-posterior odds $L_{n,trl}$ for the observed stimulus sequence on each trial *trl*, given eqs. 1-2. This per-trial variable was also corrupted by a noise term $v$, such that choice probability $\hat{r}$ was computed as:

$$\hat{r}_{trl} = \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{L_{n,trl}}{\sqrt{2}v}\right) \tag{9}$$

In addition to the noise term, we allowed for the possible presence of three further deviations of the participants' away from the ideal observer: misestimation of the hazard rate, $H$; a bias in the mapping of stimulus location to *LLR*; and, a bias in the weighting of evidence samples that were (in)consistent with an existing belief state (see following section for motivation and details).

We fit the model to each participant's data by minimizing the cross-entropy between human and model choices:

$$e = -\sum_{trl}\left(1 - r_{trl}\right)\log\left(1 - \hat{r}_{trl}\right) + r_{trl}\log\left(\hat{r}_{trl}\right) \tag{10}$$

where $r_{trl}$ indicates the human participant's choice on trial *trl* as above, and $\hat{r}_{trl}$ is the model's choice probability calculated as per eq. 9. The sum of the cross-entropy $e$ with any regularization penalty terms (see below) was minimized via particle swarm optimization[80], setting wide bounds on all parameters and running 300 pseudorandomly-initialized particles for 1500 search iterations.

The relative goodness of fit of different model variants was assessed by calculating the Bayes Information Criterion (BIC):

$$BIC = 2e + k\log\left(n\right) \tag{11}$$

where $k$ was the number of free parameters (see below), $n$ was the number of trials and $e$ was the cross-entropy as per eq. 10 (equivalent here to the negative log-likelihood of the data given the fitted model).

***Motivation of modelling choices for normative model***

As described in the preceding section, the best-fitting version of the normative model allowed for four deviations from the purely ideal observer. We motivate the inclusion of each of these here:

*1. Choice selection noise:* As shown in eq. 9, we applied a noise term only to the final log-posterior on a given trial (i.e. to the translation of final belief into a choice, after accumulation of all presented samples had taken place). We favored this noise variant to maintain consistency with previous implementations of the normative model[20,73] and because it is tractable for model fits. Moreover, we did not find evidence for a pre-dominance of noise at the evidence accumulation stage, as reported recently for other tasks[81]: Participants' choice variability relative to the best-fitting model was invariant to trial sequence length

(Supplementary Figure 8), indicating that sample-wise 'accumulation noise' was not necessary to explain the behavioral data.

*2. Misestimation of H:* In line with previous work on change-point tasks[20,82], we allowed for participant-specific subjectivity in the hazard rate, *H*. Indeed, the model fits indicated that participants had a tendency to underestimate *H* (subjective *H*=0.039 ± 0.005 (s.e.m.); $t_{16}$=-7.7, $p<10^{-6}$, two-tailed one-sample *t*-test of subjective *H* against true *H*; Supplementary Figure 4). This systematic bias could reflect prior expectations toward relative environmental stability at the stimulus presentation rate used here.

*3. Non-linear stimulus-to-LLR mapping:* In line with observations from other tasks[83,84], we allowed for a non-linearity in the mapping of the decision-relevant stimulus dimension (polar angle) onto *LLR*. This was motivated by an analysis of our data in which we estimated the weight ascribed by participants to objective stimulus positions using an approach described elsewhere[4]. This analysis estimated the subjective weight of evidence associated with samples falling into evenly spaced bins (bin spacing = 0.6 in true *LLR* space) using logistic regression:

$$\mathrm{logit}\left[p\left(r_{trl}=right\right)\right]=\sum_{k\in bins}\beta_k N_{k,trl} \tag{12}$$

where $N_{k,trl}$ was the number of samples appearing on trial *trl* with a true *LLR* falling in bin *k*. Fitting this regression model to choices produced by both our human participants and the normative accumulation rule with matched noise revealed that human participants tended to give particularly strong weight to extreme samples (Supplementary Figure 4). To account for this effect in our model fits without making assumptions about the shape of the participants' weighting functions, we estimated the subject-specific mappings of stimulus polar angle to subjective *LLR* as a non-parametric function that was fit to the observers' choices alongside the other free parameters described here. We expressed the subjective *LLR* as an interpolated function of stimulus polar angle *x* whereby values were estimated at *x*={12.5, 25, 37.5, 50, 62.5, 75, 82.5, 90}, we assumed symmetry around *x*=0 and *LLR*(*x*=0)=0, and interpolation was performed using cubic splines. For fitting, we used Tikhonov regularization of the first derivative of the function (thus promoting smoothness) by adding a penalty term to the objective function (see above): $\gamma\sum_i\left(\Delta LLR_i\right)^2$, where *i* indexes the value of *x* for which a subjective *LLR* was estimated and $\gamma$ = 1/20 was determined through ad hoc methods[20]. These fits revealed that the participants tended to over-weight evidence samples in the extrema of the stimulus space, an effect that has been shown previously to countermand the decrease in accuracy incurred by noise[83].

*4. Bias in weighting of (in)consistent evidence:* Lastly, we allowed for the possibility that humans might give greater weight, relative to the normative accumulation process, to samples of evidence that were (in)consistent with their existing belief state[85,86]. To do so, we applied a multiplicative gain factor *g* selectively to *LLR*s associated with inconsistent samples, such that the effective evidence strength $LLR_n = LLR_n \cdot g$ for any sample *n* where $sign\left(LLR_n\right)\neq sign\left(\psi_n\right)$. Thus, *g* > 1 corresponds to a relative upweighting of inconsistent samples, while 1 > *g* ≥ 0 corresponds to a relative upweighting of consistent samples. We found that participants assigned higher weight than the normative model to samples that were inconsistent with their existing beliefs (fitted weight=1.40 ± 0.04 (s.e.m.); $t_{16}$=8.2, $p<10^{-6}$, two-tailed one-sample *t*-test of fitted weights against normative weight of 1; Supplementary Figure

4), perhaps reflecting constraints on the neural circuit implementation of non-linear evidence accumulation.

In total, the full fits of the normative model plus bias terms consisted of 11 free parameters per observer (1 noise term, 1 subjective *H*, 8 stimulus-to-*LLR* mapping parameters, and 1 gain factor on inconsistent samples). We also fit more constrained variants of the normative model that lacked various combinations of the deviations from the ideal observer described above (Supplementary Figure 4). To facilitate fair comparison across model variants, for this set of analyses we re-parameterized the stimulus-to-*LLR* mapping function for each sample *n* as a scaled exponential:

$$LLR_n = |LLR_n|^{\kappa} \cdot \beta \cdot sign\left(LLR_n\right) \tag{13}$$

which was more constrained than the interpolated function in previous fits but can produce convex ($\kappa > 1$), concave ($\kappa < 1$) or linear ($\kappa = 1$) mapping functions using only two free parameters (exponent $\kappa$ and scale parameter $\beta$).

***Fitting the L to ψ mapping***

To assess how well the critical non-linearity in the normative model (eq. 2) captured the belief updating dynamics of the human participants, we also fit a model in which we estimated this function directly from the observers' choice data without constraining its shape[20]. As with the subjective stimulus-to-*LLR* mapping function above, we here estimated the mapping of $L_{n-1}$ to $\psi_n$ as an interpolated non-parametric function. We assumed symmetry of the function around $L_{n-1}=0$ and $\psi(L_{n-1}=0)=0$, and $\psi_n$ was estimated for values of $L_{n-1}$ that were spread evenly between 1 and 10 in steps of one. We applied Tikhonov regularization to the first derivative as described above, here with $\gamma = 1/2$. This model had a total of 20 free parameters, with the new mapping function replacing the subjective *H* parameter from the previous fits.

***Alternative accumulator model fits***

We additionally fit two alternative, sub-optimal accumulator models that each lack a key characteristic of the normative accumulation process. The first of these is a linear approximation of the normative model that employs leaky accumulation but lacks the stabilizing, non-absorbing bounds in the normative *L* to *ψ* mapping. This linear model has been shown to be capable of approximating some operating regimes of the normative model[20,38], and substitutes eq. 2 for the following *L* to *ψ* mapping:

$$\psi_n = L_{n-1}\left(1-\lambda\right) \tag{14}$$

where $\lambda$ is a free parameter reflecting the amount of leak.

The second sub-optimal model employs perfect, lossless evidence accumulation toward non-absorbing bounds. This model can capture the critical non-linearity in the normative *L* to *ψ* mapping, but lacks the leaky accumulation that is an important prescription of the normative model when beliefs are uncertain and *H* is high[20]. This model substitutes eq. 2 for the following:

$$\psi_n = \begin{cases} -A & L_{n-1} \le -A \\ L_{n-1} & -A < L_{n-1} < A \\ A & L_{n-1} \ge A \end{cases} \tag{15}$$

where *A* is a free parameter reflecting the height of the non-absorbing bounds.

Our fits of both alternative models variously incorporated noise, non-linear *LLR* mapping and an inconsistency bias in the same ways as described above (Supplementary Figure 4). Thus, with $\lambda$ or $A$ replacing subjective $H$, these models had the same number of free parameters as our fits of the normative model plus bias terms.

### *Psychophysical kernels*

We estimated "psychophysical kernels", which quantify the impact (regression weight) of evidence on choice as a function of time[37]. To do so, we fit the following logistic regression model to choices on full-length, 12-sample trials (**Model 2**):

$$\text{logit}\left[p\left(r_{trl}=right\right)\right]=\beta_0+\sum_{i=1}^{12}\beta_{1,i}\cdot LLR_{i,trl}+\sum_{j=2}^{12}\left(\beta_{2,j}\cdot LLR_{j,trl}\cdot CPP_{j,trl}+\beta_{3,j}\cdot LLR_{j,trl}\cdot -|\psi|_{j,trl}\right) \tag{16}$$

where $i$ and $j$ indexed sample position within the stimulus sequence on trial *trl*, and *LLR* was the true *LLR*s given the generative task statistics (i.e. not subjected to the non-linearity or inconsistency bias in the model fits). For the humans the dependent variable was the empirically observed choice (*left*=0, *right*=1); for the models it was the choice probability calculated as per eq. 9. The set of coefficients $\beta_1$ estimated the time-dependent leverage of evidence on choice. The additional set of interaction terms $\beta_2$ and $\beta_3$ estimated the modulation of evidence weighting by change-point probability and uncertainty, respectively. As in **Model 1** (eq. 6), *CPP* was log-transformed, and both log(*CPP*) and -|$\psi$| were z-scored before multiplication with *LLR* to reduce collinearity. Additionally, all final regressors were z-scored across the trial dimension to yield fitted coefficients on the same scale. Cluster-based permutation testing[87] was used to identify sample positions for each of the three sets of terms at which fitted weights differed significantly from zero (two-tailed one-sample *t*-test; 10,000 permutations; cluster-forming threshold of $p<0.05$). Differences between the *CPP* and -|$\psi$| weights were tested via paired *t*-test after averaging weights over sample position.

### *Circuit modelling*

We simulated the choice behavior of an established cortical circuit model for decision-making[26] on our task. The model consisted of 1600 pyramidal cells and 400 inhibitory interneurons, all of which were spiking neurons with multiple different conductances (see below). The pyramidal cells were organized into three distinct populations: 240 neurons selective for the 'left' choice (population $D_1$), 240 neurons selective for the 'right' choice ($D_2$), and the remaining neurons non-selective. Neurons within the choice-selective populations sent recurrent connections to neurons in the same population as well as to a common pool of inhibitory interneurons ($I$), which fed back onto both $D_1$ and $D_2$. Pyramidal neurons projected to AMPA and NMDA receptors (with fast and slow time constants, respectively) on target cells, and interneurons projected to $GABA_A$ receptors. The parameterization and update equations of the circuit model were taken from their original description in ref. [26], except for three changes described in the following.

First, in the original implementation of the model, stimulus-driven inputs to the choice-selective populations varied linearly with stimulus strength, were symmetric around 40 Hz, and always summed to 80 Hz. Thus, both choice-selective populations received average input of 40 Hz when stimulus strength was zero, while one received 80 Hz and the other 0 Hz at maximal stimulus strength. This stimulus input function produced excessive primacy in evidence weighting, inconsistent with the data. Stimulus inputs (sample-wise *LLR*) needed to be stronger for changes-of-mind to occur in response to inconsistent evidence (see also ref. [26]).

Thus, we used a threshold-linear input function that was also symmetric around 40 Hz for the two choice-selective populations, but imposed no upper bound on the input:

$$\begin{aligned} Input_{D_1} &= \max\left(0, 40 - LLR \cdot c\right) \\ Input_{D_2} &= \max\left(0, 40 + LLR \cdot c\right) \end{aligned} \tag{17}$$

where $Input_x$ was the input to choice-selective population $x$, and $c$ was a multiplicative scaling factor applied to the sample-wise *LLR*. We set $c$ to 19, yielding an input of ~110 Hz to the favored choice-selective population for the strongest possible stimulus (and 0 Hz to the other population). We verified that strong stimulus input, rather than the threshold-linear function itself, was the key factor for approximating normative decision-making in our task: An input function that was again symmetric around 40 Hz, but had different slopes for $LLR<0$ and $LLR>0$ and produced an input of 0 Hz to the non-favored population only for the most extreme evidence strength, yielded very similar behavior as what we report here (data not shown; see ref. [26] for further examination of this form of input function).

