Evidence for a probabilistic, brain-distributed, recursive mechanism for decision-making

JA Caballero[1,2,†], MD Humphries[1*], KN Gurney[2*]
[1] Faculty of Life Sciences, University of Manchester, Manchester, UK
[2] Dept. of Psychology, University of Sheffield, Sheffield, UK
[*] Joint senior author
[†] corresponding author, j.caballero@manchester.ac.uk

Abstract

Decision formation recruits many brain regions, but the computation they jointly perform is unknown. We introduce a recursive ideal Bayesian observer that makes choices using evidence streams with the statistics of spike trains in macaque sensory cortex. We show it decides faster than monkeys, indicating they lose information from their sensory cortex; conversely, if equivalently deprived of information, its choice behaviour quantitatively matches that of monkeys. The algorithm's recursive architecture maps to the recurrent cortico-basal-ganglia-thalamo-cortical loops, whose components have all been implicated in decision-making. With this mapping, we show the dynamics of variables within the algorithm qualitatively match those recorded from neurons in the monkey cortex and striatum during decisions, and predict the dynamics of basal ganglia output and thalamus. Our principled, single-equation algorithm is probabilistic, distributed, recursive and parallel. Its success at capturing anatomy, behaviour and electrophysiology suggests that the mechanism approximated by the brain has these same characteristics.

1 Introduction

Decisions rely on evidence that is collected for, accumulated about, and contrasted between available options. Neural activity congruent with accumulation over time has been reported in parietal and frontal cortex [14, 16, 27, 31, 46]. Other regions, particularly the basal ganglia [21, 32, 38, 44], have long been implicated in decision making. Consistent with this, decision-correlated cumulative activity has been recorded within the striatum [15], and appears to be causal, as perturbations of it affect decisions [17]. A diverse range of models can account for cumulative neural signals in cortex and associated behavioural data [1, 6, 18, 22, 24, 39, 52]. Few address the role of subcortical structures [8, 38], and none provide a complete account of subcortical and cortical signals in decision-making. Consequently, given these diverse and incomplete accounts, it is unclear what decision-making algorithm is being approximated by the brain.

Our solution was to start from a principled approach. Normative models of neural computations provide useful benchmarks for understanding how well the brain approximates an optimal algorithmic computation. An attractive normative framework for understanding decision-making is the family of multi-hypothesis sequential probability ratio tests (MSPRT) [5]. These evidence accumulation based algorithms are the ideal Bayesian observer for two-alternative decisions, and an asymptotic approximation to it for decisions between more than two [5, 8]. Prior work has shown how the algorithm's components can be plausibly mapped onto a range of neural circuits centered on the basal ganglia [8, 34, 37]. Unclear is whether the MSPRT can account for both behavioural data and known neural signals across cortex and striatum during decision-making.

We introduce a recursive generalization of the MSPRT (rMSPRT) that simultaneously addresses these challenges. We show the rMSPRT's computational elements map onto the cortico-basal-ganglia-thalamo-cortical loops, providing a unified cortical-subcortical model. We show the rMSPRT and its mapping can account for both behavioural data and the cortical and striatal activity from primates performing the standard random dot motion task. A particular advantage of the principled approach is that we are able to predict which aspects of the neural dynamics are and are not part of inference during decision-making. Our results suggest that the mammalian brain approximates a sequential Bayes-based decision-making algorithm, within the constraints of available sensory information.

2 Results

We tested our models against behavioural and electrophysiological data recorded by [14], from monkeys performing 2- and 4-alternative reaction-time versions of the random dot motion task (Fig. 1b,c; Methods). In all forms of the task, the monkey observes the motion of dots and indicates the dominant direction of motion with a saccade to a target in that direction. Task difficulty is controlled by the coherence of the motion: the percentage of dots moving in the target's direction.

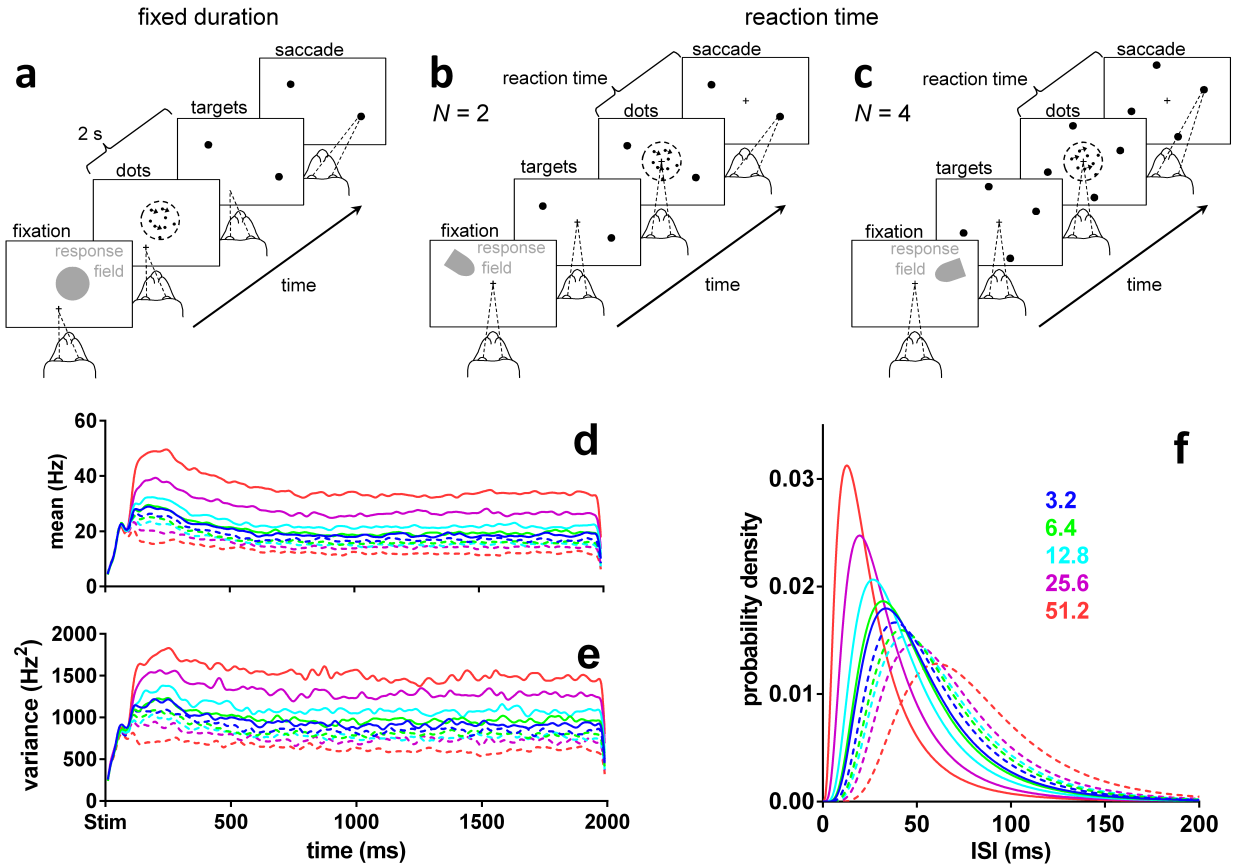

Figure 1: Random dot motion task and MT data. (a) Fixed duration task for MT recordings [10]. (b, c) Reaction time task for LIP recordings, $N = 2, 4$ [14]. (d, e) Population mean and variance of the firing rate in MT during the fixed duration dot motion task, aligned at onset of the dot stimulus (Stim), for a variety of coherence percentages (colour-coded as in the legend in panel f). Solid lines are statistics when dots were moving in the preferred direction of the MT neuron. Dashed lines are statistics when dots were moving in the opposite, null direction. Data from [10], re-analysed. (f) Lognormal density functions for the inter-spike intervals (ISI) specified by the statistics over the approximately stationary segment of (d, e) before smoothing (parameter set Ω in table 1; see Methods). Preferred and null motion directions by line type as in (d, e).

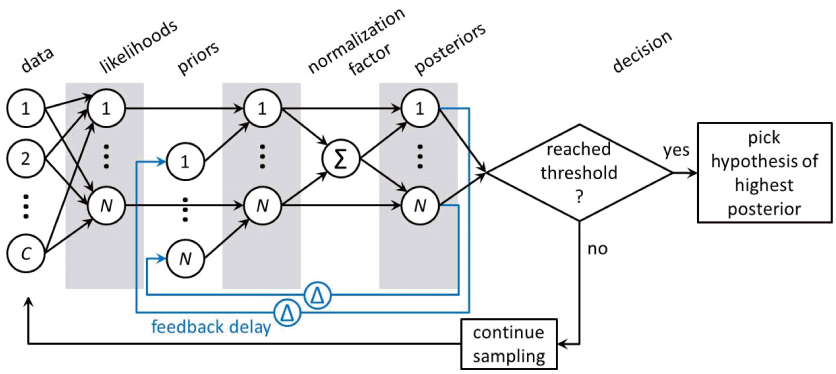

Figure 2: The MSPRT and rMSPRT in schematic form. Circles joined by arrows are the Bayes rule. All C evidence streams (data) are used to compute every one of the N likelihood functions. The product of the likelihood and prior probability of every hypothesis is normalized by the sum ($\sum$) of all products of likelihoods and priors, to produce the posterior probability of that hypothesis. All posteriors are then compared to a constant threshold. A decision is made every time with two possible outcomes: if a posterior reached the threshold, the associated hypothesis is picked, otherwise, sampling from the evidence streams continues. The MSPRT as in [5, 8] only requires what is shown in black. The general recursive MSPRT introduced here re-uses the posteriors Δ time steps in the past for present inference; hence the rMSPRT is shown in black and blue. If we are to work with the negative-logarithm of the Bayes rule and its computations —as we do in this article— all relations between computations are preserved, but products of computations become additions of their logarithms and the divisions for normalization become their negative logarithm. Eq. 9 shows this for the rMSPRT. The rMSPRT itself is formalized by Eq. 10.

During the dot motion task, neurons in the middle-temporal visual area (MT) respond more vigorously to visual stimuli moving in their "preferred" direction than in the opposite "null" direction [10]. Both the mean (Fig. 1d) and variance (Fig. 1e) of their response are proportional to the coherence of the motion. MT responses are assumed to be the uncertain evidence upon which a choice is made in the random dots task [39, 46].

2.1 Recursive MSPRT

Here we schematically review the MSPRT and introduce our recursive generalization, the rMSPRT (Fig. 2), giving full mathematical definitions and deductions in the Methods. The MSPRT decides which of N parallel, competing alternatives (or hypotheses) is the best choice, based on C sequentially sampled streams of evidence (or data). For modelling the dot-motion task, we have $N = 2$ or $N = 4$ hypotheses —the possible saccades to available targets (Fig. 1b,c)— and the C uncertain evidence streams are assumed to be simultaneous spike trains produced by visual-motion-sensitive MT neurons [13, 39, 46].

Every time new evidence arrives, the MSPRT refreshes 'on-line' the likelihood of each hypothesis: the plausibility of the combined evidence streams assuming that hypothesis is true. The likelihood is then multiplied by the probability of that hypothesis based on past experience (the prior). This product for every hypothesis is then normalized by the sum of the products from all N hypotheses; this normalisation is crucial for decision, as it provides the competition between hypotheses. The result is the probability of each hypothesis given current evidence (the posterior) —a *decision variable* per hypothesis. Finally, posteriors are compared to a threshold, and a decision is made to either choose the hypothesis whose posterior probability crosses the threshold, or to continue sampling the evidence streams. Crucially, the (r)MSPRT allows us to use the same algorithm irrespective of the number of alternatives, and thus directly compare the $N = 2$ and $N = 4$ dot-motion task variants.