Second, the recurrent connectivity between pyramidal cells in the model was structured to enforce stronger coupling between cells within the *same* choice-selective population than between cells in *different* choice-selective populations. Specifically, within a choice-selective population, $w_j=w_+$, where $w_+>1$ determined the strength of "potentiated" synapses relative to the baseline level of potentiation. Between two different choice-selective populations, and from the non-selective population to the selective ones, $w_j=w_-$, where $w_-<1$ determined the relative strength of synaptic depression. $w_-$ was directly determined by $w_+$ such that the overall recurrent excitatory synaptic drive without external stimulus input remained constant as a function of $w_+$ is varied[26]. In the original implementation of the model $w_+=1.7$, which produced relatively strong and stable attractor states that, as with weakly scaled stimulus input, prohibited changes-of-mind in response to inconsistent evidence for all but the strongest evidence strengths. We thus set $w_+=1.68$, with the resulting mildly weakened attractor dynamics allowing the model to better approximate normative decision-making on our task (stronger recency, and increased sensitivity to *CPP*).

Third, in the original model implementation simulations were run with an integration time step $dt$=0.02 ms. Because we needed to simulate ~25,000 trials of 5.6 s each and the model is slow to simulate, we set $dt$=0.2ms. We verified through simulations at the original $dt$ that this did not significantly change the behavior or population rate trajectories of the model.

We ran simulations of the model for all full-length (12-sample) trials presented to our human participants. The network was initialized at 0.4 s before the onset of the first evidence sample and the simulation ended 0.4 s after offset of the final evidence sample. Each evidence sample was assumed to provide external input to the choice-selective populations according to the stimulus input function described above, from the time of its onset to 0.4 s thereafter; during the 0.4 s periods before and after the sample sequence, external input was set to zero. The instantaneous mean population firing rates of the choice-selective populations, $r_{D_1}$ and $r_{D_2}$, were calculated by summing all spikes across a population within a 50 ms window centered on the time-point of interest, and dividing by the number of neurons in the population and the time window. The evolving model decision variable $X$ was defined as the difference between the instantaneous firing rates of the two populations: $X = r_{D_2} - r_{D_1}$. The model's choice was determined by *sign*($X$) at the end of each simulated trial; thus, like the human participants, the model needed to maintain a memory of its decision state during a post-evidence sequence delay period without external input. The model's updated decision variable given a new evidence sample was taken to be $X$ at 0.4 s after onset of that sample. We estimated the

model's psychophysical kernels as per eq. 16, using *CPP* and -|*ψ*| metrics from the ideal observer variant of the normative model.

We also simulated the choice behavior of a reduction of the above biophysical circuit model that was described by the diffusion of a decision variable *X* in the double-well potential *φ*:

$$\frac{dX}{dt} = -\frac{d\varphi}{dX} + \sigma\xi_t$$
$$\varphi(X) = -k\mu_t X - \frac{aX^2}{2} + \frac{bX^4}{4} \tag{18}$$

where $\mu_t$ was the differential stimulus input to the choice-selective populations at time point *t* relative to trial onset (in our case the per-sample *LLR,* which changed every 0.4 s) that was linearly scaled by parameter *k*; $\xi_t$ was a zero-mean, unit-variance Gaussian noise term that was linearly scaled by parameter *σ*; and *a* and *b* shape the potential[88]. The model was initialized on each trial at the onset of the first sample with *X* set to 0, simulated with time-step *dt* set to 25 ms, and its choice was determined by *sign*(*X*) at 0.4 s after the onset of the final sample in each sequence. As with the detailed spiking neuron model, we fed the reduced model the stimulus sequences for all full-length trials presented to our human participants, and calculated psychophysical kernels as per eq. 16.

The reduced model helped us to explore boundary conditions for approximating human behavior. We manually explored the parameter space to find a set of parameters at which the reduced model produced psychophysical kernels that approximately matched those of our human participants (*k*=2.2, *σ*=0.8, *a*=2, *b*=1). The shape of the potential corresponding to this combination of *a* and *b* indicated weak bi-stable attractors (Supplementary Figure 7, top). Keeping *k* and *σ* constant, we then simulated the behavior of the model for three qualitatively different dynamical regimes (Prat Ortega et al., 2020): a single attractor centered on *X*=0 (*a*=0, *b*=1), which produced extreme recency in evidence weighting; perfect integration (*a*=0, *b*=0), which produced a flat psychophysical kernel and no sensitivity to *CPP* or -|*ψ*|; and strong winner-take-all dynamics (*a*=5, *b*=1), which produced extreme primacy in evidence weighting (Supplementary Figure 7).

**Data acquisition**

MEG data were acquired on a whole-head CTF system (275 axial gradiometer sensors, CTF Systems, Inc.) in a magnetically shielded room at a sampling rate of 1,200Hz. The location of the participant's head was recorded and visualized in real-time using three fiducial coils, one fixed at the outer part of each ear canal and one on the nasal bridge. Participants were instructed to minimize movement during task performance. A template head position was registered at the beginning of each participants' first session, and the experimenter guided the participant back into that position before initializing each task block.

We used Ag/AgCl electrodes to measure the electrocardiolgram (ECG), vertical electrooculogram (vEOG), and electroencephalogram (EEG) from three scalp locations (Fz, Cz and Pz according to the 10/20 system) with a nasion reference, though these data are not analyzed here. Eye movements and pupil diameter were recorded during task performance at 1,000Hz using an MEG-compatible Eyelink 1000 Long Range Mount system (SR Research).

T1-weighted structural magnetic resonance images (MRIs) were acquired from all subjects to generate individual head models for source reconstruction (see below).

## MEG data analysis

### *Preprocessing*

MEG data were analyzed in MATLAB (MathWorks) and Python using a combination of the Fieldtrip toolbox[89], MNE[90,91] and custom-made scripts.

Continuous data were first segmented into task blocks, high-pass filtered (zero-phase, forward-pass FIR) at 0.5Hz, and bandstop filtered (two-pass Butterworth) around 50, 100 and 150 Hz to remove line noise. Data were then resampled to 400 Hz and re-segmented into single trials with the following task-dependent timings: from 1 s before trial onset to the onset of the 'Go' cue for the decision-making task; and from 1s before to 2 s after lexical cue onset for the motor localizer task. Trials containing any of the following artifacts were discarded from further analysis: (i) head motion of any of the three fiducial coils exceeding a translation of 6 mm from the first trial of the recording; (ii) blinks (detected using the standard Eyelink algorithm); (iii) saccades (detected with velocity threshold=30°$s^{-1}$, acceleration threshold=2000°$s^{-2}$) exceeding 1.5° in magnitude; (iv) squid jumps (detected by applying Grubb's test for outliers to the intercepts of lines fitted to single-trial/-sensor log-power spectra from data without a high-pass filter); (v) sensor(s) with a min-max data range exceeding 7.5 pT, usually caused by cars driving past the MEG laboratory; (vi) muscle artifacts (detected by applying a 110-140 Hz Butterworth filter to all MEG sensors, z-scoring across time, and applying a threshold of $z$=20 to each sensor).

### *Spectral analysis*

In order to isolate induced (non-phase-locked) activity components from the activity of each sensor, we first subtracted each sensor's trial-averaged ("phase-locked") response from its single-trial activity time courses. Thus, all results reported in this paper reflect modulations of cortical activity that are not phase-locked to external events, but rather are likely generated by recurrent synaptic interactions in cortical circuits[62] (see also *Discussion*: *Frequency-specific encoding of evidence and decision variables*). Repeating the analyses on the 'raw' signals (i.e. including both phase-locked and non-phase-locked activity components) did not reveal any additional features (data not shown). This likely reflects the fact that the computation of all variables assessed here entailed recurrent interactions in cortical micro- and macro-circuits[39,62]. Note that while phase-locked responses to single samples in early visual cortex were present in our data, this does not imply that these responses encode the to-be-accumulated evidence for the decision process; rather, deriving decision-relevant evidence (*LLR*) from sensory responses encoding polar angle required a transformation that incorporates knowledge of the stimulus statistics in our task.

We used sliding-window Fourier transform to compute time-frequency representations (TFRs) of the single-trial activity of each MEG sensor from both the decision-making and motor preparation tasks. Specifically, for low frequencies (1-35 Hz in steps of 1 Hz), we used a sliding window Fourier transform with one Hanning taper (window length of 0.4s in steps of 0.05s; frequency smoothing of 2.5 Hz). For high frequencies (36-100 Hz in steps of 4 Hz), we used the multi-taper method with a sequence of discrete proloid slepian tapers, a window length of 0.25 s in steps of 0.05 s, and 6 Hz frequency smoothing. We converted the complex-valued time-frequency representations of activity into units of power by taking the absolute values and squaring.

For sensor-level analyses, the axial gradiometer data were decomposed into horizontal and vertical planar gradients prior to time-frequency decomposition and these were combined afterwards to yield readily interpretable topographies.

We converted power estimates into units of modulation relative to pre-trial baseline as follows. Power estimates for each time-point *t*, frequency *f* and sensor *c* were normalized and baseline-corrected via the decibel (dB) transform $dB_{t,f,c}=10*\log_{10}(power_{t,f,c}/baseline_{f,c})$. Here, $baseline_{f,c}$ refers to the trial-averaged power collapsed across the interval from -0.4 to -0.2 s relative to trial onset. This was done to normalize (via the log-transform entailed in the above equation) the power estimates for linear regression analyses reported below.

***Construction of linear filters for motor preparatory activity***

We used data from the delayed motor response task ('motor localizer') to construct a set of filters for isolating hand movement-specific motor preparatory activity in the data from the decision-making task. Those filters are referred to as 'motor filters' in the following.

First, any motor localizer trial on which a manual response was executed before the 'Go' cue was excluded from analysis. This was the case on the majority of trials for 5 participants who apparently misunderstood the task instructions, and their data were not used for construction of the filters. Motor localizer MEG datasets from the remaining 12 participants were segregated into sensors covering the left side of the head (labelled 'ML*' by the CTF system) and those covering the right side of the head ('MR*'). For each matching pair of left/right sensors (e.g. left frontal sensor 'MLF45' matches with right frontal sensor 'MRF45'; 131 pairs in total), we calculated a lateralization index *LI* for each time-point, frequency and trial by subtracting the power modulation estimate $dB_{t,f,c}$ for the left sensor from the power estimate for the right sensor. We then fit the following linear regression model to each participant's data:

$$LI_{t,f,c',trl} = \beta_0 + \beta_1 \cdot r_{trl} + \sum_{i=1}^{n-1} \beta_{2,i} \cdot session_i \qquad (19)$$

where *t* indicated time-point relative to trial onset, *f* indicated frequency, *c'* indicated sensor pair, *trl* indicated trial, $r_{trl}$ was the single-trial response executed by the participant (*left*=0, *right*=1), and $session_i$ was a group of binary nuisance regressors included to absorb any main effect of experimental session (with *n* denoting the total number of sessions for a given participant).

The quantity of interest was the t-score associated with $\beta_1$, which provided a reliability-weighted measure of the strength with which *LI* encodes the motor response. We averaged these t-scores across the interval from 0.7 to 1.1 s post-lexical cue to generate a single sensor*frequency t-map per participant. This interval captured the period of the task during which planning, but not execution, of the motor response takes place. We then used cluster-based permutation testing (10,000 permutations with cluster-forming threshold of $p<0.01$, two-tailed) to identify spatio-spectral clusters that were significantly different from zero at the group level. This procedure yielded a single cluster ($p<0.001$; $p>0.57$ for all other clusters) with an associated sensor*frequency matrix *M* of group-level t-scores where all spatio-spectral points lying outside the cluster bounds were set to zero. The matrix *M* was used to construct the motor filters.

We generated three sets of motor filters: spectral filters, spatial filters, and spatio-spectral filters. To generate weights $w_f$ for a *spectral* filter, we integrated *M* over the spatial dimension such that:

$$w_f = \frac{\sum_{c'} M_{c',f}}{\sum_{c'}\sum_{f} M_{c',f}} \tag{20}$$

where the denominator normalized the weights to integrate to one. Weights for *spatial* and *spatio-spectral* filters were generated analogously by integrating over the spatial dimension in the numerator or not at all. The group-average spectral and spatial filters are shown in Figure 3a (right).

The resulting filters could then be applied to independent *LI* data by computing the dot product between data and filter along the desired dimension(s), yielding a filtered motor preparatory signal that we refer to as *motor* below. For example, the spectral filter computed through eq. 20 could be applied to yield a spatially-resolved motor preparatory signal as follows:

$$motor_{c'} = LI_{f,c'}\ w_f \tag{21}$$

***Model-based analysis of motor preparatory activity during decision-making task***

We applied the movement-selective motor preparation filters to sensor-level *LI* data from full-length (12-sample) trials of the decision-making task in the manner described above. Application of the *spatial* filter to these data in a time-resolved fashion generated TFRs of the relative motor preparation for each choice alternative during decision formation (Figure 3b). These were then segmented from 0 to 1s relative to the onset of each of the 12 samples per trial to yield a four-dimensional matrix (time*frequency*sample*trial), through which we assessed the sensitivity of the motor preparatory signal to decision-relevant computational variables.

It became apparent that fits of the normative model that allowed for non-linear stimulus-to-*LLR* mappings invariably yielded computational variables (in particular, *LLR*, *CPP* and *L*) with a small proportion of highly deviant values. This was caused by the tendency of human observers to assign especially strong weight to infrequently-encountered stimuli at the extrema of the stimulus range (Supplementary Figure 4). These outliers were in turn problematic for our analyses relating model-derived variables to neural measurements, the majority of which relied on linear regression. For all such analyses, we therefore derived computational variables from model fits in which the stimulus-to-*LLR* mapping was constrained to be linear (such that $LLR_n = \beta \cdot LLR_n$, where $\beta$ sets the slope of the mapping function and is a free parameter). Although this model variant yielded marginally worse goodness-of-fit to observers' choices compared to the full model (Supplementary Figure 4), the sample-wise computational variables generated by each were highly correlated ($\psi$: $\rho$ = 0.988 ± 0.014; *L*: $\rho$ = 0.988 ± 0.015; *CPP*: $\rho$ = 0.986 ± 0.018; -|$\psi$|: $\rho$ = 0.959 ± 0.045).

To determine the sensitivity of movement-selective motor preparation to each of the key components of normative belief updating established previously (Figure 1f; **Model 1**), we fit the following linear model (**Model 3**):

$$\begin{aligned} Motor_{t,f,s,trl} = \beta_0 + \beta_1 \cdot \psi_{s,trl} + \beta_2 \cdot LLR_{s,trl} + \beta_3 \cdot LLR_{s,trl} \cdot CPP_{s,trl} \\ + \beta_4 \cdot LLR_{s,trl} \cdot -|\psi|_{s,trl} + \sum_{i=1}^{n-1} \beta_{5,i} \cdot session_{i,trl} \end{aligned} \tag{22}$$

where $motor_{t,f,s,trl}$ was the motor preparatory activity (see above) at time *t* relative to sample onset, frequency *f*, sample *s* and trial *trl*. We computed t-scores for $\beta_{1\text{-}4}$ and averaged each

across the sample dimension, thereby obtaining single time-frequency maps per participant reflecting the strength with which each computational quantity is encoded in the varying motor preparation signal (Figure 3d). Clusters of significant encoding across time and frequency in the sample-averaged maps were identified via cluster-based permutation test (two-tailed one-sample *t*-test; 10,000 permutations, cluster-forming threshold of $p<0.05$), as were significant differences between the *CPP* and -|$\psi$| interaction terms (two-tailed paired *t*-test; same threshold and permutation number).

To complement the above analysis, we constructed two regression models, each of which contain complementary parts of the 'full' model from eq. 22. The idea was to fit neural signals that encoded either new evidence ($LLR_s$), or the non-linearly updated decision variable ($\psi_{s+1}$), associated with sample *s*. We then fit these models to $motor_{t,f,s,trl}$ on full-length trials (**Models 4-5**):

$$Motor_{t,f,s,trl} = \beta_0 + \beta_1 \cdot \psi_{s+1,trl} + \sum_{i=1}^{n-1} \beta_{2,i} \cdot session_{i,trl} \tag{23}$$

$$Motor_{t,f,s,trl} = \beta_0 + \beta_1 \cdot LLR_{s,trl} + \sum_{i=1}^{n-1} \beta_{2,i} \cdot session_{i,trl} \tag{24}$$

For each model, time-point *t*, frequency *f* and sample *s*, we computed a 'super-BIC' score reflecting that model's goodness-of-fit at the group level. This metric was calculated as per eq. 11, here with the negative-log-likelihood, number of free parameters and number of observations determined by the sum of these values across participants. We then averaged across the sample dimension and subtracted BIC scores for the 'belief model' (eq. 23) from those for the 'evidence model' (eq. 24). This generated a single time-frequency map of relative goodness-of-fit in which positive values indicate stronger encoding of belief relative to sensory evidence in the motor preparation signal, and *vice versa* for negative values. We further compared the belief model of eq. 23 to an alternative in which the non-linearly transformed decision variable after accumulating sample *s* ($\psi_{s+1}$) was replaced by the untransformed posterior log-odds ($L_s$). The same model comparison approach described above revealed that $\psi$ was the superior predictor of the motor preparation signal (data not shown), as expected given the observed modulation of the motor preparation signal by change-point probability and uncertainty (Figure 3d).

We also characterized the shape of the relationship between $motor_{t,f,s,trl}$ and model-derived belief state as follows. We applied the *spatio-spectral* motor preparation filter to the decision-making task data to yield a single scalar value per time point reflecting the relative motor preparation for each alternative (Figure 3c). This metric was then averaged from 0.4-0.6s after each sample *s* on full-length trials (thus capturing latencies at which the signal is modulated by *CPP* while minimizing contamination by responses to subsequent samples; Figure 3d), and *z*-scored across trials separately for each sample position and session. We also normalized the posterior log-odds $L_s$ in an analogous fashion to remove any differences in the range of *L* across participant-specific model fits. Next, for each sample position and participant we sorted the normalized motor preparation by normalized $L_s$ into 11 equal-sized bins, and calculated the mean of both metrics per bin *b*. We then averaged across sample positions and participants to yield a single *belief encoding function* for the motor preparation signal which, for visualization, we rescaled with sign preservation to have upper and lower bounds of +1 and -1, respectively (Figure 3f). We fit a unit square sigmoid to quantify the shape of this function:

$$motor_b = \frac{L_b^{e^{\xi}}}{L_b^{e^{\xi}} + \left(1 - L_b\right)^{e^{\xi}}} \tag{25}$$

where the single free parameter, the slope $\xi$, produced a function that could be sigmoidal ($\xi$>0), linear ($\xi$=0), or inverse-sigmoidal ($\xi$<0) in shape. The unit square sigmoid operated on inputs and outputs that were bounded between 0 and 1. Thus for the purposes of fitting the function only, the bin-wise motor preparation signal ($motor_b$) and belief ($L_b$) metrics $x$ were rescaled accordingly: $\overline{x}_b = \left(x_b - \min\left(x\right)\right) / \max\left(x\right)$. We estimated $\xi$ by minimizing the sum of squared residuals between the data and the fit via simplex search, and generated 95% confidence intervals for both the belief encoding function and $\xi$ fits via bootstrapping: for each participant randomly sampling, with replacement, an equal number of trials as in the data, and repeating the above procedure with the resampled trials (1,000 iterations). Bootstrapping was performed at the trial- rather than participant-level because it ensured that the entire sample could be included in the analysis; participant-level bootstrapping, on the other hand, would have required estimation of encoding functions and corresponding fits for each individual participant, and would not be meaningful for participants where effects were weak/absent.

In order to also estimate belief encoding functions for computational variables of interest from the normative model, we repeated the above procedure after replacing the motor preparation signal first with $\psi_{s+1}$, and then with $LLR_s$ (Figure 3f,g). In the case of $\psi_{s+1}$, the resulting function will generate the $L$-to-$\psi$ mapping for the normative model described by eq. 2, after application of the normalization, binning and averaging process described above. In the case of $LLR_s$, the function captured the shape of the relationship between the posterior log-odds and the most recently encountered sample of sensory evidence, which we expected to be characteristic of neural signals that faithfully encoded the momentary evidence. We further repeated the procedure using the difference in firing rates between choice-selective populations in the biophysical model, the variable reflecting this model's decision variable (see above). For this analysis we extracted difference measures at 0.4s after each sample onset.

### *Source reconstruction*

We used linearly constrained minimum variance (LCMV) beamforming to estimate activity time courses at the level of cortical sources[92]. We first constructed individual three-layer head models from subject-specific MRI scans using Fieldtrip[89] (functions: ft_volumesegment and ft_prepare_mesh). Second, head models were aligned to the MEG data by a transformation matrix that aligned the average fiducial coil position in the MEG data and the corresponding locations in each head model. Transformation matrices were generated using the MNE software package[91], and we computed one transformation matrix per recording session. Third, we reconstructed cortical surfaces from individual MRIs using FreeSurfer[93,94] and aligned two different atlases to each surface (see *Functional ROIs* below). In a fourth step we used the MNE package to compute LCMV filters for projecting data into source space. LCMV filters combined a forward model based on the head model and a source space constrained to the cortical sheet (4096 vertices per hemisphere, recursively subdivided octahedron) with a data covariance matrix estimated from the cleaned and segmented data. We computed one filter per vertex, based on the covariance matrix computed on the time-points from trial onset until 6.2 s after stimulus onset across all trials. We chose the source orientation with maximum output source power at each cortical location. In a final step, we computed TFRs of the segmented MEG data (same method as described in *Spectral analysis* above) and projected the complex time series into source space. In source space we computed TFR power at each vertex location by first aligning the polarity of time-series at neighboring vertices (because the beamformer output potentially included arbitrary sign flips for different vertices) and then

converting the complex Fourier coefficients in each vertex into power (taking absolute value and squaring).

We finally averaged the estimated power values across all vertices within a given ROI (see *Functional ROIs* below). As with the sensor-level analysis, source-power estimates were baseline-corrected using the dB transform (from -0.4 to -0.2 s relative to trial onset). We then computed source-level lateralization indices $LI_{t,f,trl,roi}$ for each time-point, frequency, trial and ROI by subtracting the power estimate for the left hemisphere ROI from the power estimate for the right hemisphere ROI.

The code that produced source estimates is available at www.github.com/Donnerlab/pymeg.

### *Regions of interests*

We used a total of 31 regions of interest (ROIs), which were delineated in previous functional MRI work[28,42,43] and likely participated in the visuo-motor transformation that mapped patch locations to behavioral choice in our task. These regions were composed of (i) retinotopically organized visual cortical field maps provided by the atlas from Wang et al.[42] and (ii) three regions exhibiting hand movement-specific lateralization of cortical activity: aIPS, IPS/PCeS and the hand sub-region of M1[28]; and (iii) a dorsal/ventral premotor cortex cluster of regions from a whole-cortex atlas[43].

***Table 1:*** *ROI definition along the sensorimotor pathway.*

| Cluster | Functional areas | Source |
|---|---|---|
| **V1** | Dorsal and ventral parts of V1 | Wang et al., 2015 |
| **V2-V4** | Dorsal and ventral parts of V2, V3, V4 | |
| **V3A/B** | V3A, V3B | |
| **IPS0/1** | IPS0, IPS1 | |
| **IPS2/3** | IPS2, IPS3 | |
| **LOC1/2** | LO1, LO2 | |
| **MT+** | TO1, TO2 | |
| **Ventral occipital** | V01, V02 | |
| **PHC** | PHC1, PHC2 | |
| **aIPS** | aIPS1 | de Gee et al., 2017 |
| **IPS/PCeS** | IPS/Post-central sulcus | |
| **M1 (hand)** | M1 | |
| **PMd/v** | 55b, 6d, 6a, FEF, 6v, 6r, PEF | Glasser et al., 2016 |

Following a scheme proposed by Wandell and colleagues[95], we grouped visual cortical field maps with a shared foveal representation into clusters (see Table 1), thus increasing the spatial distance between ROI centers and minimizing the risk of signal leakage[96] (due to limited filter resolution or volume conduction). ROI masks from both atlases[42,43] as well as ref. [28] were co-registered to individual MRIs.

### *Model-based analysis of different ROIs during decision-making task*

At the level of these functional ROIs, we repeated the model-based regression analyses described above for the sensor-level motor preparatory activity (see *Analysis of motor preparatory activity during decision-making task*). Here, we replaced the sensor-level motor

preparation signal $motor_{t,f,s,trl}$ with source-level hemispheric lateralization indices ($LI_{t,f,s,trl}$) for each ROI. We fitted the linear regression model comprising each of the key components of normative belief updating (**Model 3**; eq. 22) to each ROI. We plotted time-frequency maps of the t-scores reflecting sample-wise encoding of *LLR* ($\beta_2$) to initially highlight qualitative differences between individual ROIs (Figure 4a), and averaged within early visual (V1-V4), dorsal/ventral visual (V3A/B, IPS0-3, LOC1/2, MT+, Ventral occipital, PHC) and 'decision-encoding' (IPS/PCeS, PMd/v, M1; see *Encoding of normative computational variables in distinct frequency bands and cortical regions*) sets for the remaining terms (Figure 4b).