The MSPRT is a feedforward algorithm, and cannot account for the ubiquitous presence of feedback in neural circuits. To solve this, we introduce the recursive MSPRT (Fig. 2), which uses a generalized, recursive form of the Bayes rule we deduced from first principles (Eq. 5). The MSPRT is a special case of the rMSPRT (in its general form in Eqs. 5 and 10) when priors do not change or, equivalently, for an infinite recursion delay; this is, $\Delta \to \infty$. Also, the previous recurrent extension of MSPRT [19, 34] is a special case of the rMSPRT when $\Delta = 1$. Hence, the rMSPRT generalizes both in allowing the re-use of posteriors from *any* given time in the past as priors for present inference. This uniquely allows us to map the rMSPRT onto neural circuits containing arbitrary feedback delays, in particular solving the problem of decomposing the decision-making algorithm into

distributed components across multiple brain regions. We show below how this allows us to map the rMSPRT onto the cortical-basal-ganglia-thalamo-cortical loops.

Inference using recursive and non-recursive forms of Bayes' rule gives the same results (*e.g.* see [48]), and so both MSPRT and rMSPRT perform identically. Thus, like MSPRT [5, 8], for $N = 2$ rMSPRT also collapses to the sequential probability ratio test of [49] and is thereby optimal in the sense that it requires the smallest expected number of observations to decide, at any given error rate [50]; in other words, for $N = 2$ rMSPRT is also the ideal (Bayesian, normative) observer.

2.2 Upper bounds of decision time predicted by the (r)MSPRT

We first use a particular instance of rMSPRT (Eqs. 9 and 10) to determine bounds on the decision time in the dot motion task. We can then ask how well monkeys approximate such bounds. This instance uses a minimal amount of sensory information, by assuming as many evidence streams as alternatives; that is, $C = N$. Thus, it gives the upper bound on optimal expected decision times (exact for $N = 2$ alternatives, approximate for $N = 4$) per *condition* (given combination of coherence and N); assuming $C > N$ would predict shorter decision times [13].

Following [13], we assume that during the random dot motion task (Fig. 1a-c), the evidence for every possible saccade comes as simultaneous sequences of inter-spike intervals (ISI) produced in MT. In our simulations, evidence sequences are randomly drawn from ISI distributions extracted from the responses of MT neurons facing either their preferred or null directions (Fig. 1f). By repeating the simulations for thousands of trials per condition, we can compare model and monkey performance.

Using these data-determined MT statistics, the (r)MSPRT predicts that the mean decision time on the dot motion task is a decreasing function of coherence (Fig. 3a). For comparison with monkey reaction times, the model's reaction times are the sum of its decision times and estimated non-decision time, encompassing sensory delays and motor execution. For macaques 200–300 ms of non-decision time is a plausible range [20, 45]. Within this range, monkeys tend not to reach the predicted upper bound of reaction time.

The (r)MSPRT framework suggests that decision times depend on the discrimination information in the evidence [13]. Discrimination information here is measured as the divergence between the ISI distributions of MT neurons for dots moving in their preferred and null directions. Intuitively, the larger this divergence, the easier and hence faster the decision. We can estimate how much discrimination information monkeys used by asking how much the (r)MSPRT would require to obtain the same reaction times on correct trials as the monkeys, per condition. We thus find that monkeys tended to use less discrimination information than that in ISI distributions in their MT, when making the decision (Fig. 3b; Methods). In contrast, the (r)MSPRT uses the full discrimination information available. This implies that the decision-making mechanism in monkeys lost large proportions of MT discrimination information (Fig. 3c). Since the (r)MSPRT decision times are upper bounds, this in turn means that this loss is the minimum.

2.3 (r)MSPRT with depleted information reproduces monkey performance

To verify if this information loss alone could account for the monkeys' deviation from the (r)MSPRT upper bounds, we depleted the discrimination information of its input distributions to exactly match the estimated monkey loss in Fig. 3c per condition. We did so only by modifying the mean and standard deviation of the null direction ISI distribution, to make it more similar to the preferred distribution (exemplified in Fig. 3d; see Methods).

Using these information-depleted statistics, the mean reaction times predicted by the (r)MSPRT in correct trials closely match those of monkeys (Fig. 4a). Strikingly, although this information-depletion procedure is based only on data from correct trials, the (r)MSPRT now also matches closely the mean reaction times of monkeys from error trials (Fig. 4b). Moreover, for both correct and error trials the (r)MSPRT accurately captures the relative scaling of mean reaction time by the number of alternatives (Fig. 4a,b).

The reaction time distributions of the model closely resemble those of monkeys in that they are positively skewed and exhibit shorter right tails for higher coherence levels (Fig. 4c-f). These qualitative features are captured across both correct and error trials, and 2 and 4-alternative tasks. Together, these results support the hypothesis that the primate brain approximates an algorithm similar to the rMSPRT.

2.4 rMSPRT maps onto cortico-basal ganglia circuitry

Beyond matching behaviour, we then asked whether the rMSPRT could account for the simultaneously recorded neural dynamics during decision making. To do so, we first must map its components to a neural circuit. Being able to handle arbitrary delays means the rMSPRT could in principle map to a range of feedback neural circuits.

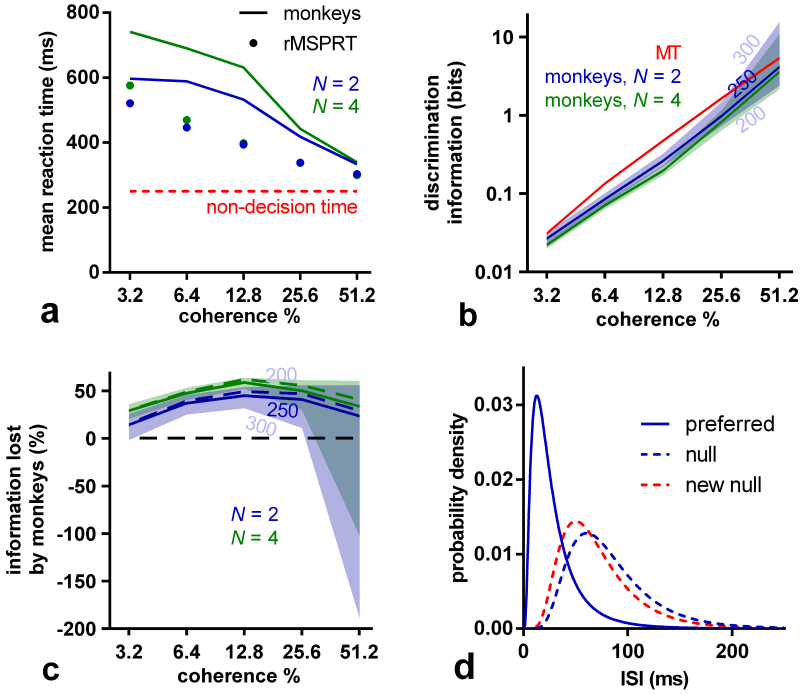

Figure 3: (r)MSPRT predicts information loss during decision making. (a) Comparison of mean reaction time of monkeys for 2 and 4 alternatives (lines) with that predicted by (r)MSPRT (markers), both for correct trials. Red line: assumed 250 ms of non-decision time. Simulation values are means over 100 Monte Carlo experiments each comprising 3200, 4800 total trials for $N = 2, 4$, correspondingly, under the parameter set Ω extracted from MT recordings. (b) Discrimination information per ISI in MT statistics (red) compared to the (r)MSPRT's predictions of the discrimination information available to the monkeys. Central lines are for a non-decision time of 250 ms; the edges of the correspondingly coloured shaded regions are for non-decision times of 300 and 200 ms. (c) As per panel (b), but expressed as a percentage of information lost by monkeys with respect to the information available in MT for the three assumed non-decision times (solid lines and shadings). The information lost if the reaction time match is perfected (see Supplementary note S1.1) is shown as dashed lines (assuming 250 ms of non-decision time). (d) Example ISI density functions before (blue) and after (solid blue and dashed red) information depletion; $N = 2$, 51.2 % coherence and 250 ms of non-decision time. The null distribution was adjusted to become the 'new null' by changing its mean and standard deviation to make it more similar to the preferred distribution. After adjustment, the discrimination information between the preferred and 'new null' distributions matches that estimated from the monkeys performance (panel b). See Methods.

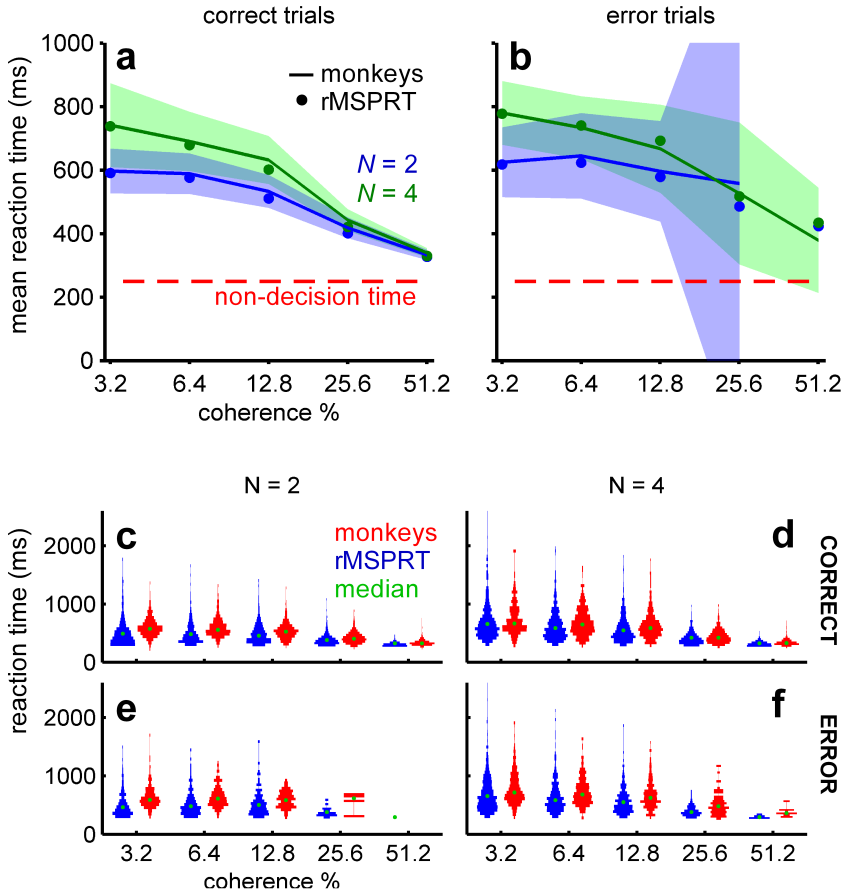

Figure 4: Monkey reaction times are consistent with (r)MSPRT using depleted discrimination information. (a, b) Mean reaction time of monkeys (lines) with 99 % Chebyshev confidence intervals (shading) and (r)MSPRT predictions for correct (a; Eq. 11) and error trials (b; Eq. 12) when using information-depleted statistics (MT parameter set Ω_d). (r)MSPRT results are means of 100 simulations with 3200, 4800 total trials each for $N = 2, 4$, respectively. Confidence intervals become larger in error trials because monkeys made fewer mistakes for higher coherence levels. (c-f) 'Violin' plots of reaction time distributions (vertically plotted histograms reflected about the y-axis) from monkeys (red; 766–785, 1170–1217 trials for $N = 2, 4$, respectively) and (r)MSPRT when using information-depleted statistics (blue; single example Monte Carlo simulation with 800, 1200 total trials for $N = 2, 4$).