The above analysis revealed that signatures of the normative decision variable were widely distributed across cortex in the alpha frequency band (8-14 Hz). We then dissected the nature of this decision-related alpha-band activity in more detail. For a set of early/dorsal visual, intraparietal and motor regions which we assumed were ordered hierarchically[44,45] (V1, V2-4, V3A/B, IPS0/1, IPS2/3, aIPS, IPS/PCeS, PMd/v, M1), we estimated belief encoding functions for alpha-band activity in each ROI via the procedure described above (using the average *LI* between 8-14 Hz and 0.4-0.6 s post-sample for the 'decision-encoding' ROIs with the same rationale as for the motor signal; but between 0.3-0.5 s for the remaining ROIs, to minimize contamination by faster sensory responses to subsequent samples that tended to be present in these regions). We fit the slope of the encoding functions as per eq. 25. Next, we fit the linear model described by eq. 22 to *LI* data for each ROI segmented from 0 to 1.4s around sample onset, extracted t-scores for $\beta_2$, and averaged across sample positions and 8-14 Hz to yield a vector reflecting the temporal evolution of alpha-band *LLR* encoding. We flipped the sign of this vector such that the peak encoding was positive, interpolated to millisecond temporal resolution using cubic splines, and normalized such that the maximum of the vector was equal to one (normalized vector $\overline{x}_t = x_t / \max(x)$; Figure 5c). We then calculated the latency to half-maximum of the vector for each ROI (Figure 5d), as well as the timescale $\tau$ of the *LLR* encoding by fitting a decaying exponential:

$$f(\overline{x}_t) = e^{-\frac{t}{\tau}} \tag{26}$$

to the right portion of the peak-aligned vector (Figure 5e; fit by minimizing the sum of squared residuals between vector and fit via simplex).

We used bootstrapping (1,000 iterations at trial-level; as above for analysis of motor preparatory activity) to compute 95% confidence intervals for each of the three parameters (belief encoding slope, evidence encoding latency and timescale) per ROI. We tested for parameter differences between a given pair ROIs via two-tailed bootstrap hypothesis test (calculating the difference in fitted parameters between the ROI pair for each bootstrap iteration and computing a *p*-value as the proportion of the resulting distribution that was above or below zero, divided by two). Spatial gradients in the estimated parameters across the visual cortical hierarchy (V1, V2-V4, V3A/B, IPS0/1, IPS2/3, green shading in Fig. 5) were also tested for via two-tailed bootstrap hypothesis tests, fitting a line to the hierarchically-ordered parameters for each iteration of the bootstrap and computing a *p*-value as the proportion of the resulting distribution of slopes that was above or below zero, divided by two.

### *Impact of residual fluctuations on choice*

To interrogate the relevance of residual fluctuations in the ROI-specific neural signals for choice, we extracted for each ROI the time point-, frequency- and sample-specific standardized residuals $resid_{t,f,s}$ from the fit of eq. 22 – i.e. a vector that captures across-trial fluctuations in the neural signal that are not explained by decision-relevant computational variables. We then examined whether these residual signal fluctuations predicted variance in

behavioral choice over and above the choice-predictive factors identified previously, via the following logistic regression (**Model 6**):

$$\text{logit}\left[p\left(r_{trl}=right\right)\right]=\beta_0+\sum_{i=1}^{12}\beta_{1,i}\cdot LLR_{i,trl}+\sum_{j=2}^{12}\begin{pmatrix}\beta_{2,j}\cdot LLR_{j,trl}\cdot CPP_{j,trl}+\\ \beta_{3,j}\cdot LLR_{j,trl}\cdot -|\psi|_{j,trl}+\\ \beta_{4,j}\cdot resid_{t,f,j,trl}\end{pmatrix} \quad (27)$$

where $\beta_4$ captured the effect of interest per sample position *j*, and all other terms were identical to those in eq. 16. We averaged the fitted $\beta_4$ across sample positions 10-12 to increase signal-to-noise, under the reasoning that the recency effects in evidence weighting observed in participants' psychophysical kernels (Figure 2b) would correspond to stronger choice-predictive residual fluctuations toward the end of the trial. Indeed, we observed no reliable choice-predictive fluctuations for samples 2-4 (data not shown). For each ROI, we then used cluster-based permutation testing to identify time points at which the fitted weights differed significantly from zero (two-tailed one-sample *t*-test; 10,000 permutations; cluster-forming threshold of $p<0.05$; Figure 6).

***Power law exponents of intrinsic activity fluctuations***

We also assessed regional differences in the contributions of fast vs. slow frequencies to 'intrinsic' activity fluctuations (Figure 5f). For each cerebral hemisphere, ROI, and trial, we computed the power spectrum (from 1-120 Hz) of activity in the 1 s 'baseline' interval preceding trial onset using Fast Fourier transform. Power spectra were then averaged over trials and hemispheres. These regionally-specific power spectra of intrinsic activity fluctuations were modeled as a linear superposition of two functional processes[97]: an aperiodic component modeled as a decaying exponential; and a variable number of periodic components (band-limited peaks, e.g., the ~10 Hz peak). Power *P* at frequency *f* of the aperiodic component was modeled as a power law: $\log_{10}\left(P\left(f\right)\right)=10^b f^{-\chi}$, where *b* was the broadband offset of the spectrum and $\chi$ was the so-called scaling exponent[98]. The periodic components were modeled as Gaussians with parameters *A*, $\mu$ and $\sigma$, describing the amplitude, peak frequency and band-width of the component, respectively. We fit this model to the power spectra from each ROI and participant using the FOOOF toolbox[97] (default constraints and approach, without so-called 'knees' which were absent in the measured spectra, presumably due to the short intervals). The bands 49-51 Hz and 99-101 Hz, which were contaminated by line noise, were excluded from the fits. The model fits are shown in Supplementary Figure 6.

Our subsequent analyses focused on the power law scaling exponent $\chi$. We used this parameter to quantify the relative contributions of fast vs. slow activity fluctuations to the aperiodic component (large exponent equates to greater contribution of slow fluctuations). We assessed whether there was a spatial gradient in the fitted exponents across the visual cortical hierarchy (V1, V2-V4, V3A/B, IPS0/1, IPS2/3, green shading in Fig. 5) via permutation (shuffling) tests (since, unlike for previous analyses of spatial gradients relying on trial-level bootstrapping, meaningful $\chi$ estimates were here available for all participants and ROIs). We fitted a line to the hierarchically-ordered, group-average parameters. Repeating this procedure for 10,000 random shuffles of the ROI labels (shuffling at the level of individual participants) yielded a null distribution of slopes, from which we extracted the *p*-value of the slopes across ROIs obtained for unshuffled labels.

***Limitations of approach***

Despite the richness of our results, the current approach also has limitations. In particular, MEG source estimation is limited by signal leakage, due to a combination of the limited spatial resolution of the source reconstruction and volume conduction in the neural tissue. We sought to minimize these effects by restricting our analysis to a selection of ROIs (see above) with relatively low spatial granularity. Furthermore, we focused the interregional comparisons on *differences* (rather than *similarities*) – in other words, the heterogeneity of computational properties along the visuo-motor pathway (Figures 4 and 5). For example, signal leakage cannot explain the observed similarity between primary visual cortex and anterior decision-encoding regions, and their respective dissimilarity with the anatomically-intermediate clusters IPS0/1 and IPS2/3. Because of signal leakage, the interregional heterogeneity we report here constitutes a lower bound for the true heterogeneity of neural activity.

A second limitation of MEG is low sensitivity for subcortical sources. Invasive recording techniques will be required to illuminate the contributions of the striatum, superior colliculus, thalamus, and cerebellum to the decision computations that we characterize here.

## Pupillometry data analysis

### *Preprocessing*

Eye blinks and other noise transients were removed from pupillometric time series using a custom linear interpolation algorithm in which artifactual epochs were identified via both the standard Eyelink blink detection algorithm and thresholding of the first derivative of the per-block z-scored pupil time-series (threshold = $\pm 3.5 zs^{-1}$). The pupil time-series for each block was then band-pass filtered between 0.06-6Hz (Butterworth), re-sampled to 50Hz and z-scored across time. We then computed the first derivative of the resulting time-series, which was the focus of all subsequent analyses. As such, our preprocessing of the pupil signal promoted sensitivity to fast, evoked changes in pupil size and removed slower fluctuations within and across task blocks.

### *Sensitivity to change-point occurrence*

We first visualized the sensitivity of the pupil derivative signal to change-point occurrence in the following manner. The signal was segmented around -0.5–5.8 s relative to trial onset for all full-length (12-sample) trials, thus encompassing the duration of decision formation on these trials. We then discarded all trials on which >1 change-point occurred, averaged over subsets of trials on which single change-points occurred at sample positions 2-4, 5-6, 7-8, 9-10 or 11-12, and from each of the resulting traces subtracted the average signal from trials on which no change-point occurred. This procedure created 5 time-series of the pupil derivative around change-points that occurred at different points in time over the trial, relative to the case where no change-point occurred (Figure 8b). On the same plot we overlaid relative *CPP* for the same subsets of trials for comparison, as calculated from the best-fitting normative model fits.

### *Sensitivity to sample-wise computational variables and relevance for choice*

Next, we assessed the sensitivity of the pupil signal to decision-relevant computational variables derived for each presented sample from the normative model fits. Pupil size is a signal with demonstrated sensitivity to various forms of surprise and uncertainty in other task contexts[35,52,53], but not to 'signed' choice-selective variables (such as *LLR* and $\psi$ in the present case). Thus, unlike the lateralized MEG signals above for which *CPP* and $-|\psi|$ *modulate* the influence of new information on signal change, for the pupil we interrogated direct relationships

with these variables. We did so by segmenting the pupil derivative from 0 to 1s after the onset of each individual sample *s* on full-length trials, and fitting the following linear regression model (**Model 7**; Figure 8c):

$$pupil_{t,s,trl} = \beta_0 + \sum_{g=s-1}^{s} \left( \beta_{1,g} \cdot CPP_{g,trl} + \beta_{2,g} \cdot -|\psi|_{g,trl} + \beta_{3,g} \cdot |LLR|_{g,trl} \right) + \beta_4 \cdot x_gaze_{t,s,trl} + \beta_5 \cdot y_gaze_{t,s,trl} \quad (28)$$

where *t* indicated time point relative to sample onset, and *x_gaze* and *y_gaze* were the instantaneous horizontal and vertical gaze positions included to absorb a possible artifactual effect of gaze position on measured pupil size. |*LLR*| is included to capture a possible relationship between pupil and a low-level form of surprise (since |*LLR*| exhibits a monotonic negative relationship in our task with the unconditional probability of observing a given stimulus over the entire experiment; Supplementary Figure 2). We further include previous sample *CPP*, -|*ψ*| and |*LLR*| in the model because the pupil response to an eliciting impulse is relatively slow[99], and thus correlations with model-based variables from the previous sample can cause spurious correlations with those from the current sample within the time window considered here (in particular, a positive correlation that was present between $CPP_{s-1}$ and $-|\psi|_s$). As above for the MEG signals, significant effects for the terms of interest were assessed via two-tailed cluster-based permutation testing after averaging the associated t-scores over sample positions. We also delineated the sensitivity of the evoked pupil response to other candidate surprise measures, as described in Supplementary Figure 2.

We interrogated the relevance of variability in the sample-wise pupil response for choice via a similar approach to that described for the MEG signals above (**Model 8**; Figure 8d). Specifically, we fit versions of eq. 27 in which the final term ($\beta_4$) was $LLR_{s,trl} \cdot resid_{t,s,trl}$, where *resid* here refers to the standardized residuals from fits of eq. 28. Here this term takes the form of a modulation of *LLR* weighting and not a direct effect on choice as in eq. 27, which accounts for the fact that the pupil response is a modulatory signal and, unlike the lateralized MEG signals, is not sensitive to the relative evidence/belief for different choice alternatives.

***Relationship with neural signals***

Lastly, we investigated the relationship between the pupil response to each sample, and the previously identified neural signals that encode evidence strength and belief. We extracted a per-sample scalar metric of the magnitude of the pupil response at the time of peak *CPP* encoding in the pupil (*t* = 0.57s post-sample; Figure 8c), z-scored over trials at each sample position, and assessed whether this metric modulated the neural encoding of the associated evidence sample via the following linear model (**Model 9**):

$$LI_{t,f,s,trl} = \beta_0 + \beta_1 \cdot \psi_{s,trl} + \beta_2 \cdot LLR_{s,trl} + \beta_3 \cdot LLR_{s,trl} \cdot CPP_{s,trl} + \beta_4 \cdot LLR_{s,trl} \cdot -|\psi|_{s,trl} + \beta_5 \cdot LLR_{s,trl} \cdot pupil_{s,trl} + \sum_{i=1}^{n-1} \beta_{6,i} \cdot session_{i,trl} \quad (29)$$

where $LI_{t,f,s,trl}$ was the power lateralization for a given ROI. This model was an extension of eq. 22 where the term of central interest ($\beta_5$) captures the extent to which the pupil response to sample *s* enhances or suppresses the encoding of that sample in the lateralized signal, over and above the variance in the signal captured by variables from the model fits that reflect evidence strength and belief updating.

## References


1 Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen, J. D. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* **113**, 700-765 (2006).

2 Brunton, B. W., Botvinick, M. M. & Brody, C. D. Rats and humans can optimally accumulate evidence for decision-making. *Science* **340**, 95-98, doi:10.1126/science.1233912 (2013).

3 Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535-574 (2007).

4 Yang, T. & Shadlen, M. N. Probabilistic reasoning by neurons. *Nature* **447**, 1075-1080, doi:10.1038/nature05852 (2007).

5 Wald, A. & Wolfowitz, J. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics* **19**, 326–339, doi:10.1214/aoms/1177730197 (1948).

6 Brody, C. D. & Hanks, T. D. Neural underpinnings of the evidence accumulator. *Curr. Opin. Neurobiol.* **37**, 149-157, doi:10.1016/j.conb.2016.01.003 (2016).