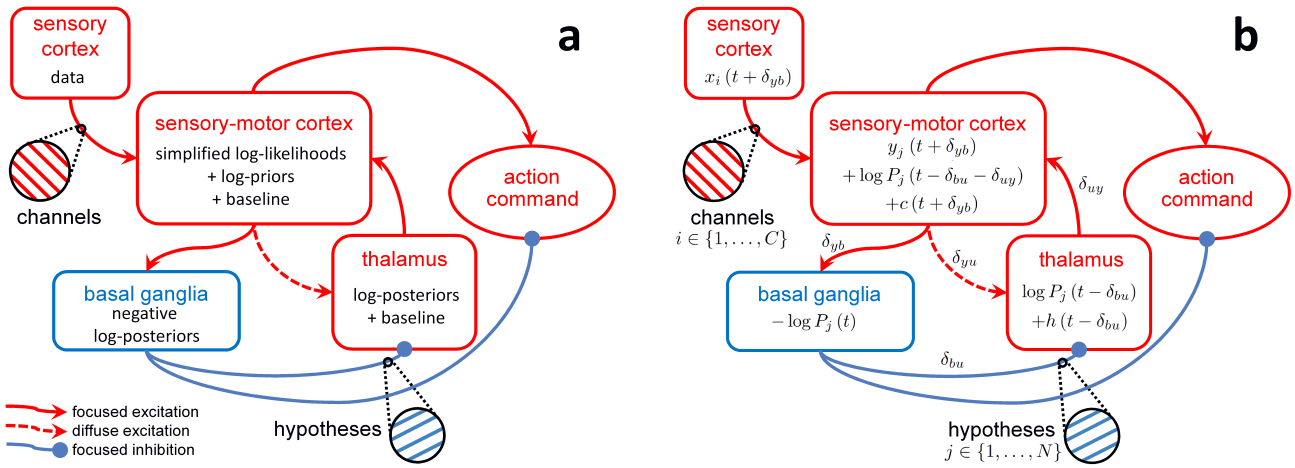

Figure 5: Mapping of rMSPRT computations to the cortico-basal-ganglia-thalamo-cortical loops. (a) Mapping of the negative logarithm of rMSPRT components from Fig. 2. Sensory cortex (*e.g.* MT) produces fresh evidence for the decision, delivered to sensory-motor cortex in C parallel channels (*e.g.* MT spike trains). Sensory-motor cortex (*e.g.* LIP or FEF) computes in parallel the simplified log-likelihoods of all hypotheses given this evidence and adds log-priors —or fed-back log-posteriors after the delay Δ has elapsed. It also adds a hypothesis-independent baseline comprising a simulated constant background activity (*e.g.* from LIP before stimulus onset) and a time increasing term from the interaction with the thalamus. The basal ganglia bring the computations of all hypotheses together into new negative log-posteriors that are then tested against a threshold. Negative log-posteriors will tend to decrease for the best-supported hypothesis, and increase otherwise. This is consistent with the idea that basal ganglia output selectively removes inhibition from a chosen motor program while increasing inhibition of competing programs [4, 8, 32, 44]. Further details of this computation are in Supplementary Fig. S1. Finally, the thalamus conveys the updated log-posterior from basal ganglia output to be used as a log-prior by sensory-motor cortex. Thalamus' baseline is given by its diffuse feedback from sensory-motor cortex. (b) Corresponding formal mapping of rMSPRT's computational components, showing how Eq. 9 decomposes. All computations are delayed with respect to the basal ganglia via the integer latencies δ_{pq}, from p to q; where $p, q \in \{y, b, u\}$, y stands for the sensory-motor cortex, b for the basal ganglia and u for the thalamus. $\Delta = \delta_{yb} + \delta_{bu} + \delta_{uy}$ with the requirement $\Delta \geq 1$.

Because cortex [14, 16, 27, 31, 46], basal ganglia [15, 44] and thalamus [51] have been implicated in decision-making, we sought a mapping that could account for their collective involvement.

In the visuo-motor system, MT projects to the lateral intra-parietal area (LIP) and frontal eye fields (FEF) —the 'sensory-motor cortex'. The basal ganglia receives topographically organized afferent projections [28] from virtually the whole cortex, including LIP and FEF [25, 43, 47]. In turn, the basal ganglia provide indirect feedback to the cortex through thalamus [3, 41]. This arrangement motivated the feedback embodied in rMSPRT.

Multiple parallel recurrent loops connecting cortex, basal ganglia and thalamus can be traced anatomically [3, 41]. Each loop in turn can be sub-divided into topographically organised parallel loops [2, 41]. Based on this, we conjecture the organization of these circuits into N functional loops, for decision formation, to simultaneously evaluate the possible hypotheses (see [34]).

Our mapping of computations within the rMSPRT to the cortico-basal-ganglia-thalamo-cortical loop is shown in Fig. 5. Its key ideas are, first, that areas like LIP or FEF in the cortex evaluate the plausibility of all available alternatives in parallel, based on the evidence produced by MT, and join this to any initial bias. Second, that as these signals traverse the basal ganglia, they compete, resulting in a decision variable per alternative. Third, that the basal ganglia output nuclei uses these to assess whether to make a final choice and what alternative to pick. Fourth, that decision variables are returned to LIP or FEF via thalamus, to become a fresh bias carrying all conclusions on the decision so far. In addition, our mapping of the rMSPRT provides an account for the spatially diffuse cortico-thalamic projection [40]; it predicts the projection conveys an hypothesis-independent signal that does not affect the inference carried out by the overall loop (see Supplementary note S1.2).

2.5 Electrophysiological comparison

With the mapping above, we could compare the proposed rMSPRT computations to recorded activity during decision-making in area LIP and striatum. We first consider the dynamics around decision initiation. During the dot motion task, the mean firing rate of LIP neurons deviates from baseline into a stereotypical dip soon after stimulus onset, possibly indicating the reset of a neural integrator [22, 46]. LIP responses become choice- and coherence-modulated after the dip [46]. We therefore reasoned that LIP neurons engage in decision formation from the bottom of the dip and model their mean firing rate from then on. After this, mean firing rates "ramp-up" for $\sim$ 40 ms, then "fork": they continue ramping-up if dots moved towards the response (or movement) field of the neuron (inRF trials; Fig. 6a, solid lines) or decrease (perhaps plateau) if the dots were moving away from its response field (outRF trials; dashed lines) [14, 46]. The magnitude of LIP firing rate is also proportional to the number of available alternatives (Fig. 6a,b) [14].

The model LIP in rMSPRT (sensory-motor cortex) captures each of these properties: activity ramps from the start of the integrator, forks between putative in- and out-RF responses, and scales with the number of alternatives (Fig. 6c). Under this model, inRF responses in LIP occur when the likelihood function represented by neurons was best matched by the uncertain MT evidence; correspondingly, outRF responses occur when the likelihood function was not well matched by the evidence.

The rMSPRT provides a mechanistic explanation for the ramp-and-fork pattern. Initial accumulation (steps 0–2) occurs before the feedback has arrived at the model sensory-motor cortex, resulting in a ramp. The forking (at step 3) is the point at which the posteriors from the output of the model basal ganglia first arrive at sensory-motor cortex to be re-used as priors. By contrast, non-recursive MSPRT (without feedback of posteriors) predicts well-separated neural signals throughout (Fig. 6e). Consequently, the rMSPRT suggests that the fork represents the time at which updated signals representing the competition between alternatives —here posterior probabilities— are made available to the sensory-motor cortex.

The rMSPRT further predicts that the scaling of activity in sensory-motor cortex by the number of alternatives is due to cortico-subcortical loops becoming organized as N parallel functional circuits, one per hypothesis. This would determine the baseline output of the basal ganglia. Until task related signals reach the model basal ganglia, their output codes the initial priors for the set of N hypotheses (see Methods). Their output is then an increasing function of the number of alternatives (Fig. 6f). This increased inhibition of thalamus in turn reduces baseline cortical activity as a function of N. The direct proportionality of basal ganglia output firing rate to N recorded in the substantia nigra pars reticulata (SNr) of macaques during a fixed-duration choice task [4] lends support to this hypothesis.

The rMSPRT also captures key features of dynamics at decision termination. For inRF trials, the mean firing rate of LIP neurons peaks at or very close to the time of saccade onset (Fig. 6b). By contrast, for outRF trials mean rates appear to fall just before saccade onset. The rMSPRT can capture both these features (Fig. 6d) when we allow the model to continue updating after the decision rule (Eq. 10) is met. The decision rule is implemented at the output of the basal ganglia and the model sensory-motor cortex peaks just before the final posteriors have reached the cortex. The rMSPRT thus predicts that the activity in LIP lags the actual decision.

This prediction may explain an apparent paradox of LIP activity. The peri-saccadic population firing rate peak in LIP during inRF trials (Fig. 6b) is commonly assumed to indicate the crossing of a threshold and thus decision termination. Visuo-motor decisions must be terminated well before saccade to allow for the delay in the execution of the motor command, conventionally assumed in the range of 80–100 ms in macaques [39, 45]. It follows that LIP peaks too close to saccade onset ($\sim$ 15 ms before) for this peak to be causal. The rMSPRT suggests that the inRF LIP peak is not indicating decision termination, but is instead a delayed read-out of termination in an upstream location.

LIP firing rates are also modulated by dot-motion coherence (Fig. 7a,b,i,j). Following stimulus onset, the response of LIP neurons tends to fork more widely for higher coherence levels (Fig. 7a,i) [14, 46]. The increase in activity before a saccade during inRF trials is steeper for higher coherence levels, reflecting the shorter average reaction times (Fig. 7b,j) [14, 46]. The rMSPRT shows coherence modulation of both the forking pattern (Fig. 7c,k) and slope of activity increase (Fig. 7d,l). The apparent non-convergence to a common activity level in the model we take up in the Discussion.

Similar modulation of population firing rates during the dot motion task has been observed in the striatum [15]. Naturally, the striatum in rMSPRT, which relays cortical input (Supplementary Fig. S1), captures this modulation (Supplementary Fig. S2).

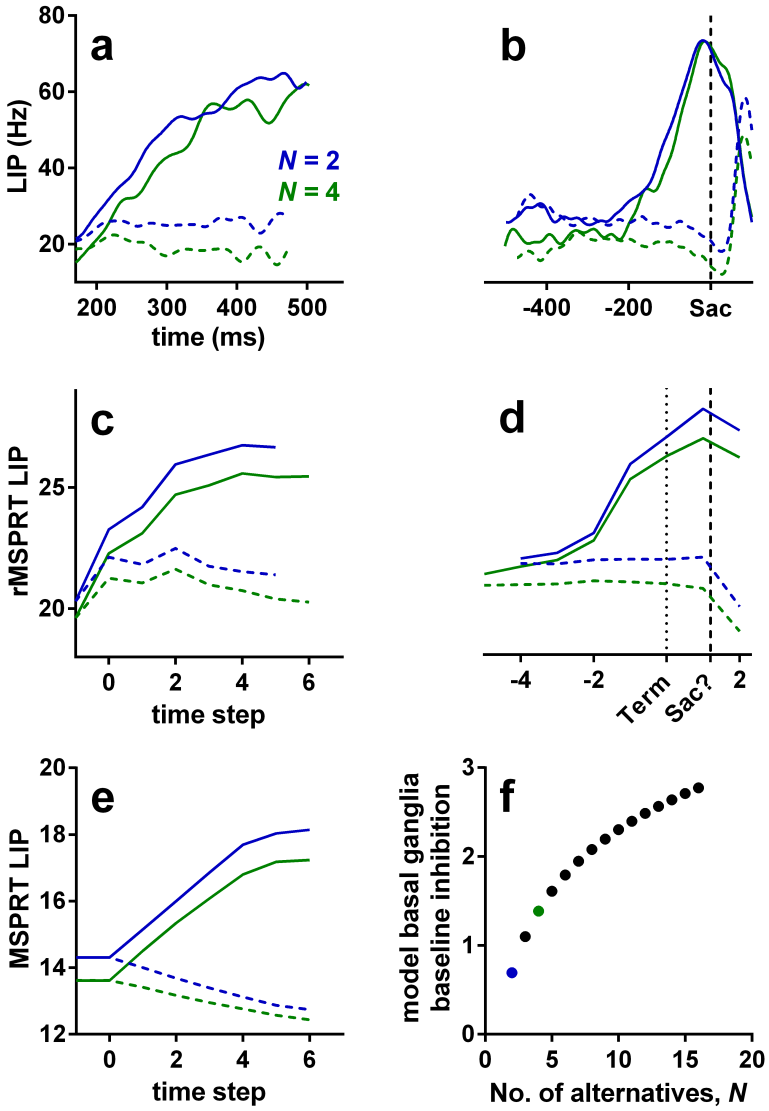

Figure 6: Example LIP firing rate patterns and predictions of rMSPRT and MSPRT at 25.6 % coherence. (a, b) Mean population firing rate of LIP neurons during correct trials on the reaction-time version of the dot motion task. By convention, inRF trials are those when recorded neurons had the motion-cued target inside their response field (solid lines); outRF trials are those when recorded neurons had the motion-cued target outside their response field (dashed lines). (a) Aligned at stimulus onset, starting at the stereotypical dip, illustrating the "ramp-and-fork" pattern between average inRF and outRF responses. (b) Aligned at saccade onset (vertical dashed line). (c, d) Mean time course of the model sensory-motor cortex in rMSPRT aligned at decision initiation (c; $t = 1$) and termination (d; Term; dotted line), for correct trials. Initiation and termination are with respect to the time of basal ganglia output. Note the suggested saccade time "Sac?", close to the peak of inRF computations. Simulations are a single Monte Carlo experiment with 800, 1200 total trials for $N = 2, 4$, respectively, using parameter set Ω_d. We include an additional step at -1 determined only by initial priors and baseline, where no inference is carried out ($y_i\left(t + \delta_{yb}\right) = 0$ for all i). Conventions as in (a). (e) Same as in (c), but for the standard, non-recursive MSPRT (defined here as the first case of Eq. 9). (f) Baseline output of the model basal ganglia increases as a function of the number of alternatives, thus increasing the initial inhibition of thalamus and cortex. For uniform priors, the rMSPRT predicts this function is: $-\log P\left(H_i\right) = -\log\left(1/N\right)$. Coloured dots indicate $N = 2$ (blue) and $N = 4$ (green).