7 Cisek, P. & Kalaska, J. F. Neural mechanisms for interacting with a world full of action choices. *Annu. Rev. Neurosci.* **33**, 269-298, doi:10.1146/annurev.neuro.051508.135409 (2010).

8 de Lange, F. P., Rahnev, D. A., Donner, T. H. & Lau, H. Prestimulus oscillatory activity over motor cortex reflects perceptual expectations. *J. Neurosci.* **33**, 1400-1410 (2013).

9 Donner, T. H., Siegel, M., Fries, P. & Engel, A. K. Buildup of choice-predictive activity in human motor cortex during perceptual decision making. *Curr. Biol.* **19**, 1581-1585, doi:10.1016/j.cub.2009.07.066 [doi] (2009).

10 Fischer, A. G., Nigbur, R., Klein, T. A., Danielmeier, C. & Ullsperger, M. Cortical beta power reflects decision dynamics and uncovers multiple facets of post-error adaptation. *Nat. Commun.* **9**, 5038, doi:10.1038/s41467-018-07456-8 (2018).

11 Hanks, T. D. *et al.* Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature* **520**, 220-223, doi:10.1038/nature14066 (2015).

12 Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78-84, doi:10.1038/nature12742 (2013).

13 Scott, B. B. *et al.* Fronto-parietal cortical circuits encode accumulated evidence with a diversity of timescales. *Neuron* **95**, 385-398, doi:10.1016/j.neuron.2017.06.013 (2017).

14 Zhou, Y. & Freedman, D. J. Posterior parietal cortex plays a causal role in perceptual and categorical decisions. *Science* **365**, 180-185, doi:10.1126/science.aaw8347 (2019).

15 Wyart, V., de Gardelle, V., Scholl, J. & Summerfield, C. Rhythmic fluctuations in evidence accumulation during decision making in the human brain. *Neuron* **76**, 847-858, doi:10.1016/j.neuron.2012.09.015 (2012).

16 O'Connell, R. G., Dockree, P. M. & Kelly, S. P. A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat. Neurosci.* **15**, 1729-1735, doi:10.1038/nn.3248 (2012).

17 Hernandez, A. *et al.* Decoding a perceptual decision process across cortex. *Neuron* **66**, 300-314, doi:10.1016/j.neuron.2010.03.031 (2010).

18 Latimer, K. W., Yates, J. L., Meister, M. L., Huk, A. C. & Pillow, J. W. Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science* **349**, 184-187, doi:10.1126/science.aaa4056 (2015).

19 Yu, A. J. & Dayan, P. Uncertainty, neuromodulation, and attention. *Neuron* **46**, 681-692, doi:10.1016/j.neuron.2005.04.026 (2005).

20 Glaze, C. M., Kable, J. W. & Gold, J. I. Normative evidence accumulation in unpredictable environments. *eLife* **4**, doi:10.7554/eLife.08825 (2015).

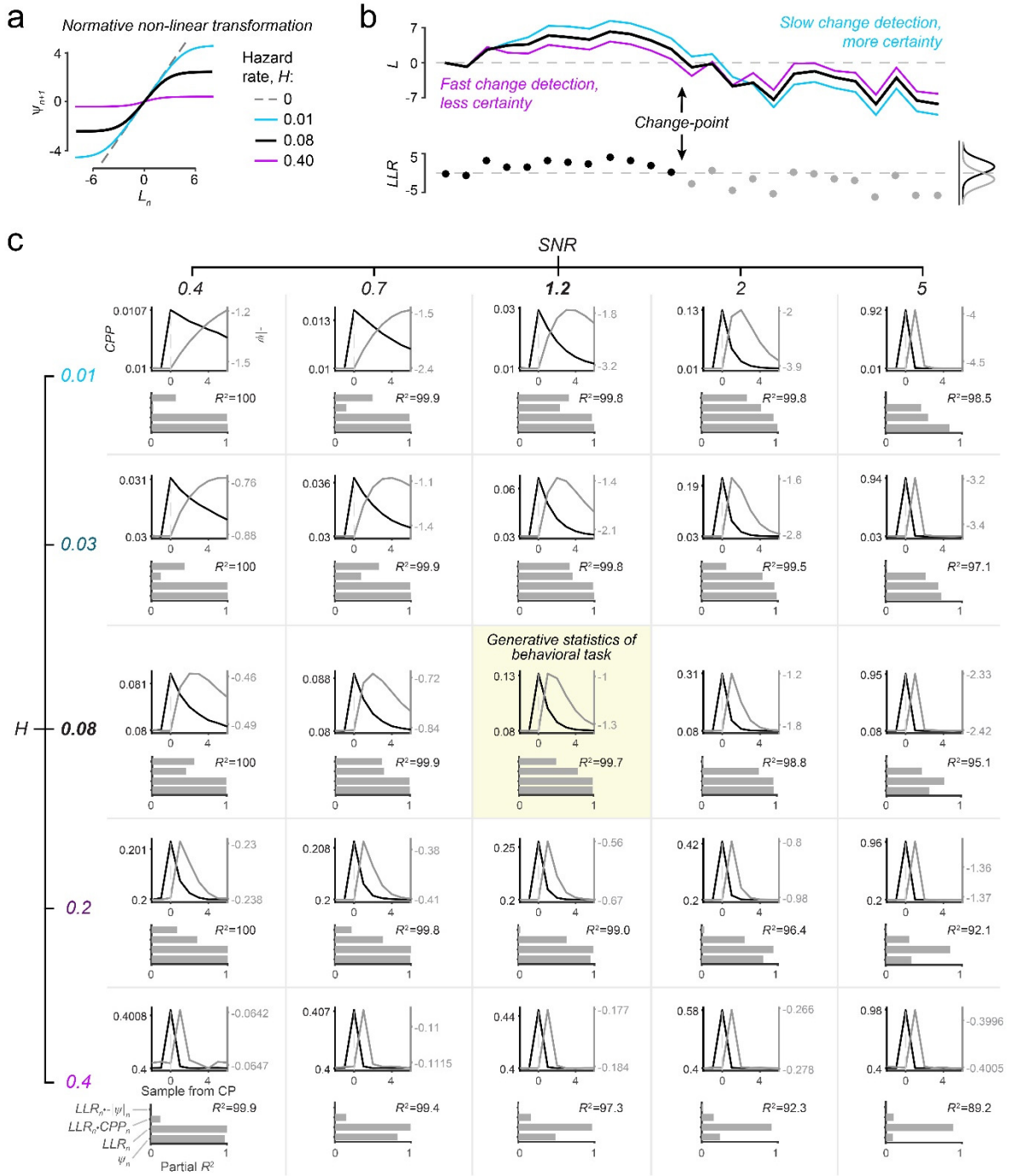
21 Ossmy, O. *et al.* The timescale of perceptual evidence integration can be adapted to the environment. *Curr. Biol.* **23**, 981-986, doi:10.1016/j.cub.2013.04.039 (2013).
22 Chaudhuri, R., Knoblauch, K., Gariel, M. A., Kennedy, H. & Wang, X. J. A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. *Neuron* **88**, 419-431, doi:10.1016/j.neuron.2015.09.008 (2015).
23 Markov, N. T. *et al.* A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cereb. Cortex* **24**, 17-36, doi:10.1093/cercor/bhs270 (2014).
24 Murray, J. D., Jaramillo, J. & Wang, X. J. Working memory and decision-making in a frontoparietal circuit model. *J. Neurosci.* **37**, 12167-12186, doi:10.1523/jneurosci.0343-17.2017 (2017).
25 Wimmer, K. *et al.* Sensory integration dynamics in a hierarchical network explains choice probabilities in cortical area MT. *Nat. Commun.* **6**, 6177, doi:10.1038/ncomms7177 (2015).
26 Wang, X. J. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* **36**, 955-968, doi:10.1016/s0896-6273(02)01092-9 (2002).
27 Aston-Jones, G. & Cohen, J. D. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* **28**, 403-450 (2005).
28 de Gee, J. W. *et al.* Dynamic modulation of decision biases by brainstem arousal systems. *eLife* **6**, doi:10.7554/eLife.23232 (2017).
29 Eckhoff, P., Wong-Lin, K.-F. & Holmes, P. Optimality and robustness of a biophysical decision-making model under norepinephrine modulation. *J. Neurosci.* **29**, 4301-4311 (2009).
30 Murphy, P. R., Boonstra, E. & Nieuwenhuis, S. Global gain modulation generates time-dependent urgency during perceptual choice in humans. *Nat. Commun.* **7**, 13526, doi:10.1038/ncomms13526 (2016).
31 Mobbs, D., Trimmer, P. C., Blumstein, D. T. & Dayan, P. Foraging for foundations in decision neuroscience: insights from ethology. *Nat. Rev. Neurosci.* **19**, 419-427, doi:10.1038/s41583-018-0010-7 (2018).
32 Veliz-Cuba, A., Kilpatrick, Z. P. & Josic, K. Stochastic models of evidence accumulation in changing environments. *SIAM Review* **58**, 264-289 (2016).
33 Marr, D. & Poggio, T. From understanding computation to understanding neural circuitry. (Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Cambridge, MA, 1976).
34 Dayan, P. & Yu, A. J. Phasic norepinephrine: a neural interrupt signal for unexpected events. *Network* **17**, 335-350, doi:10.1080/09548980601004024 (2006).
35 Nassar, M. R. *et al.* Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat. Neurosci.* **15**, 1040-1046 (2012).
36 Nienborg, H. & Cumming, B. G. Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature* **459**, 89-92, doi:10.1038/nature07821 (2009).
37 Okazawa, G., Sha, L., Purcell, B. A. & Kiani, R. Psychophysical reverse correlation reflects both sensory and decision-making processes. *Nat. Commun.* **9**, 3479, doi:10.1038/s41467-018-05797-y (2018).
38 Piet, A. T., El Hady, A. & Brody, C. D. Rats adopt the optimal timescale for evidence integration in a dynamic environment. *Nat. Commun.* **9**, 4265, doi:10.1038/s41467-018-06561-y (2018).
39 Siegel, M., Donner, T. H. & Engel, A. K. Spectral fingerprints of large-scale neuronal interactions. *Nat. Rev. Neurosci.* **13**, 121-134, doi:10.1038/nrn3137 [doi] (2012).
40 Bondy, A. G., Haefner, R. M. & Cumming, B. G. Feedback determines the structure of correlated variability in primary visual cortex. *Nat. Neurosci.* **21**, 598-606, doi:10.1038/s41593-018-0089-1 (2018).

41 Siegel, M., Buschman, T. J. & Miller, E. K. Cortical information flow during flexible sensorimotor decisions. *Science* **348**, 1352-1355, doi:10.1126/science.aab0551 (2015).
42 Wang, L., Mruczek, R. E., Arcaro, M. J. & Kastner, S. Probabilistic maps of visual topography in human cortex. *Cereb. Cortex* **25**, 3911-3931, doi:10.1093/cercor/bhu277 (2015).
43 Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171-178, doi:10.1038/nature18933 (2016).
44 Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1-47, doi:10.1093/cercor/1.1.1 (1991).
45 Michalareas, G. *et al.* Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron* **89**, 384-397, doi:10.1016/j.neuron.2015.12.018 (2016).
46 Murray, J. D. *et al.* A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* **17**, 1661-1663, doi:10.1038/nn.3862 (2014).
47 Wilming, N., Murphy, P. R., Meyniel, F. & Donner, T. H. Disentangling decision-related feedback and feedforward signal in human cortex. *bioRxiv* (2020).
48 Arnsten, A. F., Paspalas, C. D., Gamo, N. J., Yang, Y. & Wang, M. Dynamic Network Connectivity: A new form of neuroplasticity. *Trends Cogn. Sci.* **14**, 365-375, doi:10.1016/j.tics.2010.05.003 (2010).
49 Brunel, N. & Wang, X. J. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J. Comput. Neurosci.* **11**, 63-85, doi:10.1023/a:1011204814320 (2001).
50 Froemke, R. C. Plasticity of cortical excitatory-inhibitory balance. *Annu. Rev. Neurosci.* **38**, 195-219, doi:10.1146/annurev-neuro-071714-034002 (2015).
51 Pfeffer, T. *et al.* Catecholamines alter the intrinsic variability of cortical population activity and perception. *PLoS biology* **16**, e2003453, doi:10.1371/journal.pbio.2003453 (2018).
52 Krishnamurthy, K., Nassar, M. R., Sarode, S. & Gold, J. I. Arousal-related adjustments of perceptual biases optimize perception in dynamic environments. *Nat. Hum. Behav.* **1**, doi:10.1038/s41562-017-0107 (2017).
53 Muller, T. H., Mars, R. B., Behrens, T. E. & O'Reilly, J. X. Control of entropy in neural models of environmental state. *eLife* **8**, doi:10.7554/eLife.39404 (2019).
54 Breton-Provencher, V. & Sur, M. Active control of arousal by a locus coeruleus GABAergic circuit. *Nat. Neurosci.* **22**, 218-228, doi:10.1038/s41593-018-0305-z (2019).
55 Joshi, S., Li, Y., Kalwani, R. M. & Gold, J. I. Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron* **89**, 221-234, doi:10.1016/j.neuron.2015.11.028 (2016).
56 Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H. & Balsters, J. H. Pupil diameter covaries with BOLD activity in human locus coeruleus. *Hum. Brain Mapp.* **35**, 4140-4154 (2014).
57 Reimer, J. *et al.* Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nat. Commun.* **7**, 13289, doi:10.1038/ncomms13289 (2016).
58 Behrens, T. E., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. Learning the value of information in an uncertain world. *Nat. Neurosci.* **10**, 1214-1221, doi:10.1038/nn1954 (2007).
59 Mathys, C. D. *et al.* Uncertainty in perception and the Hierarchical Gaussian Filter. *Front. Hum. Neurosci.* **8**, 825, doi:10.3389/fnhum.2014.00825 (2014).
60 McGuire, J. T., Nassar, M. R., Gold, J. I. & Kable, J. W. Functionally dissociable influences on learning rate in a dynamic environment. *Neuron* **84**, 870-881, doi:10.1016/j.neuron.2014.10.013 (2014).
61 O'Reilly, J. X. *et al.* Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E3660-3669, doi:10.1073/pnas.1305373110 (2013).

62 Donner, T. H. & Siegel, M. A framework for local cortical oscillation patterns. *Trends Cogn. Sci.* **15**, 191-199, doi:10.1016/j.tics.2011.03.007 (2011).
63 Bastos, A. M. *et al.* Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* **85**, 390-401, doi:10.1016/j.neuron.2014.12.018 (2015).
64 van Kerkoerle, T. *et al.* Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 14332-14341, doi:10.1073/pnas.1402773111 (2014).
65 Erlich, J. C., Brunton, B. W., Duan, C. A., Hanks, T. D. & Brody, C. D. Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat. *eLife* **4**, doi:10.7554/eLife.05457 (2015).
66 Katz, L. N., Yates, J. L., Pillow, J. W. & Huk, A. C. Dissociated functional significance of decision-related activity in the primate dorsal stream. *Nature* **535**, 285-288, doi:10.1038/nature18617 (2016).
67 Cisek, P. Cortical mechanisms of action selection: the affordance competition hypothesis. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **362**, 1585-1599, doi:10.1098/rstb.2007.2054 (2007).
68 Inagaki, H. K., Fontolan, L., Romani, S. & Svoboda, K. Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature* **566**, 212-217, doi:10.1038/s41586-019-0919-7 (2019).
69 Amit, D. J. & Brunel, N. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex* **7**, 237-252, doi:10.1093/cercor/7.3.237 (1997).
70 Machens, C. K., Romo, R. & Brody, C. D. Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science* **307**, 1121-1124, doi:10.1126/science.1104171 (2005).
71 Harris, K. D. & Thiele, A. Cortical state and attention. *Nat. Rev. Neurosci.* **12**, 509-523, doi:10.1038/nrn3084 (2011).
72 Iigaya, K. Adaptive learning and decision-making under uncertainty by metaplastic synapses guided by a surprise detection system. *eLife* **5**, doi:10.7554/eLife.18073 (2016).
73 Glaze, C. M., Filipowicz, A. L. S., Kable, J. W., Balasubramanian, V. & Gold, J. I. A bias–variance trade-off governs individual differences in on-line learning in an unpredictable environment. *Nat. Hum. Behav.* **2**, 213–224, doi:https://doi.org/10.1038/s41562-018-0297-4 (2018).
74 Niyogi, R. K. & Wong-Lin, K. Dynamic excitatory and inhibitory gain modulation can produce flexible, robust and optimal decision-making. *PLoS Comput. Biol.* **9**, e1003099, doi:10.1371/journal.pcbi.1003099 (2013).
75 Teufel, H. J. & Wehrhahn, C. Evidence for the contribution of S cones to the detection of flicker brightness and red–green. *J. Opt. Soc. Am. A. Opt. Image Sci. Vis.* **17**, 994-1006 (2000).
76 Roitman, J. D. & Shadlen, M. N. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.* **22**, 9475-9489, doi:22/21/9475 [pii] (2002).
77 Brainard, D. H. The psychophysics toolbox. *Spat. Vis.* **10**, 433–436 (1997).
78 Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
79 Ratcliff, R., Smith, P. L., Brown, S. D. & McKoon, G. Diffusion decision model: current issues and history. *Trends Cogn. Sci.* **20**, 260-281, doi:10.1016/j.tics.2016.01.007 (2016).
80 Birge, B. in *Proc. IEEE Swarm Intell. Symp.* 182-186 (IEEE Swarm Intelligence Symposium, Indianapolis, IN, 2003).
81 Drugowitsch, J., Wyart, V., Devauchelle, A. D. & Koechlin, E. Computational precision of mental inference as critical source of human choice suboptimality. *Neuron* **92**, 1398-1411, doi:10.1016/j.neuron.2016.11.005 (2016).

82 Nassar, M. R., Wilson, R. C., Heasly, B. & Gold, J. I. An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J. Neurosci.* **30**, 12366-12378, doi:10.1523/jneurosci.0822-10.2010 (2010).
83 Spitzer, B., Waschke, L. & Summerfield, C. Selective overweighting of larger magnitudes during noisy numerical comparison. *Nat. Hum. Behav.* **1**, doi:https://doi.org/10.1038/s41562-017-0145 (2017).
84 Tsetsos, K., Chater, N. & Usher, M. Salience driven value integration explains decision biases and preference reversal. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 9659-9664, doi:10.1073/pnas.1119569109 (2012).
85 Cheadle, S. *et al.* Adaptive gain control during human perceptual choice. *Neuron* **81**, 1429-1441, doi:10.1016/j.neuron.2014.01.020 (2014).
86 Talluri, B. C., Urai, A. E., Tsetsos, K., Usher, M. & Donner, T. H. Confirmation bias through selective overweighting of choice-consistent evidence. *Curr. Biol.* **28**, 3128-3135.e3128, doi:10.1016/j.cub.2018.07.052 (2018).
87 Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177-190, doi:10.1016/j.jneumeth.2007.03.024 (2007).
88 Roxin, A. & Ledberg, A. Neurobiological models of two-choice decision making can be reduced to a one-dimensional nonlinear diffusion equation. *PLoS Comput. Biol.* **4**, e1000046, doi:10.1371/journal.pcbi.1000046 (2008).
89 Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J. M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 156869, doi:10.1155/2011/156869 (2011).
90 Gramfort, A. *et al.* MEG and EEG data analysis with MNE-Python. *Frontiers in neuroscience* **7**, 267, doi:10.3389/fnins.2013.00267 (2013).
91 Gramfort, A. *et al.* MNE software for processing MEG and EEG data. *Neuroimage* **86**, 446-460, doi:10.1016/j.neuroimage.2013.10.027 (2014).
92 Van Veen, B. D., van Drongelen, W., Yuchtman, M. & Suzuki, A. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans. Biomed. Eng.* **44**, 867-880, doi:10.1109/10.623056 (1997).
93 Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179-194 (1999).
94 Fischl, B., Sereno, M. I. & Dale, A. M. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* **9**, 195-207 (1999).
95 Wandell, B. A., Dumoulin, S. O. & Brewer, A. A. Visual field maps in human cortex. *Neuron* **56**, 366-383, doi:10.1016/j.neuron.2007.10.012 (2007).
96 Gross, J., Timmermann, L., Kujala, J., Salmelin, R. & Schnitzler, A. Properties of MEG tomographic maps obtained with spatial filtering. *Neuroimage* **19**, 1329-1336, doi:10.1016/s1053-8119(03)00101-0 (2003).
97 Haller, M. *et al.* Parameterizing neural power spectra. *bioRxiv* **299859**, doi:https://doi.org/10.1101/299859 (2018).
98 Miller, K. J., Sorensen, L. B., Ojemann, J. G. & den Nijs, M. Power-law scaling in the brain surface electric potential. *PLoS Comput. Biol.* **5**, e1000609, doi:10.1371/journal.pcbi.1000609 (2009).
99 Hoeks, B. & Levelt, W. J. Pupillary dilation as a measure of attention: A quantitative system analysis. *Behav. Res. Meth. Instrum. Comput.* **25**, 16–26, doi:https://doi.org/10.3758/BF03204445 (1993).

## Supplementary Figures

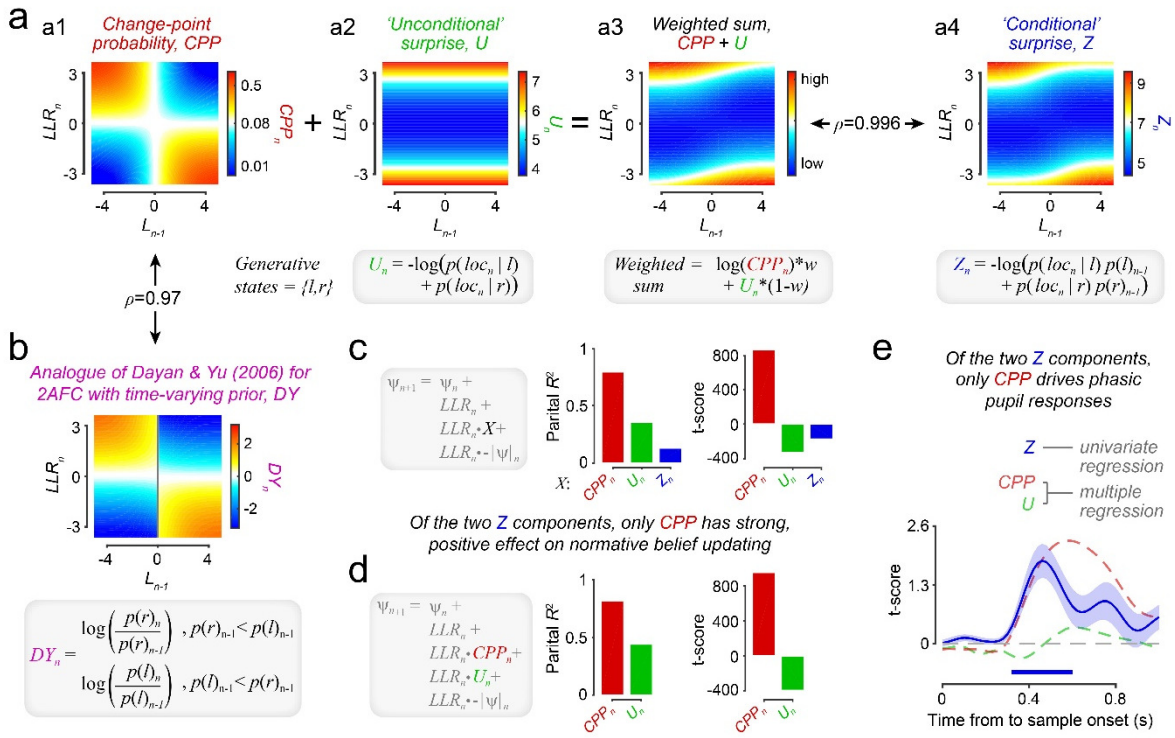



**Supplementary Figure 1. Sensitivity of normative evidence accumulation to change-point probability and uncertainty across a range of generative task statistics.**

(**a**) Non-linearity in normative model for different hazard rates ($H$): Posterior belief after accumulating the most recent sample ($L_n$) is converted into prior belief for next sample ($\psi_{n+1}$, both variables expressed as log-odds for each alternative) through a non-linear transform, which saturates (slope≈0) for strong $L_n$ and entails more moderate information loss (0<<slope<1) for weak $L_n$. In a static environment ($H$=0), the model combines new evidence (log-likelihood ratio for sample $n$, $LLR_n$) with prior ($\psi_n$) into updated belief without information loss (i.e. $L_n=\psi_n+LLR_n$; $\psi_{n+1}=L_n$). In an unpredictable environment ($H$=0.5), the new belief is determined solely by the most recent evidence sample (i.e. $L_n=LLR_n$; $\psi_{n+1}=0$) so that no evidence accumulation takes place.

(**b**) Example trajectories of the model decision variable (*L*) as function of the same evidence stream, but for different levels of *H*.

(**c**) Upper panels of each grid segment: Change point-triggered dynamics of change-point probability (*CPP*) and uncertainty (-|*ψ*|) derived from normative model as function of *H* (grid rows) and evidence signal-to-noise ratio (SNR, difference in generative distribution means over their standard deviation; grid columns). Lower panels of each grid point: Contribution of computational variables (including *CPP* and -|*ψ*|) to normative belief updating, expressed as coefficients of partial determination for terms of a linear model predicting the updated prior belief for a forthcoming sample. Yellow background (center of grid): generative statistics of our task; reproduced from main Figure 1e for comparison. Data for each parameter combination were derived from a simulated sequence of 10 million observations.