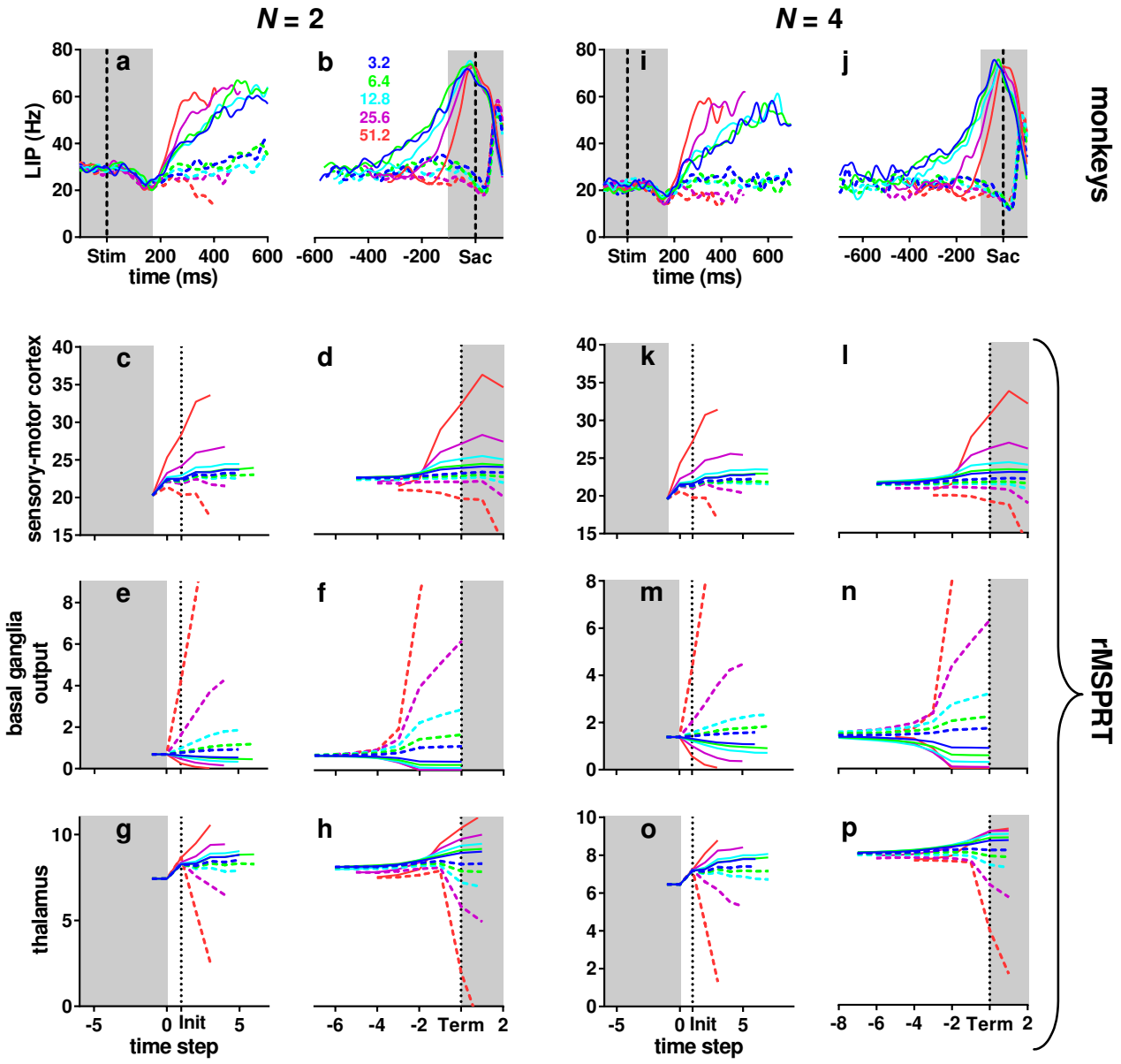

Figure 7: Modulation of activity by coherence throughout the cortico-basal-ganglia-thalamo-cortical loop. (a-h) $N = 2$. (i-p) $N = 4$. Top row: mean firing rate in LIP over time, aligned to stimulus onset (Stim; a, i) or saccade onset (Sac; b, j) (vertical dashed lines). (c-h, k-p) mean rMSPRT computations as mapped in Fig. 5, aligned at decision initiation or termination (Init/Term; dotted lines); single Monte Carlo experiment with 800, 1200 total trials for $N = 2, 4$, respectively. (e, f, m, n) predicted simultaneous course of mean firing rate in SNr. (g, h, o, p) predicted course in thalamic relay nuclei. Solid: inRF. Dashed: outRF. Coherence % as in legend. Unshaded regions indicate approximate periods where a mechanism of decision formation should aim to reproduce the recordings.

2.6 Electrophysiological predictions

Our proposed mapping of the rMSPRT's components (Fig. 5) makes testable qualitative predictions for the mean responses in basal ganglia and thalamus during the dot motion task. For the basal ganglia output, likely from the oculomotor regions of the SNr, rMSPRT (like MSPRT) predicts a drop in the activity of output neurons in inRF trials and an increase in outRF ones. It also predicts that these changes are more pronounced for higher coherence levels (Fig. 7e,f,m,n). These predictions are consistent with recordings from macaque SNr neurons showing that they suppress their activity during visually or memory guided saccade tasks, in putative support of saccades towards a preferred region of the visual field [4, 26], and enhance it otherwise [26].

For visuo-motor thalamus, rMSPRT predicts that the time course of the mean firing rate will exhibit a ramp-and-fork pattern similar to that in LIP (Fig. 7g,h,o,p). The separation of in- and out-RF activity is consistent with the results of [51] who found that, during a memory-guided saccade task, neurons in the macaque medio-dorsal nucleus of the thalamus (interconnected with LIP and FEF), responded more vigorously when the target was flashed within their response field than when it was flashed in the opposite location.

2.7 Predictions for neural activity features not crucial for inference

Understanding how a neural system implements an algorithm is complicated by the need to identify which features are core to executing the algorithm, and which are imposed by the constraints of implementing computation using neural elements —for example, that neurons cannot have negative firing rates, so cannot straightforwardly represent negative numbers. The three free parameters in the rMSPRT allow us to propose which functional and anatomical properties of the cortico-basal-ganglia-thalamo-cortical loop are workarounds within these constraints, but do not affect inference.

One free parameter enforces the baseline activity that LIP neurons maintain before and during the initial stimulus presentation (Fig. 7a,i). Varying this parameter, l, scales the overall activity of LIP, but does not change the inference performed (Fig. 8a). Consequently, other than the dependence on N, this suggests that the inference algorithm in the brain is insensitive to the baseline activity of LIP.

The second free parameter, w_{yt}, sets the strength of the spatially diffuse projection from cortex to thalamus. Varying this weight changes the forking between inRF and outRF computations but does not affect inference (Fig. 8b). The third free parameter, n, sets the scale at which input ISIs are processed (Methods). Changing it varies the slope of sensory-motor computations (Fig. 8c) without affecting inference, even allowing all-decreasing predicted mean firing rates. By definition, the log-likelihood of a sequence tends to be negative and decrease monotonically as the sequence lengthens. Hence, n may capture a workaround of the decision-making circuitry to represent positive simplified log-likelihoods and avoid signal 'underflow', by means of scaling the input data.

Previous studies have reported that the cortical outRF response in the dot-motion task can either increase [16], plateau [14, 46] (see Fig. 7a,i) or decrease [31] after the fork; similar patterns of plateau or decrease have been described for cortical signals representing the non-favoured hypothesis in other evidence accumulation tasks [27, 29, 33]. Consequently, the rMSPRT predicts that the precise slope of responses to favoured and non-favoured hypotheses is unrelated to the inference procedure.

3 Discussion

We sought a principled model for decision making that could provide a candidate theory for the mammalian brain's decision-making algorithm while at the same time respecting the constraints of a neural implementation, including feedback. We find it remarkable that, starting from data-constrained spike-trains [13], a monolithic statistical test in the form of the rMSPRT can simultaneously account for much of the anatomy, behaviour and electrophysiology of decision-making.

Here we have pushed the analogy between a single-equation statistical test and the neural decision-making algorithm a long way before it broke down. While it is not plausible that the brain implements exactly a specific algorithm, our results suggest the rMSPRT captures some essential elements of the mammalian brain's decision-making mechanism. These, we propose, are the following. First, that the mechanism is probabilistic in nature —the brain utilizes the uncertainty in neural signals, rather than suffering from it. Second, that the mechanism works entirely 'on-line', continuously updating representations of hypotheses that can be queried at any time to make a decision. Third, that this processing is distributed, recursive, and parallel, producing a decision variable for each available hypothesis. And fourth, that this recursion allows the mechanism to be adaptive to changes in the statistics of the environment, as it can re-use updated probabilities about hypotheses as priors for decision-making.

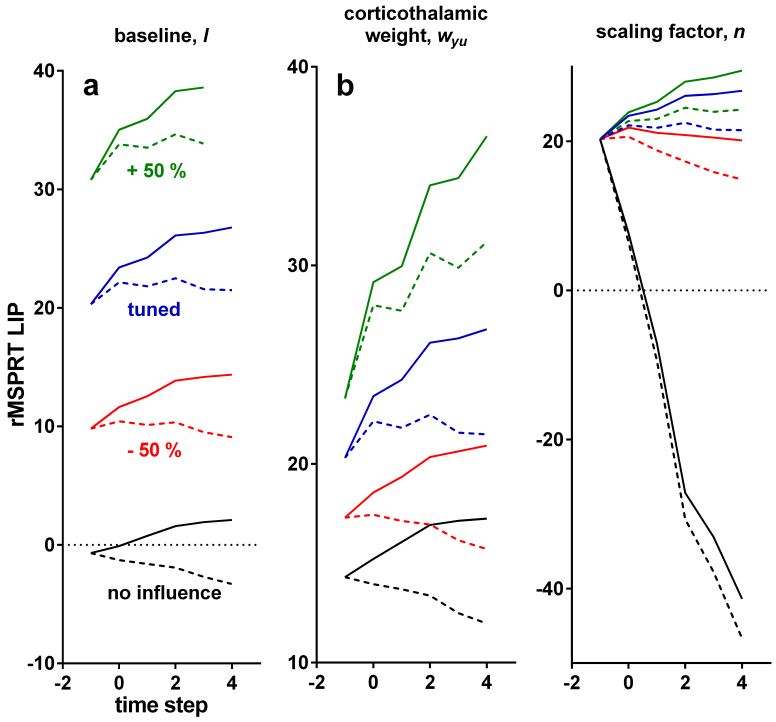

Figure 8: Effect of variations of free parameters on the time course of the model LIP in rMSPRT. Each solid and dashed set of lines is the mean of correct trials in a single Monte Carlo experiment, with 800 total trials, 25.6 % coherence and $N = 2$. Computations aligned at decision initiation. Solid: inRF. Dashed: outRF. Blue: with parameters as tuned for this study (see Methods). Green: increasing parameter value by 50 %, keeping other parameters as tuned. Red: decreasing it by 50 %, keeping others as tuned. Black: removing the effect of the tested parameter ($l = 0$, $w_{yu} = 0$, $n = 1$), keeping others as tuned. (a) Varying the baseline, l. (b) Varying the cortico-thalamic weight, w_{yu}. (c) Varying the data scaling factor n.

3.1 Why implement a recursive algorithm in the brain?

Prior work proposed that the cortex and basal ganglia alone could implement the non-recurrent MSPRT [8, 37]. However, the looped cortico-basal-ganglia-thalamo-cortical architecture implies a recursive computation. This raises the question of why such a complex, distributed feedback architecture exists.

First, recursion makes trial-to-trial adaptation of decisions possible. Priors determined by previous stimulation (fed-back posteriors), can bias upcoming decisions towards the expected best choice, even before any new evidence is collected. This can shorten reaction times in future familiar settings without compromising accuracy.

Second, recursion provides a robust memory. A posterior fed-back as a prior is a sufficient statistic of all past evidence observations. That is, it has taken 'on-board' all sensory information since the decision onset. This means that the sensory-motor cortex in the rMSPRT only need keep track of observations in a moving time window of maximum width Δ —the delay around the loop– rather than keeping track of the entire sequence of observations. For a physical substrate subject to dynamics and leakage, like a neuron in LIP or FEF, this has obvious advantages: it would reduce the demand for keeping a perfect record (*e.g.* likelihood) of all evidence, from the usual hundreds of milliseconds in decision times to the ~ 30 ms of latency around the cortico-basal-ganglia-thalamo-cortical loop (adding up estimates from [23, 30, 42]).

3.2 Lost information and perfect integration

The rMSPRT predicts that monkeys do not make full use of the discrimination information available in MT (Fig. 3b). Here rMSPRT needs only $C = N$ MT spike-trains to outperform monkeys: as this is the upper bound of rMSPRT mean decision time, this implies the monkey's predicted loss (Fig. 3c) is a minimum.

This gap arises because rMSPRT is a generative model of the task, which assumes initial knowledge of the coherence and N that determine appropriate likelihoods for the task at hand. Any deviation from this ideal will tend to degrade performance [7], whether it comes from one or more of the inherent leakiness of neurons, correlations or the condition to likelihood mapping [13]. LIP neurons change their coding during learning of the dot motion task and MT neurons do not [35], implying that learning the task requires mapping of MT to LIP populations by synaptic plasticity [36]. Consequently, even if the MT representation is perfect, the learnt mapping only need satisfice the task requirements, not optimally perform.

Excellent matches to performance in both correct and error trials were obtained solely by accounting for lost information in the evidence streams. In doing so, no noise was added within the rMSPRT itself. Prior experimental work reported perfect, noiseless integration by both rat and human subjects performing an auditory task, attributing all effects of noise on task performance to the variability in the sensory input [12]. Our results extend this observation to primate performance on the dot motion task, and further support the idea that neural decision-making mechanism can perform perfect integration of uncertain evidence.

3.3 Neural response patterns during decision formation

Neurons in LIP, FEF [16] and striatum [15] exhibit a ramp-and-fork pattern during the dot motion task. Analogous choice-modulated patterns have been recorded in the medial premotor cortex of the macaque during a vibro-tactile discrimination task [29] and in the posterior parietal cortex and frontal orienting fields of the rat during an auditory discrimination task [27]. The rMSPRT indicates that such slow dynamics emerge from a delayed inhibitory feedback looped architecture in the decision circuits. Together, this suggests that decision formation in mammals may approximate a common recursive computation.

A random dot stimulus pulse delivered earlier in a trial has a bigger impact on LIP firing rate than a later one [31]. This highlights the importance of capturing the initial, early-evidence ramping-up before the forking. However, multiple models omit it, focusing only on the forking (*e.g.* [6, 18, 39]). Other, heuristic models account for LIP activity from the onset of the choice targets, through dots stimulation and up until saccade onset (*e.g.* [1, 22, 24, 52]). Nevertheless, their predicted firing rates rely on two fitted heuristic signals that shape both the dip and the ramp-and-fork pattern. In contrast, the ramp-and-fork dynamics emerge naturally from the delayed inhibitory feedback in rMSPRT during decision formation.

rMSPRT replicates the ramp-and-fork pattern for individual coherence levels and given N (Fig. 6). However, comparing its predictions across N (Fig. 6d) or over coherence levels (Fig. 7d,l) reveals that the model sensory-motor cortex does not converge to a common value around decision termination in inRF trials, as the LIP does (Fig. 6b, Fig. 7b,j and [14]). Here we have reached the limits where the direct comparison of the single-equation statistical test to neural signals breaks down. Our mapping of rMSPRT onto the cortico-basal-ganglia-thalamo-cortical circuitry suggests the core underlying computation contributed per site during decision formation (Fig.

5), such as the negative log-posterior probability at the basal ganglia. Even if these suggestions are accurate, we would still expect other factors to influence the dynamics of recorded neural signals: these regions after all engage in multiple computations.

One possibility is that the convergence of LIP activity to a common value just prior to saccade onset may result from monotonic transformations of these core computations. For instance, the successful fitting of previous computational models to neural signals has been critically dependent on the addition of heuristic quantities [1, 22, 24, 52]. The incorporation of similar heuristic signals may also convert the present qualitative resemblance of the recordings, by rMSPRT, to a quantitative reproduction. This is, however, beyond the scope of this study, whose aim is to test how closely a normative mechanism can explain behaviour and electrophysiology during decisions.

3.4 Emergent predictions

Inputs to the rMSPRT were determined solely from MT responses during the dot-motion task, and it has only three free parameters, none of which affect inference. It is thus surprising that it renders emergent predictions that are consistent with experimental data. First, our information-depletion procedure used exclusively statistics from correct trials. Yet, after depletion, rMSPRT matches monkey behaviour in correct and error trials (Fig. 4), suggesting a mechanistic connection between them in the monkey that is naturally captured by rMSPRT. Second, the values of the three free parameters were chosen solely so that the model LIP activity resembled the ramp-and-fork pattern observed in our LIP data-set (Fig. 6a,c). As demonstrated in Fig. 8, the ramp-and-fork pattern is a particular case of two-stage patterns that are an intrinsic property of the rMSPRT, guaranteed by the feedback of the posterior after the delay Δ has elapsed (Eq. 5). Nonetheless, the model also matches LIP dynamics when aligned at decision termination (panels b and d). Third, the predictions of the time course of the firing rate in SNr and thalamic nuclei naturally emerge from the functional mapping of the procedure onto the cortico-basal-ganglia-thalamo-cortical circuitry. These are already congruent with existing electrophysiological data; however, their full verification awaits recordings from these sites during the dot motion task. These and other emergent predictions are an encouraging indicator of the explanatory power of a family of exact inference algorithms represented by the rMSPRT.

4 Methods

4.1 Experimental paradigms

Behavioural and neural data was collected in two previous studies [10, 14], during two versions of the random dots task (Fig. 1a-c). Detailed experimental protocols can be found in such studies. Below we briefly summarize them.

4.1.1 Fixed duration

Three rhesus macaques (*Macaca mulatta*) were trained to initially fixate their gaze on a visual fixation point (cross in Fig. 1a). A random dot kinematogram appeared covering the response field of the MT neuron being recorded (grey patch); task difficulty was controlled per trial by the proportion of dots (coherence %) that moved in one of two directions: that to which the MT neuron was tuned to —its preferred motion direction— or its opposite —null motion direction. After 2 s the fixation point and kinematogram vanished and two targets appeared in the possible motion directions. Monkeys received a liquid reward if they then saccaded to the target towards which the dots in the stimulus were predominantly moving [10].

4.1.2 Reaction time

Two macaques learned to fixate their gaze on a central fixation point (Fig. 1b,c). Two (Fig. 1b) or four (Fig. 1c) eccentric targets appeared, signalling the number of alternatives in the trial, N; one such target fell within the response (movement) field of the recorded LIP neuron (grey patch). This is the region of the visual field towards which the neuron would best support a saccade. Later a random dot kinematogram appeared where a controlled proportion of dots moved towards one of the targets. The monkeys received a liquid reward for saccading to the indicated target when ready [14].

4.2 Data analysis

For comparability across databases, we only analysed data from trials with coherence levels of 3.2, 6.4, 12.8, 25.6, and 51.2 %, unless otherwise stated. We used data from all neurons recorded in such trials. This is, between 189 and 213 visual-motion-sensitive MT neurons (see table 1; data from [10, 11]) and 19 LIP neurons whose activity was previously determined to be choice- and coherence-modulated (data from [14]). The behavioural data analysed was that associated to LIP recordings. For MT, we analysed the neural activity between the onset and the vanishing of the stimulus. For LIP we focused on the period between 100 ms before stimulus onset and 100 ms after saccade onset.

To estimate moving statistics of neural activity we first computed the spike count over a 20 ms window sliding every 1 ms, per trial. The moving mean firing rate per neuron per condition was then the mean spike count over the valid bins of all trials divided by the width of this window; the standard deviation was estimated analogously. LIP recordings were either aligned at the onset of the stimulus or of the saccade; after or before these (respectively), data was only valid for a period equal to the reaction time per trial. The population moving mean firing rate is the mean of single-neuron moving means over valid bins; analogously, the population moving variance of the firing rate is the mean of single neuron moving variances. For clarity, population statistics were then smoothed by convolving them with a Gaussian kernel with a 10 ms standard deviation. The resulting smoothed population moving statistics for MT are in Fig. 1d,e. In Figs. 6a,b and 7a,b,i,j smoothed mean LIP firing rates are plotted only up to the median reaction time plus 80 ms, per condition.

Analogous procedures were used to compute the moving mean of rMSPRT computations, per time step. These are shown up to the median of termination observations plus 3 time steps in Figs. 6c-e and 7c-h,k-p.

4.3 Definition of rMSPRT

Let $\mathbf{x}(t) = (x_1(t), \ldots, x_C(t))$ be a vector random variable composed of observations, $x_j(t)$, made in C channels at time $t \in \{1, 2, \ldots\}$. Let also $\mathbf{x}(r:t) = (\mathbf{x}(r)/n, \ldots, \mathbf{x}(t)/n)$ be the sequence of i.i.d. vectors $\mathbf{x}(t)/n$ from r to t ($r < t$). Here $n \in \{\mathbb{R} > 0\}$ is a constant data scaling factor. If $n > 1$, it scales down incoming data, $x_j(t)$. Note that this is only effective from the likelihood on and does not affect the format or time interpretation of the data. This is key to reveal that the dynamics in rMSPRT computations match those of cortical recordings (Fig. 8c). Crucially, since n is hypothesis-independent, it does not affect inference.

Assume there are $N \in \{2, 3, \ldots\}$ alternatives or hypotheses about the uncertain *evidence*, $\mathbf{x}(1:t)$ —say possible courses of action or perceptual interpretations of sensory data. The task of a decision maker is to determine which hypothesis H_i ($i \in \{1, \ldots, N\}$) is best supported by this evidence as soon as possible, for a given level of accuracy. To do this, it requires to estimate the posterior probability of each hypothesis given the data, $P(H_i|\mathbf{x}(1:t))$, as formalized by Bayes rule. The mechanism we seek must be recursive to match the nature of the brain circuitry. This implies that it can use the outcome of previous choices to inform a present one, thus becoming adaptive and engaged in ongoing learning. Formally, $P(H_i|\mathbf{x}(1:t))$ will be initially computed upon starting priors $P(H_i)$ and likelihoods $P(\mathbf{x}(1:t)|H_i)$; however, after some time $\Delta \in \{1, 2, \ldots\}$, it will re-use past posteriors, $P(H_i|\mathbf{x}(1:t-\Delta))$, Δ time steps ago, as priors, along with the likelihood function $P(\mathbf{x}(t-\Delta+1:t)|H_i)$ of the segment of $\mathbf{x}(1:t)$ not yet accounted by $P(H_i|\mathbf{x}(1:t-\Delta))$. A mathematical induction proof of this form of Bayes rule follows.