**Supplementary Figure 2. Change-point probability modulates normative belief updating and phasic arousal responses more strongly than alternative surprise metrics.**

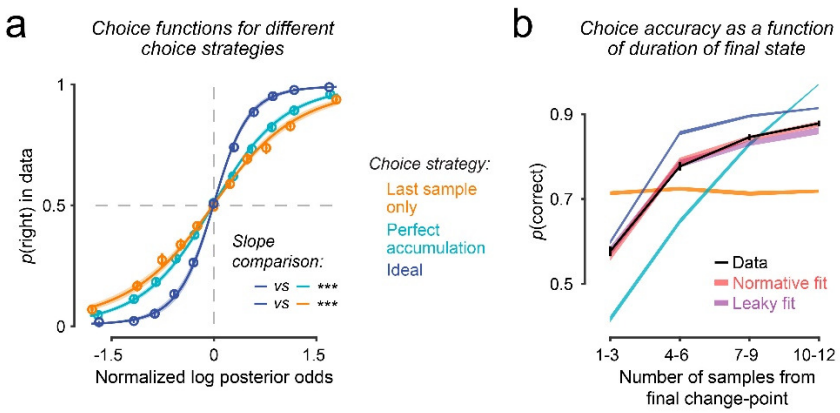
(**a**) Alternative surprise metrics as function of posterior belief after previous sample ($L_{n-1}$) and evidence provided by the current sample ($LLR_n$). **(a1)** *CPP* (see main text). **(a2)** 'Unconditional' surprise, *U*, defined as Shannon surprise (negative log probability) associated with new sample location ('$loc_n$'), given only knowledge of the generative distributions associated with each environmental state *S={l,r}.* In our task, *U* varied monotonically with absolute evidence strength *(|LLR|)*. **(a4)** 'Conditional' surprise *Z*: Shannon surprise associated with new sample, conditioned on *both* knowledge of the generative distributions *and* one's current belief about the environmental state. **(a3)** *Z* was closely approximated (Spearman's $\rho$=0.996) by a linear combination of *log*(*CPP*) and *U* (weights: *w* and 1-*w*, respectively; determined by Nelder-Mead simplex optimization as *w*=0.274).

(**b**) Characterization of another form of surprise derived from a model of phasic locus coeruleus activity by ref. 34, hence denoted as 'Dayan & Yu surprise' and abbreviated as *DY*. For the oddball target detection task modelled in ref. 34, *DY* was defined as the ratio between the fixed prior probability of an unlikely environmental state and its posterior probability given a new stimulus. In our 2AFC task, where the prior varies during decision formation, we defined *DY* as indicated below the heatmap. We found that this *DY* measure was highly similar to *CPP* ($\rho$=0.97; compare panels b and a1).

(**c**) Modulations of normative belief updating by *CPP*, *U* and *Z*. For each surprise metric, we fit a linear model to updated prior belief (same form as **Model 1;** inset; *X*, surprise metric). Left: coefficients of partial determination; right: contributions expressed as t-scores. *CPP* exerted the strongest and only *positive* modulatory effect on evidence weighting, consistent with the observed effects on behavior, cortical dynamics, and pupil (see main text).

(**d**) As panel c, but for expanded linear model that included modulations of evidence weighting by *CPP* and *U*, which combine linearly into *Z*. Only *CPP* yielded a robust, positive effect on evidence weighting.

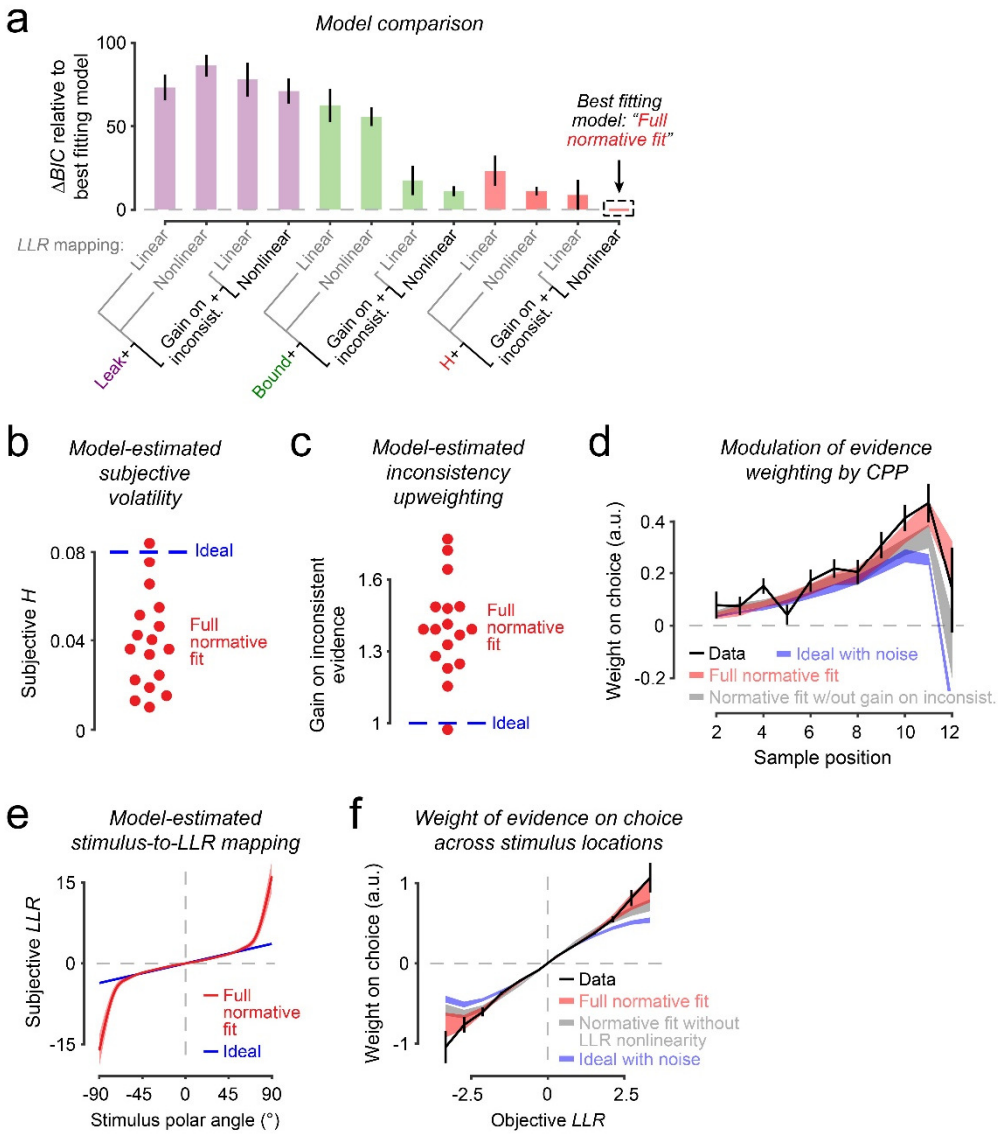
(**e**) Encoding of surprise metrics in pupil response (temporal derivative) to evidence samples. Encoding time course for *Z* was computed by a regression model that included X- and Y-gaze positions as nuisance variables. Encoding time courses for *CPP* and *U* were derived from an expanded regression model that included both terms. Only *CPP* had a significant effect, indicating that *CPP* alone accounted for the Z encoding effect observed via the simpler regression model. Shaded area, s.e.m., significance bars, p<0.05 (two-tailed cluster-based permutation test).




**Supplementary Figure 3. Consistency of human choices with those of idealized choice strategies, and dependence of choice accuracy on final state duration.**

(**a**) Human choices as function of log-posterior odds (z-scored) derived from each of three alternative strategies: basing choices only on final evidence sample (orange); perfect evidence accumulation (cyan); and an ideal observer with perfect knowledge of the generative task statistics and employing the normative accumulation process for the task (blue). Points and error bars show observed data ± s.e.m.; lines and shaded areas show mean ± s.e.m. of fits of sigmoid functions to the data. Slopes of fitted sigmoids were steepest for ideal observer (ideal vs. perfect accumulation: $t_{16}=5.9$, $p<0.0001$; ideal vs. last-sample: $t_{16}=9.6$, $p<10^{-7}$; two-tailed paired t-tests), indicating that human choices were most consistent with those of the ideal observer.

(**b**) Choice accuracy as a function of duration of the final environmental state for the human participants (black), idealized strategies from panel a, and fits of the normative model (light red) and a leaky accumulator model (light purple) to the human choices. Participants' choice accuracy increased with the number of samples presented after the final state change on each trial, consistent with temporal accumulation. Error bars and shaded regions indicate s.e.m.




**Supplementary Figure 4. Model comparison and qualitative signatures indicate approximation of normative belief updating by measured human behavior.**

(**a**) Bayes Information Criteria (BIC) for all models fit to the human choice data, relative to best-fitting model. Bars show mean ± s.e.m. Model constraints are specified by the tick labels. Colors denote leaky accumulator (purple), perfect accumulator with non-absorbing bounds (green), and the normative accumulation process with a subjective hazard rate (red). Labels refer to the following: 'Linear' ('Nonlinear')'=linear (non-linear) scaling of stimulus-to-LLR mapping function; 'Gain on inconist.'=multiplicative gain term applied to inconsistent samples; 'Leak'=accumulation leak; 'Bound'= height of non-absorbing bounds; 'H'=subjective hazard rate. All models included a noise term applied to the final log-posterior odds per trial. See *Methods* for additional model details. Labels in black text highlight models plotted in main Figure 2 and Supplementary Figure 5. Best fitting model ('full normative fit' in remaining panels) employs normative accumulation with subjectivity in hazard rate, a gain term on inconsistent samples, and non-linear stimulus-to-LLR mapping.

(**b**) Subjective hazard rates from full normative fits, with true hazard rate in blue. Participants underestimated the volatility in the environment ($t_{16}$=-7.7, $p<10^{-6}$, two-tailed one-sample t-test of subjective *H* against true *H*).

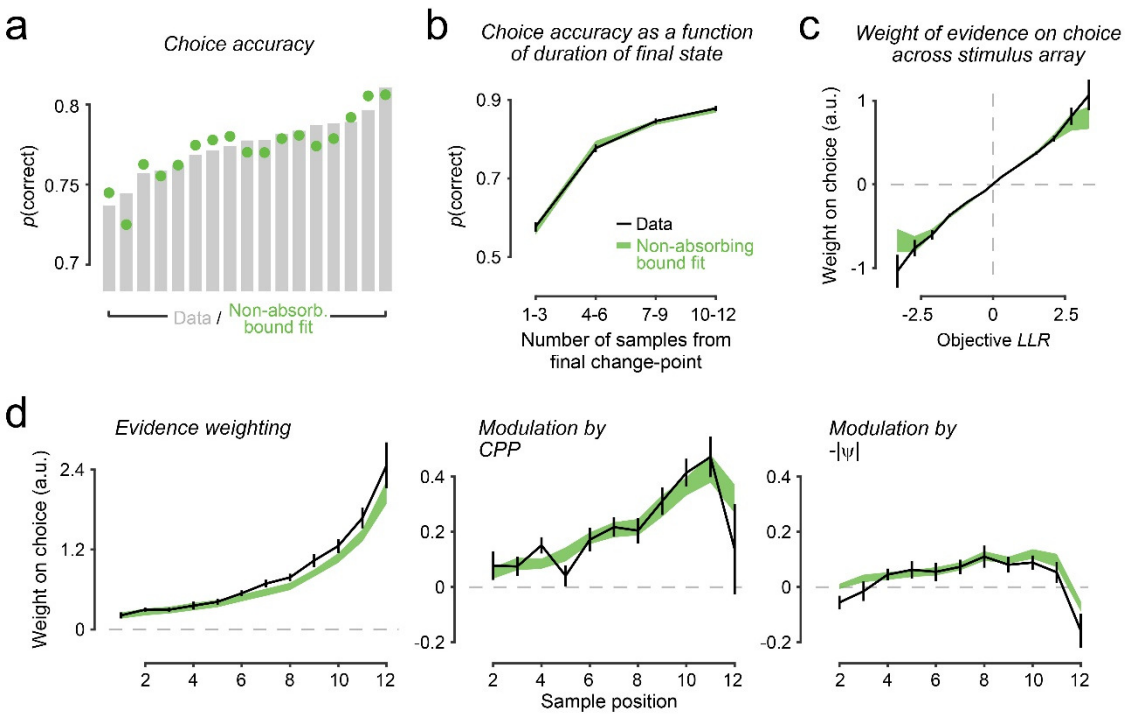
(**c**) Multiplicative gain factors applied to evidence samples that were inconsistent with the existing belief state, estimated from full normative fits. Gain factor>1 reflects relative up-weighting of inconsistent samples beyond that prescribed by normative model; gain factor<1 reflects relative down-weighting of

inconsistent samples. The ideal observer employing the normative accumulation process uses gain factor=1 (dashed blue line). Participants assigned higher weight to inconsistent samples than the ideal observer ($t_{16}$=8.2, $p<10^{-6}$, two-tailed one-sample *t*-test of fitted weights against 1).

(**d**) Regression coefficients reflecting modulation of evidence weighting by change-point probability (*CPP*). Shown are coefficients for the human participants (black line), full normative fits (light red), normative model fits without gain factor applied to inconsistent evidence (grey), and the ideal observer with matched noise. Error bars and shaded regions, s.e.m.

(**e**) Mapping of stimulus location (polar angle; x-axis) onto evidence strength (*LLR*; y-axis) across the full range of stimulus locations. Blue line reflects the true mapping given the task generative statistics, used by the ideal observer. Red line and shaded areas show mean ± s.e.m of subjective mappings used by the human participants, estimated as an interpolated non-parametric function (*Methods*).