If say $\Delta = 2$, in the first time step, $t = 1$:

$$P(H_i|\mathbf{x}(1)) = \frac{P(\mathbf{x}(1)|H_i)\,P(H_i)}{P(\mathbf{x}(1))} \tag{1}$$

By $t = 2$:

$$P(H_i|\mathbf{x}(2), \mathbf{x}(1)) = \frac{P(\mathbf{x}(2), \mathbf{x}(1)|H_i)\,P(H_i)}{P(\mathbf{x}(2), \mathbf{x}(1))}$$

Note that we are still using the initial fixed priors $P(H_i)$. Now, for $t = 3$:

$$P(H_i|\mathbf{x}(3), \mathbf{x}(2), \mathbf{x}(1)) = \frac{P(\mathbf{x}(3), \mathbf{x}(2), \mathbf{x}(1)|H_i)\,P(H_i)}{P(\mathbf{x}(3), \mathbf{x}(2), \mathbf{x}(1))} \tag{2}$$

According to the product rule, we can segment the probability of the sequence $\mathbf{x}(1:t)$ as:

$$P(\mathbf{x}(1:t)) = P(\mathbf{x}(t-\Delta+1:t), \mathbf{x}(1:t-\Delta)) = P(\mathbf{x}(t-\Delta+1:t)|\mathbf{x}(1:t-\Delta))\,P(\mathbf{x}(1:t-\Delta)) \tag{3}$$

And, since $\mathbf{x}(t)$ are i.i.d., the likelihood of the two segments is:

$$P\left(\mathbf{x}\left(1:t\right)|H_i\right) = P\left(\mathbf{x}\left(t-\Delta+1:t\right)|H_i\right) P\left(\mathbf{x}\left(1:t-\Delta\right)|H_i\right) \tag{4}$$

If we substitute the likelihood in Eq. 2 by Eq. 4, its normalization constant by Eq. 3 and re-group, we get:

$$P\left(H_i|\mathbf{x}(3),\mathbf{x}(2),\mathbf{x}(1)\right) = \left(\frac{P\left(\mathbf{x}(3),\mathbf{x}(2)|H_i\right)}{P\left(\mathbf{x}(3),\mathbf{x}(2)|\mathbf{x}(1)\right)}\right)\left(\frac{P\left(\mathbf{x}(1)|H_i\right)P\left(H_i\right)}{P\left(\mathbf{x}(1)\right)}\right)$$

It is evident that the right factor is $P\left(H_i|\mathbf{x}(1)\right)$ as in Eq. 1. Hence, in this example, by $t=3$ we start using past posteriors as priors for present inference as:

$$P\left(H_i|\mathbf{x}(3),\mathbf{x}(2),\mathbf{x}(1)\right) = \frac{P\left(\mathbf{x}(3),\mathbf{x}(2)|H_i\right)P\left(H_i|\mathbf{x}(1)\right)}{P\left(\mathbf{x}(3),\mathbf{x}(2)|\mathbf{x}(1)\right)}$$

So, in general:

$$P\left(H_i|\mathbf{x}\left(1:t\right)\right) = \begin{cases} \dfrac{P\left(\mathbf{x}\left(1:t\right)|H_i\right)P\left(H_i\right)}{P\left(\mathbf{x}\left(1:t\right)\right)} & \text{for } t \leq \Delta \\ \dfrac{P\left(\mathbf{x}\left(t-\Delta+1:t\right)|H_i\right)P\left(H_i|\mathbf{x}\left(1:t-\Delta\right)\right)}{P\left(\mathbf{x}\left(t-\Delta+1:t\right)|\mathbf{x}\left(1:t-\Delta\right)\right)} & \text{for } t > \Delta \end{cases} \tag{5}$$

where the normalization constants are

$$P\left(\mathbf{x}\left(1:t\right)\right) = \sum_{j=1}^{N} P\left(\mathbf{x}\left(1:t\right)|H_j\right)P\left(H_j\right)$$
$$P\left(\mathbf{x}\left(t-\Delta+1:t\right)|\mathbf{x}\left(1:t-\Delta\right)\right) = \sum_{j=1}^{N} P\left(\mathbf{x}\left(t-\Delta+1:t\right)|H_j\right)P\left(H_j|\mathbf{x}\left(1:t-\Delta\right)\right)$$

We stress that by $t > \Delta$, Eq. 5 starts making use of a past posterior as prior.

It is apparent that the critical computations in Eq. 5 are the likelihood functions. The forms that we consider ahead are based on the simplest shown by [13], where the number of evidence streams equals the number of hypotheses ($C = N$); however, as discussed by them, more complex ($C > N$), biologically-plausible ones can be formulated if necessary. Although not required, to simplify the notation when $C = N$, data in the channel conveying the most salient evidence for hypothesis H_i will bear its same index i, as $x_i\left(j\right)$ (see [13]). When $t \leq \Delta$ we have:

$$P\left(\mathbf{x}\left(1:t\right)|H_i\right) = a\left(t\right)\prod_{j=1}^{t}\frac{f_*\left(x_i\left(j\right)/n\right)}{f_0\left(x_i\left(j\right)/n\right)} \tag{6}$$

this is, the likelihood that $x_i\left(j\right)/n$ was drawn from a distribution, f_*, rather than from f_0, that is assumed to have originated $x_k\left(j\right)/n$ $(k \neq i)$ for the rest of the channels. In Eq. 6, $a\left(t\right) = \prod_{m=1}^{t}\prod_{k=1}^{N} f_0\left(x_k\left(m\right)/n\right)$ is a hypothesis-independent factor that does not affect Eq. 5 and thus needs not to be considered further.

When $t > \Delta$ only the observations in the time window $[t-\Delta+1, t]$ are used for the likelihood function because data before this window is already considered within the fed-back posterior, $P\left(H_i|\mathbf{x}\left(1:t-\Delta\right)\right)$. Then, the likelihood function is:

$$P\left(\mathbf{x}\left(t-\Delta+1:t\right)|H_i\right) = d\left(t\right)\prod_{j=t-\Delta+1}^{t}\frac{f_*\left(x_i\left(j\right)/n\right)}{f_0\left(x_i\left(j\right)/n\right)} \tag{7}$$

where again $d\left(t\right) = \prod_{m=t-\Delta+1}^{t}\prod_{k=1}^{N} f_0\left(x_k\left(m\right)/n\right)$ need not to be considered further.

Now, for our likelihood functions to work upon a statistical structure like that produced by neurons in MT we need to be more specific. [13] showed that ISIs in MT during the random dot motion task are best described as lognormally distributed and we assume that decisions are made upon the information conveyed by them. Thus, from now on we assume that f_* and f_0 are lognormal and that they are specified by means μ_* and μ_0 and standard deviations σ_* and σ_0, respectively. We can then re-write Eqs. 6 and 7 as the log-likelihood function, $y_i\left(t\right)$, substituting the lognormal-based form of it reported by [13]:

$$y_i\left(t\right) = \begin{cases} g_0\Delta + g_1\sum_{j=1}^{t}\left[\log\left(x_i\left(j\right)/n\right)\right]^2 + g_2\sum_{j=1}^{t}\log\left(x_i\left(j\right)/n\right) & \text{for } t \leq \Delta \\ g_0\Delta + g_1\sum_{j=t-\Delta+1}^{t}\left[\log\left(x_i\left(j\right)/n\right)\right]^2 + g_2\sum_{j=t-\Delta+1}^{t}\log\left(x_i\left(j\right)/n\right) & \text{for } t > \Delta \end{cases} \tag{8}$$

with

$$g_0 = \frac{\kappa_0^2}{2\Theta_0^2} - \frac{\kappa_*^2}{2\Theta_*^2} + \log\left(\frac{\Theta_0}{\Theta_*}\right)$$
$$g_1 = \frac{1}{2\Theta_0^2} - \frac{1}{2\Theta_*^2}$$
$$g_2 = \frac{\kappa_*}{\Theta_*^2} - \frac{\kappa_0}{\Theta_0^2}$$

where $\kappa = \log\left(\mu^2/\sqrt{\sigma^2+\mu^2}\right)$ and $\Theta^2 = \log\left(\sigma^2/\mu^2+1\right)$ with appropriate subindices $_*$, $_0$.

The terms $g_0\Delta$ in Eq. 8 are hypothesis-independent, can be absorbed into $a(t)$ and $d(t)$, correspondingly, and thus will not be considered further. As a result of this, the $y_i(t)$ used from now on is a "simplified" version of the log-likelihood.

We now take the logarithm of Eq. 5, define $-\log P_i(t) \equiv -\log P(H_i|\mathbf{x}(1:t))$ and substitute the simplified log-likelihood from Eq. 8 in the result, giving:

$$-\log P_i(t) = \begin{cases} -y_i(t) - \log P(H_i) - c(t) + \log\sum_{j=1}^{N}\exp\left(y_j(t) + \log P(H_j) + c(t)\right) & \text{for } t \le \Delta \\ -y_i(t) - \log P_i(t-\Delta) - c(t) + \log\sum_{j=1}^{N}\exp\left(y_j(t) + \log P_j(t-\Delta) + c(t)\right) & \text{for } t > \Delta \end{cases} \tag{9}$$

The term $c(t)$ models a hypothesis independent baseline: importantly, as this term is uniform across all hypotheses, it has no effect on inference. It is defined in detail below.

The rMSPRT itself (shown schematically in Fig. 2) takes the form:

$$D(t) = \begin{cases} \text{Choose hypothesis } i\text{:} & \text{if } -\log P_i(t) = \min_{j\in\{1,\ldots,N\}} -\log P_j(t) \le \theta, \text{ at } t = T \\ \text{Continue sampling:} & \text{if } \min_{j\in\{1,\ldots,N\}} -\log P_j(t) > \theta, \end{cases} \tag{10}$$

where $D(t)$ is the decision at time t, $\theta \in (0, -\log(1/N)]$ is a constant threshold, and T is the decision termination time.

4.4 Cortical and thalamic baselines

The cortical baseline $c(t)$ in Eq. 9, delayed as mapped in Fig. 5 is

$$c(t+\delta_{yb}) = h(t-\delta_{bu}-\delta_{uy}) + l$$

It houses a constant baseline l and the thalamo-cortical contribution $h(t-\delta_{bu}-\delta_{uy})$, which in turn is the delayed cortical input to the thalamus

$$h(t-\delta_{bu}) = w_{yu}\frac{\sum_{i=1}^{N}\left(y_i(t+\delta_{yb}-\delta_{yu}) + \log P_i(t-\delta_{bu}-\delta_{uy}-\delta_{yu}) + c(t+\delta_{yb}-\delta_{yu})\right)}{N}$$

Here we have chosen $h(t-\delta_{bu})$ to be a scaled average of cortical contributions; nevertheless, any other hypothesis-independent function of them can be picked instead if necessary, rendering similar results.

The definitions above introduce two free parameters $l \in \mathbb{R}^+$ and $w_{yu} \in [0,1)$ that have the purpose of shaping and revealing the dynamics of the computations within rMSPRT during decision formation. The range of w_{yu} ensures that the value of computations in the cortico-thalamo-cortical, positive-feedback loop is not amplified to the point of disrupting inference in the overall loop. Crucially, since both parameters are hypothesis-independent, none affects inference.

4.5 Model parameters

To parameterize the input stochastic processes and likelihood functions of rMSPRT, we estimated the means $\mu_* n$ and $\mu_0 n$ and standard deviations $\sigma_* n$ and $\sigma_0 n$ as those over the activity between 900 and 1900 ms after stimulus onset in the MT population, per condition (shown smoothed in Fig. 1d,e). The subscript $_*$ indicates the condition when dots were predominantly moving in the direction preferred by the neuron. The subscript $_0$ indicates when they were moving against it. We dub the resulting parameter set Ω and report it in table 1. Fig. 1f shows the lognormal ISI distributions specified by Ω; solid ones are f_*, in our notation, and dashed ones are f_0, per coherence.

		Ω, Ω_d		Ω		Ω_d, $N=2$		Ω_d, $N=4$	
Coherence %	No. neurons	$\mu_* n$	$\sigma_* n$	$\mu_0 n$	$\sigma_0 n$	$\mu_0 n$	$\sigma_0 n$	$\mu_0 n$	$\sigma_0 n$
3.2	206	54.1	33.1	59.4	34.5	59.0	34.4	58.5	34.3
6.4	211	52.0	32.2	62.9	35.3	60.6	34.7	59.8	34.4
12.5	213	46.1	30.5	65.5	36.1	60.3	34.6	58.3	34.0
25.5	208	37.7	28.0	70.2	37.2	62.0	34.9	59.9	34.3
51.2	189	29.9	26.0	83.5	40.6	75.5	38.5	71.8	37.4

Table 1: Population ISI statistics (ms) in MT per coherence (first column). Second column: number of neurons for which data was available per coherence. μ: mean. σ: standard deviation. n: scaling factor. The parameter set (Ω or Ω_d) to which each value corresponds is noted above them. Note that due to the information depletion required to produce Ω_d, $\mu_0 n$ and $\sigma_0 n$ take different values for $N = 2, 4$.

We use $l = 15$, $w_{yu} = 0.4$, $n = 40$ and $\delta_{yb}, \delta_{yu}, \delta_{bu}, \delta_{uy} = 1$ (hence $\Delta = 3$) for all of our simulations. The value of latencies was set to 1 for simplicity. The values of the first three free parameters come from a manual tuning exercise with the aim of revealing a pattern in the model LIP akin to the ramp-and-fork one in LIP recordings; note that such a two-segment pattern is already guaranteed by the two parts of Eq. 5.

4.6 Spikes and continuous time out of discrete time

We have defined rMSPRT to operate over a discrete time line; however, the brain operates over continuous time. [13] demonstrated that, as an average, such a discrete-time MSPRT requires the same number of observations as those of a more general instance of MSPRT that works over continuous time and makes choices upon ISIs in actual spike-trains. This indicated that the decision sample size for correct choices, $\langle T \rangle_c$, required by the simpler discrete-time MSPRT (or rMSPRT) for decision, can be interpreted as the continuously-distributed mean decision time

$$\tau_c = (\langle T \rangle_c + 0.5)\, \mu_* n \tag{11}$$

predicted by the more general continuous-time, spiking MSPRT, where $\mu_* n$ is the mean ISI produced by a MT neuron whose preferred motion direction was matched by the stimulus and was thus firing the fastest on average. When $\mu_* < \mu_0$, as in MT [10], the hypothesis selected in error trials is that misinformed by channels with mean $\mu_0 n$ which intuitively happened to fire faster than those whose mean was actually $\mu_* n$. Hence, the mean decision time predicted by rMSPRT in error trials would be:

$$\tau_e = (\langle T \rangle_e + 0.5)\, \mu_0 n, \tag{12}$$

where $\langle T \rangle_e$ is the mean decision sample size for error trials. An instance of rMSPRT capable of making choices upon sequences of spike-trains is straightforward from the formal framework above and that introduced by [13]; nevertheless, for simplicity here we choose to work with the discrete time rMSPRT. After all, thanks to Eqs. 11 and 12 we can still interpret its behaviour-relevant predictions in terms of continuous time. We use these to compute the decision times that originate the reaction times in Figs. 3 and 4.

4.7 Estimation of lost information

We outline here how we can use the monkeys' reaction times on correct trials and the properties of the rMSPRT here, to estimate the amount of discrimination information lost by the animals. This is, the gap between all the information available in the responses of MT neurons, as fully used by the rMSPRT in Fig. 3a, and the fraction of such information actually used by monkeys.

The expected number of observations to reach a correct decision for (r)MSPRT, $\langle T \rangle_c$, depends on two quantities. First, the mean total information required for the decision, $I(\epsilon, N)$, that depends only on the error rate ϵ, and N. Second, the 'distance' between distributions of ISIs from MT neurons facing preferred and null directions of motion [13]. This distance is the Kullback-Leibler divergence from f_* to f_0

$$D = \int_x f_*(x) \log_2 \left(\frac{f_*(x)}{f_0(x)} \right) dx$$

which measures the discrimination information available between the distributions. Using these two quantities, the decision time in the (r)MSPRT is [13]:

$$\langle T \rangle_c \geq \frac{I(\epsilon, N)}{D}, \tag{13}$$

The product of our Monte Carlo estimate of $\langle T \rangle_c$ in the rMSPRT (which originated Fig. 3a) and D from the MT ISI distributions (Fig. 1f), gives an estimate of the limit $I(\epsilon, N)$ in expression 13, denoted by $\hat{I}(\epsilon, N)$.

The 'mean decision sample size' of monkeys —hence the superscript ^{m}— within this framework corresponds to $\hat{\langle T \rangle}_c^{\mathrm{m}} = (\hat{\tau}_c^{\mathrm{m}}/\mu^* n) - 0.5$ (from Eq. 11). Here, $\hat{\tau}_c^{\mathrm{m}}$ is the estimate of the mean decision time of monkeys for correct choices, per condition; that is, the reaction time from Fig. 3a minus some constant non-decision time. With $\hat{\langle T \rangle}_c^{\mathrm{m}}$ and $\hat{I}(\epsilon, N)$, we can estimate the corresponding discrimination information available to the monkeys in this framework as $\hat{D}^{\mathrm{m}} = \hat{I}(\epsilon, N) / \hat{\langle T \rangle}_c^{\mathrm{m}}$ (from expression 13).

Fig. 3b compares D (red line) to $\hat{D}^{\mathrm{m}}$ (blue/green lines and shadings) for monkeys, using non-decision times in a plausible range of 200–300 ms. Fig. 3c shows the discrimination information lost by monkeys as the percentage of D, $\left[1 - \left(\hat{D}^{\mathrm{m}}/D\right)\right] \times 100$ %.

4.8 Information depletion procedure

Expression 13 implies that the reaction times predicted by rMSPRT should match those of monkeys if we make the algorithm lose as much information as the monkeys did. We did this by producing a new parameter set that brings f_0 closer to f_* per condition, assuming 250 ms of non-decision time; critically, simulations like those in Fig. 4 will give about the same rMSPRT reaction times regardless of the non-decision time chosen, as long as it is the same assumed in the estimation of lost information and this information-depletion procedure.

An example of the results of information depletion in one condition is in Fig. 3d. We start with the original parameter set extracted from MT recordings, Ω ('preferred' and 'null' densities in Fig. 3d), and keep μ_* and σ_* fixed. Then, we iteratively reduce/increase the differences $|\mu_0 - \mu_*|$ and $|\sigma_0 - \sigma_*|$ by the same proportion, until we get new parameters μ_0 and σ_0 that, together with $\mu_*$and σ_*, specify preferred ('preferred' in Fig. 3d) and null ('new null') density functions that bear the same discrimination information estimated for monkeys, $\hat{D}^{\mathrm{m}}$; hence, they exactly match the information loss in the solid lines in Fig. 3c. Intuitively, since the 'new null' distribution in Fig. 3d is more similar to the 'preferred' one than the 'null', the Kullback-Leibler divergence between the first two is smaller than that between the latter two. The resulting parameter set is dubbed Ω_d and reported in table 1.

4.9 Simulation procedures

For rMSPRT decisions to be comparable to those of monkeys, they must exhibit the same error rate, $\epsilon \in [0, 1]$. Error rates are taken to be an exponential function of coherence (%), s, fitted by non-linear least squares ($R^2 > 0.99$) to the behavioural psychometric curves from the analysed LIP database, including 0 and 72.4 % coherence for this purpose. This resulted in:

$$\epsilon = \begin{cases} 0.50 \exp(-0.11 s), & \text{for } N = 2 \\ 0.75 \exp(-0.08 s), & \text{for } N = 4 \end{cases} \tag{14}$$

Since monkeys are trained to be unbiased regarding choosing either target, initial priors for rMSPRT are set flat ($P(H_i) = 1/N$ for all i) in every simulation. During each Monte Carlo experiment, rMSPRT made decisions with error rates from Eq. 14. The value of the threshold, θ, was iteratively found to satisfy ϵ per condition [13]; an example of the θs found is shown in Fig. S3a in the Supplement. Decisions were made over data, $x_j(t)/n$, randomly sampled from lognormal distributions specified for all channels by means and standard deviations μ_0 and σ_0, respectively. The exception was a single channel where the sampled distribution was specified by μ_* and σ_*. This models the fact that MT neurons respond more vigorously to visual motion in their preferred direction compared to motion in a null direction, *e.g.* against the preferred. Effectively, this simulates macaque MT neural activity during the random dot motion task. The same parameters were used to specify likelihood functions per experiment. All statistics were from either the Ω or Ω_d parameter sets as noted per case. Note that the statistics actually used in the simulations are those from MT in table 1, divided by the scaling factor n.

Author Contributions JC deduced the rMSPRT, and performed all simulations and data analyses; all authors discussed the results and wrote the article.

Acknowledgements We thank Anne Churchland, Roozbeh Kiani and Michael Shadlen for sharing their experimental data and Abhinav Singh, Mathew Evans and Silvia Maggi (Humphries lab) for discussions. This work was supported by a National Council of Science and Technology (CONACyT) Fellowship (JAC) and a Medical Research Council Senior non-Clinical Fellowship (MDH).

References

[1] Albantakis, L. and Deco, G. (2009). The encoding of alternatives in multiple-choice decision making. *Proc Natl Acad Sci U S A*, 106(25):10308–10313.

[2] Alexander, G. E. and Crutcher, M. D. (1990). Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends Neurosci*, 13(7):266–271.

[3] Alexander, G. E., DeLong, M. R., and Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu Rev Neurosci*, 9:357–381.

[4] Basso, M. A. and Wurtz, R. H. (2002). Neuronal activity in substantia nigra pars reticulata during target selection. *The Journal of Neuroscience*, 22(5):1883–1894.

[5] Baum, C. and Veeravalli, V. (1994). A sequential procedure for multihypothesis testing. *IEEE Transactions on Information Theory*, 40(6):1994–2007.

[6] Beck, J. M., Ma, W., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., Shadlen, M. N., Latham, P. E., and Pouget, A. (2008). Probabilistic population codes for bayesian decision making. *Neuron*, 60(6):1142–1152.

[7] Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E., and Pouget, A. (2012). Not noisy, just wrong: the role of suboptimal inference in behavioral variability. *Neuron*, 74(1):30–39.

[8] Bogacz, R. and Gurney, K. (2007). The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Comput*, 19(2):442–477.

[9] Bogacz, R. and Larsen, T. (2011). Integration of reinforcement learning and optimal decision-making theories of the basal ganglia. *Neural Comput*.

[10] Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (1992a). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci*, 12(12):4745–4765.

[11] Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (1992b). Responses of single neurons in macaque MT/V5 as a function of motion coherence in stochastic dot stimuli, Neur Sig Arch, NSA2004.1. www.neuralsignal.org.

[12] Brunton, B. W., Botvinick, M. M., and Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science*, 340:95–98.

[13] Caballero, J. A., Lepora, N. F., and Gurney, K. N. (2015). Probabilistic decision making with spikes: From ISI distributions to behaviour via information gain. *PLoS ONE*, 10(4):e0124787.

[14] Churchland, A. K., Kiani, R., and Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nat Neurosci*, 11(6):693–702.

[15] Ding, L. and Gold, J. I. (2010). Caudate encodes multiple computations for perceptual decisions. *J Neurosci*, 30(47):15747–15759.

[16] Ding, L. and Gold, J. I. (2012a). Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. *Cerebral Cortex*, 22(5):1052–1067.

[17] Ding, L. and Gold, J. I. (2012b). Separate, causal roles of the caudate in saccadic choice and execution in a perceptual decision task. *Neuron*, 75(5):865–874.

[18] Ditterich, J. (2006). Stochastic models of decisions about motion direction: behavior and physiology. *Neural Networks*, 19(8):981–1012.

[19] Ditterich, J. (2010). A comparison between mechanisms of multi-alternative perceptual decision making: Ability to explain human behavior, predictions for neurophysiology, and relationship with decision theory. *Front Neurosci*, 4:184.

[20] Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., and Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *J Neurosci*, 32(11):3612–3628.

[21] Frank, M. J., Seeberger, L. C., and O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science*, 306(5703):1940–1943.