(**f**) Regression coefficients reflecting subjective weight of evidence associated with binned stimulus locations, for the human participants (black line), full normative model fits including the non-linear *LLR* mapping shown in panel e (light red), normative model fits allowing only a linear *LLR* mapping (grey), and the ideal observer with matched noise. Error bars and shaded regions, s.e.m.




**Supplementary Figure 5. Fit to human behavior of a model with perfect evidence accumulation toward non-absorbing bounds.**

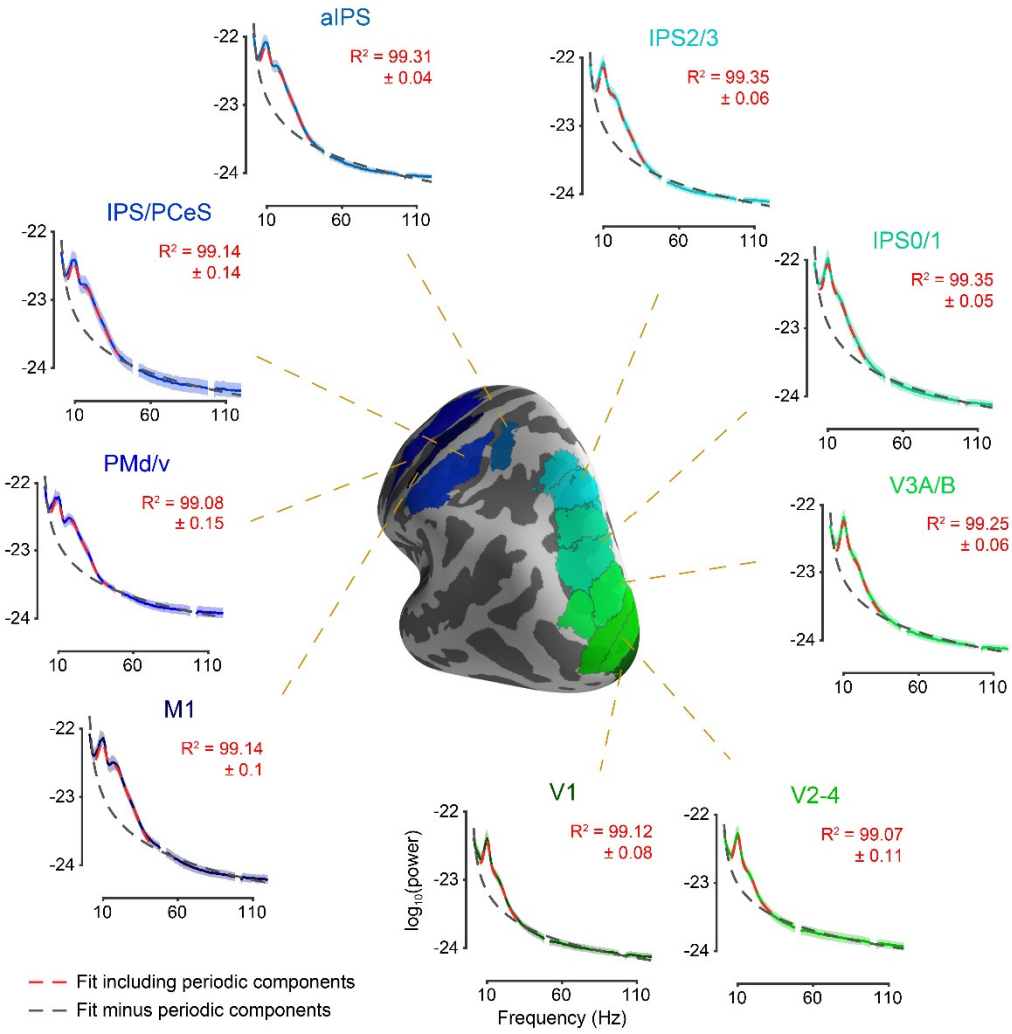
This form of bound imposes upper and lower limits on the evolving decision variable without terminating the decision process, thus enabling changes of mind even after the bound has been reached.

(**a**) Choice accuracies for human participants (grey bars) and the model (green).

(**b**) Choice accuracy as a function of duration of the final environmental state, for the participants and the model.

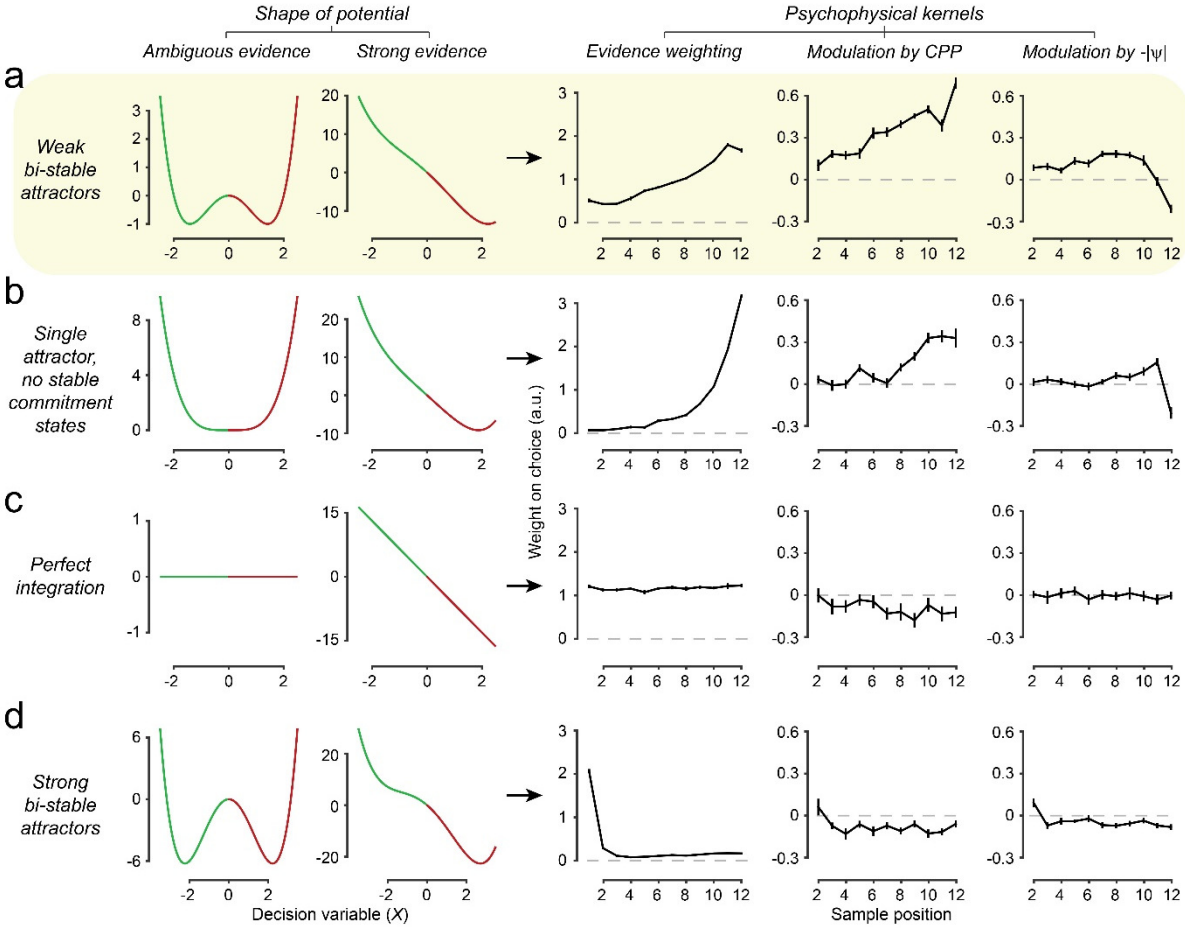
(**c**) Regression coefficients reflecting the subjective weight of evidence associated with binned stimulus locations, for the participants and the model.

(**d**) Psychophysical kernels for the data and the model reflecting the time-resolved weight of evidence on final choice (left), and the modulation of this evidence weighting by sample-wise change-point probability (middle) and uncertainty (right). In panels b-d, error bars and shaded regions indicate s.e.m. Fits of the model version used to generate model behavior in all panels included free parameters for both a non-linear stimulus-to-LLR mapping function and a gain factor on inconsistent samples (see Supplementary Figure 4). Like the normative model, perfect accumulation to non-absorbing bounds captured all qualitative features identified in the behavioral data. This is in line with the insight that this form of accumulation closely approximates the normative model in settings like ours, where strong belief states are formed often (i.e. low noise and/or low H; ref. 20). This form of accumulation also captures a key feature of the dynamics of the circuit model in ref. 26 that we interrogate in main Figure 7; i.e. a saturating decision variable in response to consecutive consistent samples of evidence.




**Supplementary Figure 6. Power spectra and associated model fits of intrinsic fluctuations across cortex.**

Power spectra were estimated for the set of ROIs spanning the visuo-motor pathway and used for main Figures 5,6, and 8. For each hemisphere, region, and trial, we computed the 'baseline' power spectrum from a 1-second period preceding trial onset and averaged these spectra across trials and hemispheres. Colored blue/green lines in each plot show mean across participants; shading indicates s.e.m. We modeled the power spectra as a linear superposition of an aperiodic component (power law scaling) and a variable number of periodic components (band-limited peaks, *Methods*). Full model fits are shown via dashed red lines with the corresponding goodness-of-fit ($R^2$ ± s.e.m.) at inset; the aperiodic component only is overlaid in gray. We plotted the scaling exponent of the aperiodic component in main Figure 5f, as a measure of the relative contributions of fast vs. slow fluctuations to the measured power spectrum.




**Supplementary Figure 7. Assessment of boundary conditions at which circuit model approximates normative evidence accumulation.**

A reduction of the spiking-neuron circuit model (main Figure 7; *Methods*) was used to explore the impact of different dynamical regimes on behavioral signatures of evidence weighting. Left two columns: shape of model's energy landscape ('potentials') in response to ambiguous evidence (*LLR*=0, left), or in response to strong evidence (*LLR*=mean+1 s.d. of *LLRs* across experiment=1.35, second from left). Middle-to-right columns: psychophysical kernels and modulations of evidence weighting by *CPP* and -|*ψ*|.

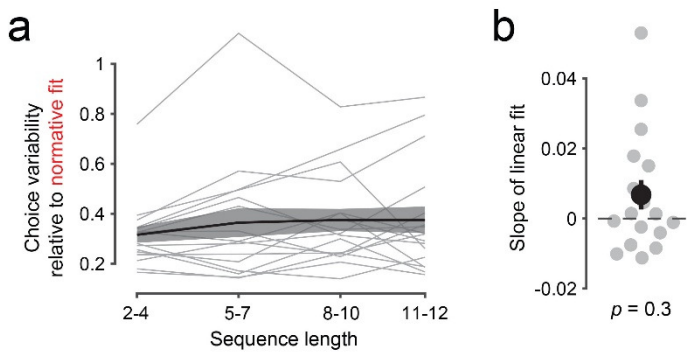
(**a**) Double-well potential with small barrier between wells, corresponding to weak bi-stable attractor dynamics. This model variant featured both a saturating decision variable in the face of multiple consistent evidence samples (akin to non-absorbing bounds), and wells that maintained commitment states but were sufficiently shallow for changes-of-mind in response to inconsistent evidence. It produced all key qualitative features of normative evidence accumulation (recency, strong modulation by CPP; weak modulation by uncertainty).

(**b**) Single-well potential with no barrier. This model variant also featured a saturating decision variable, but no stable states of commitment. Thus, it exhibited stronger recency in evidence weighting than the one in panel a.

(**c**) Flat potential with no wells, corresponding to perfect evidence accumulation without bounds. This model lacked both a saturating decision variable and stable states. As a consequence, it weighed all evidence samples equally (flat psychophysical kernel) and did not produce clear modulations by CPP or uncertainty.

(**d**) Double-well potential with high barrier between wells, corresponding to strong bi-stable attractor dynamics. In this model variant, the wells were too deep for changes-of-mind to occur for all but the most extreme inconsistent evidence. Thus, it produced extreme primacy in evidence weighting.

The evidence weighting signatures of the regimes in panels c, d were qualitatively inconsistent with the signatures of participants' behavior (compare Figure 2b). All three signatures were qualitatively reproduced by the regimes in both a and b, with the closest approximation to human behavior provided by the weak bi-stable attractor regime in panel a (highlighted in yellow).




**Supplementary Figure 8. Variability of human choices relative to choices of the normative model fits is invariant to sequence length.**

(**a**) Participants' choice variability relative to the best-fitting variant of the normative model ('full normative fit' in Supplementary Figure 4), for each of four different bins of trial sequence lengths. Choice variability is quantified as the cross entropy between participant and model choices. Thin gray lines show individual participants; thick black line and gray shading, mean ± s.e.m.

(**b**) Slopes of linear fits to the per-participant choice variability functions shown in panel a. Gray points show individual participants; black point and error bar show mean ± s.e.m. Displayed *p*-value is from a two-tailed Wilcoxon signed rank test ($H_0$: slope = 0). The finding that choice variability was invariant to sequence length indicates that sample-wise 'accumulation noise' was not necessary to explain the behavioral data.