[22] Furman, M. and Wang, X.-J. (2008). Similarity effect and optimal control of multiple-choice decision making. *Neuron*, 60(6):1153–1168.

[23] Gerfen, C. R. and Wilson, C. J. (1996). Chapter II, the basal ganglia. In Swanson, L., Björklund, A., and Hökfelt, T., editors, *Integraded systems of the CNS, part III Cerebellum, basal ganglia, olfactory system*, volume 12 of *Handbook of Chemical Neuroanatomy*, pages 371 – 468. Elsevier.

[24] Grossberg, S. and Pilly, P. K. (2008). Temporal dynamics of decision-making during motion perception in the visual cortex. *Vision research*, 48(12):1345–1373.

[25] Hamani, C., Saint-Cyr, J. A., Fraser, J., Kaplitt, M., and Lozano, A. M. (2004). The subthalamic nucleus in the context of movement disorders. *Brain*, 127(Pt 1):4–20.

[26] Handel, A. and Glimcher, P. W. (1999). Quantitative analysis of substantia nigra pars reticulata activity during a visually guided saccade task. *J Neurophysiol*, 82(6):3458–3475.

[27] Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., and Brody, C. D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature*, 520(7546):220–223.

[28] Heimer, L., Zahm, D. S., and Alheid, G. F. (1995). Basal ganglia. In Paxinos, G., editor, *The Rat Nervous System*, pages 579–628. Academic Press, USA, 2nd edition.

[29] Hernández, A., Zainos, A., and Romo, R. (2002). Temporal evolution of a decision-making process in medial premotor cortex. *Neuron*, 33(6):959–972.

[30] Hikosaka, O., Sakamoto, M., and Miyashita, N. (1993). Effects of caudate nucleus stimulation on substantia nigra cell activity in monkey. *Exp Brain Res*, 95(3):457–472.

[31] Huk, A. C. and Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *The Journal of neuroscience*, 25(45):10420–10436.

[32] Humphries, M. D., Stewart, R. D., and Gurney, K. N. (2006). A physiologically plausible model of action selection and oscillatory activity in the basal ganglia. *J Neurosci*, 26(50):12921–12942.

[33] Kira, S., Yang, T., and Shadlen, M. N. (2015). A neural implementation of wald's sequential probability ratio test. *Neuron*, 85(4):861–873.

[34] Larsen, T., Leslie, D. S., Collins, E. J., and Bogacz, R. (2010). Posterior weighted reinforcement learning with state uncertainty. *Neural Comput*, 22(5):1149–1179.

[35] Law, C.-T. and Gold, J. I. (2008). Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nature neuroscience*, 11(4):505–513.

[36] Law, C.-T. and Gold, J. I. (2009). Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nat Neurosci*, 12(5):655–663.

[37] Lepora, N. F. and Gurney, K. N. (2012). The basal ganglia optimize decision making over general perceptual hypotheses. *Neural Comput*, 24(11):2924–2945.

[38] Lo, C.-C. and Wang, X.-J. (2006). Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nat Neurosci*, 9(7):956–963.

[39] Mazurek, M. E., Roitman, J. D., Ditterich, J., and Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cereb Cortex*, 13(11):1257–1269.

[40] McFarland, N. R. and Haber, S. N. (2002). Thalamic relay nuclei of the basal ganglia form both reciprocal and nonreciprocal cortical connections, linking multiple frontal cortical areas. *J Neurosci*, 22(18):8117–8132.

[41] Middleton, F. A. and Strick, P. L. (2000). Basal ganglia and cerebellar loops: motor and cognitive circuits. *Brain Res Brain Res Rev*, 31(2-3):236–250.

[42] Nambu, A., Yoshida, S., and Jinnai, K. (1988). Projection on the motor cortex of thalamic neurons with pallidal input in the monkey. *Exp Brain Res*, 71(3):658–662.

[43] Petras, J. M. (1971). Connections of the parietal lobe. *J Psychiatr Res*, 8(3):189–201.

[44] Redgrave, P., Prescott, T. J., and Gurney, K. (1999). The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience*, 89(4):1009–1023.

[45] Resulaj, A., Kiani, R., Wolpert, D. M., and Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, 461(7261):263–266.

[46] Roitman, J. D. and Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci*, 22(21):9475–9489.

[47] Saint-Cyr, J. A., Ungerleider, L. G., and Desimone, R. (1990). Organization of visual cortical inputs to the striatum and subsequent outputs to the pallido-nigral complex in the monkey. *J Comp Neurol*, 298(2):129–156.

[48] Sivia, D. S. and Skilling, J. (2006). *Data Analysis–A Bayesian Tutorial*. Oxford Science Publications, USA, 2nd edition.

[49] Wald, A. (1947). *Sequential analysis*. Wiley publications in mathematical statistics. John Wiley, New York.

[50] Wald, A. and Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistics*, 19(3):326–339.

[51] Watanabe, Y. and Funahashi, S. (2004). Neuronal activity throughout the primate mediodorsal nucleus of the thalamus during oculomotor delayed-responses. i. cue-, delay-, and response-period activity. *J Neurophysiol*, 92(3):1738–1755.

[52] Wong, K.-F., Huk, A. C., Shadlen, M. N., and Wang, X.-J. (2007). Neural circuit dynamics underlying accumulation of time-varying evidence during perceptual decision making. *Front Comput Neurosci*, 1:6.

S1 Supplementary materials

S1.1 Information loss for a perfected match of monkey reaction times

The slight deviation of the mean reaction times of rMSPRT *vs* those of monkeys in Fig. 4a stems from the expression 13 being an inequality. Due to this, $\hat{I}(\epsilon, N)$ is a likely over-estimate of $I(\epsilon, N)$. Dividing $\hat{I}(\epsilon, N)$ by $\langle \hat{T} \rangle_c^{\text{m}}$ hence gives an over-estimate of the monkey discrimination information, $\hat{D}^{\text{m}}$. If then rMSPRT uses statistics consistent with this over-estimated $\hat{D}^{\text{m}}$, it renders under-estimated reaction times. This minor discrepancy can be corrected by further multiplying $\hat{D}^{\text{m}}$, per condition, by the corresponding ratio of the decision time of the model over that of the monkey, from Fig. 4a. Repeating the experiments with the implied parameter set would trivially render rMSPRT reaction times that more perfectly match those of monkeys (not shown). This will likely carry with it a better match in error trials which is, again, unconstrained in the procedure. Nonetheless, this exercise gives us the full information loss associated to such perfected match, shown in Fig. 3c as dashed lines for a 250 ms non-decision time (compare to solid lines); this constitutes a further refined measure of the minimum information lost by the animals according to this framework.

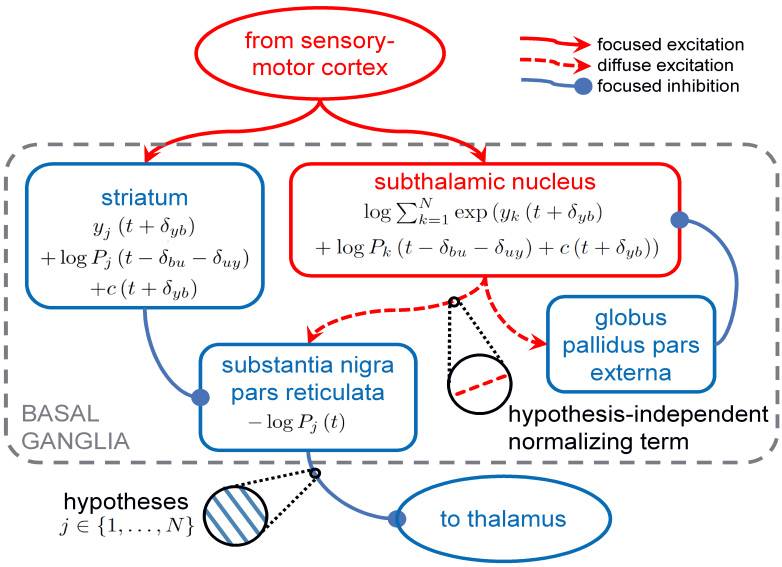

Figure S1: Mapping of rMSPRT computations to the basal ganglia. Parallel computations for N hypotheses — indexed by j— mapped onto the basal ganglia nuclei, within the grey dashed box (see [8, 9, 34]). Same conventions and notation as in Fig. 5. All computations are delayed with respect to the substantia nigra pars reticulata. $\log\sum_{k=1}^{N}\exp(y_k(t+\delta_{yb})+\log P_k(t-\delta_{bu}-\delta_{uy})+c(t+\delta_{yb}))$: normalization term (from Eq. 9), putting together the cortical computations for all hypotheses into a hypothesis-independent contribution. Note that the model striatum represents a copy of the cortical signal (as in Fig. 5) but its influence on the substantia nigra pars reticulata is the negative of such cortical input.

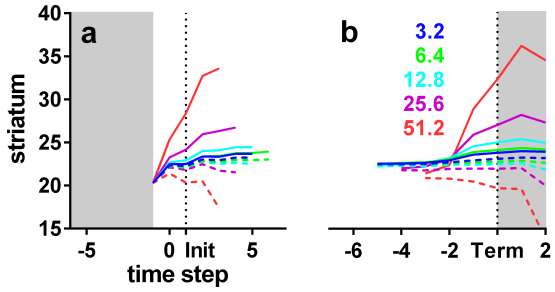

Figure S2: Qualitative prediction for the time course of striatum in rMSPRT. Same conventions as Fig. 7. (a) Aligned at decision initiation (Init; $t=1$; dotted line). (b) Aligned at decision termination (Term; dotted line). Single Monte Carlo experiment with $N=2$ and 800 total trials.

S1.2 Function of a diffuse cortico-thalamic projection

rMSPRT proposes a role for the spatially diffuse cortico-thalamic projection [40]. The thalamic baseline, contributed by cortex, $h(t-\delta_{bu})$, is predicted to constitute the offset within which updated inference results (posteriors made priors) from basal ganglia are conveyed back to cortex. This is another unique prediction of rMSPRT, not accounted for by previous sequential ideal observers. Fig. S3b shows $h(t-\delta_{bu})$ along with the corresponding output of the model thalamus in inRF and outRF settings. Without the increasing range facilitated by a diffuse cortico-thalamic contribution (red in model) the biological thalamic output (blue in model) would be at risk of saturating at 0, especially in outRF trials where the basal ganglia inhibition is strongest, wasting then the segregation of hypotheses carried by it, represented by the fed-back posteriors in the model. If the biological thalamus were thus saturated, rMSPRT also predicts that the cortex would be limited to produce likelihood-proportional responses only on the basis of the immediately incoming data, instead of making use of the whole sequence.

Lastly, note that the increasing $h(t-\delta_{bu})$ becomes $h(t-\delta_{bu}-\delta_{uy})$ at the model sensory-motor cortex. This

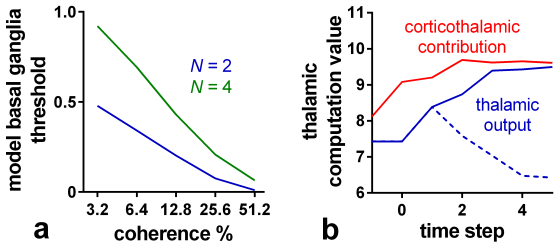

Figure S3: Threshold in basal ganglia and influence of cortico-thalamic contribution. (a) Position of the threshold, θ, set at the model basal ganglia in simulations like those of Figs. 4, 6 and 7, per condition. (b) Example mean cortico-thalamic contribution, $h\left(t-\delta_{bu}\right)$ (red), compared to the mean thalamic output in inRF settings (solid blue) and outRF ones (dashed blue) for 25 % coherence and $N=2$. Both (a) and (b) come from single Monte Carlo experiments with 800, 1200 trials for $N=2,4$, respectively.

is the time changing component of the cortical baseline, $c\left(t+\delta_{yb}\right)$ (Fig. 5 and Methods). Thus, we argue that $h\left(t-\delta_{bu}-\delta_{uy}\right)$ may correspond to the increasing choice-independent drive, dubbed "urgency signal", revealed by [20] when averaging population responses of LIP neurons during the dot motion task across choices, for separate coherence levels.