Bayesian computation through cortical dynamics

Hansem Sohn[*1], Devika Narain[*1], Nicolas Meirhaeghe[2], Mehrdad Jazayeri[1]

[1] Dept. of Brain & Cognitive Sciences, McGovern Institute for Brain Research, MIT; [2] Harvard-MIT Division of Health Sciences and Technology, Bioastronautics Training Program; * Equal contribution

Statistical regularities in the environment give rise to prior beliefs that help optimize behavior under sensory uncertainty. Bayesian theory formalizes how prior beliefs can be leveraged and has aided our understanding of sensorimotor behavior. However, we do not yet know how neural circuits implement Bayesian integration. To tackle this question, we recorded neural activity in the dorsomedial frontal cortex (DMFC) of monkeys trained to perform a time-interval reproduction task (Ready-Set-Go, 'RSG') with two distinct prior statistics and various levels of sensory uncertainty (Figure 1a,b). Across trials, the sample interval between Ready and Set, t_s, was sampled from one of two prior distributions, a 'Short' and 'Long' prior (Figure 1c,d). Because the two distributions were overlapping, the task offered an opportunity to characterize how prior beliefs are integrated with sensory measurements. Animals learned to perform the task as verified by positive regression slopes between the produced interval, t_p, and t_s (Figure 1e). Importantly, the slopes were less than unity (sign-rank test, $p<10^{-9}$) indicating that animals biased their responses toward the mean of the cued prior, consistent with Bayesian model predictions (Figure 1e). These results suggest that the RSG task provides a suitable platform for investigating the neural basis of Bayesian integration.

Response profile of individual neurons in DMFC showed rich and heterogeneous dynamics that depended on the prior condition (Figure 2a), suggesting that neurons in this area were modulated according to the animal's prior belief about t_s. To better understand neural computations across the population, we applied principal component analysis to visualize the evolution of neural trajectories in a state space. In the estimation epoch between Ready and Set, the most salient feature in the data was a low-dimensional rotation of neural trajectories (Figure 2b). Critically, this rotation was temporally tuned to the support of each prior, suggesting that rotational dynamics may play a key role in enabling Bayesian integration. In the production epoch (Figure 2c), after the Set flash rapidly displaced neural states for nearly 200 ms, neural trajectories had an orderly structure with respect to t_s and evolved toward a common terminal state (Go) at progressively slower speed for longer t_s. These features during the production epoch are consistent with our previous findings showing that the different neural states across t_s after Set serve as initial conditions to set up the speed of the ensuing trajectories that eventually control time of movement initiation.

How could rotational dynamics during the estimation epoch generate the prior-dependent biases in the initial conditions for the production epoch? An inherent property of a curved trajectory is that when projected onto a line connecting the two ends of the trajectory, equidistant points along the trajectory become warped, similar to the effect of Bayesian integration on the behavior (Figure 3a). Based on this geometrical intuition, we hypothesized that neural states along the rotating trajectory implicitly represent the moment-by-moment Bayesian estimate of elapsed time that can be decoded when projected onto a line in the state space. Consistent with this hypothesis, we found that projections of neural states during the support of the prior onto a 1D 'encoding axis' exhibited a regression to the mean compatible with Bayes-optimal behavior (Figure 3c). Neural states after Set also corresponded closely to the Bayesian estimate of t_s when projected onto a decoding axis (Figure 3d), suggesting that the Set-evoked transient response mapped neural states before Set onto the decoding axis after Set (Figure 3b). Finally, we found that neural states along the decoding axis adjust the speed of the ensuing dynamics (Figure 3e), which allow the animal to control movement initiation time (Figure 3f). Together, these step-by-step analyses suggest that the rotation during the estimation epoch provides a Bayesian estimate of elapsed time that sets the speed of dynamics during the production epoch allowing animals to produce Bayes-optimal behavior (Figure 3b).

In summary, we found that prior statistics create low-dimensional manifolds in the frontal cortex that cause the mapping of sensory inputs to motor outputs to be biased in accordance with Bayesian inference. These results uncover a simple yet general principle whereby prior beliefs are embedded in neural circuits and shape cortical latent dynamics to influence behavior.

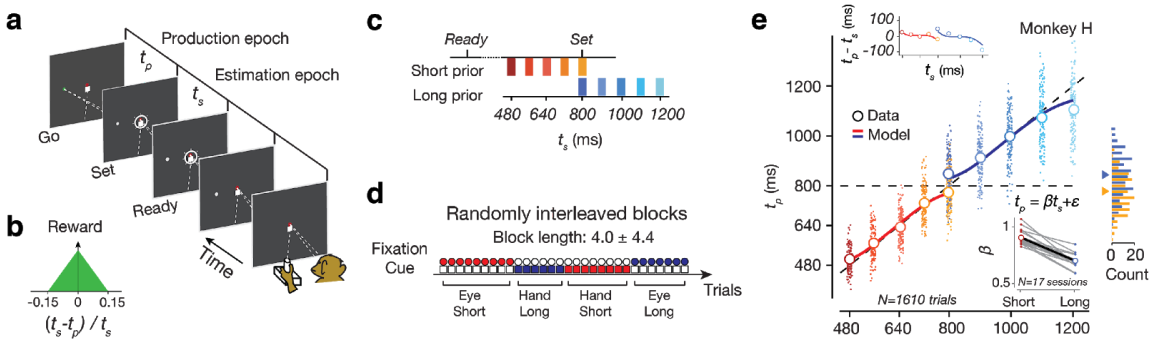

Figure 1. Task and behavior. a) Schematic of a trial of the Ready-Set-Go task. The monkey fixates a central fixation spot and holds a joystick. After a variable delay, a white target is presented randomly on the left or right and subsequently two flashes – Ready followed by Set – are presented. The animal has to estimate the sample interval, t_s, between Ready and Set (estimation epoch), and reproduce this interval (t_p) via a delayed response toward the target either by a saccade or joystick movement (production epoch). **b) Feedback**. The monkey receives juice reward (green region) if the relative error $(t_p-t_s)/t_s$ is smaller than 0.15. **c) Prior distributions of t_s**. On each trial, t_s is sampled from one of two discrete, uniform prior distributions ('Short' and 'Long') overlapping at 800 ms. **d) Trial types**. The experiment consisted of 2 prior conditions (Short and Long) x 2 effectors (Eye and Hand) x 2 target directions (Left and Right). The 4 conditions for the prior and effector were randomly interleaved across blocks of trials, cued throughout the trial by the different fixation spots. **e) Behavior**. A representative session showing individual t_p pooled across effectors and target directions (small filled circles) and corresponding averages (large open circles) for each t_s. The red and blue lines are Bayesian model predictions for Short and Long prior conditions, respectively. Right: Histograms of t_p for the overlapping t_s for the two prior conditions with the corresponding averages (triangles). Top-left inset: Average error (i.e., bias) for each t_s (data: circles; Bayesian model: lines). Bottom-right inset: Slopes of regression lines relating t_p to t_s for each prior condition (gray lines: individual sessions, black lines : averages).

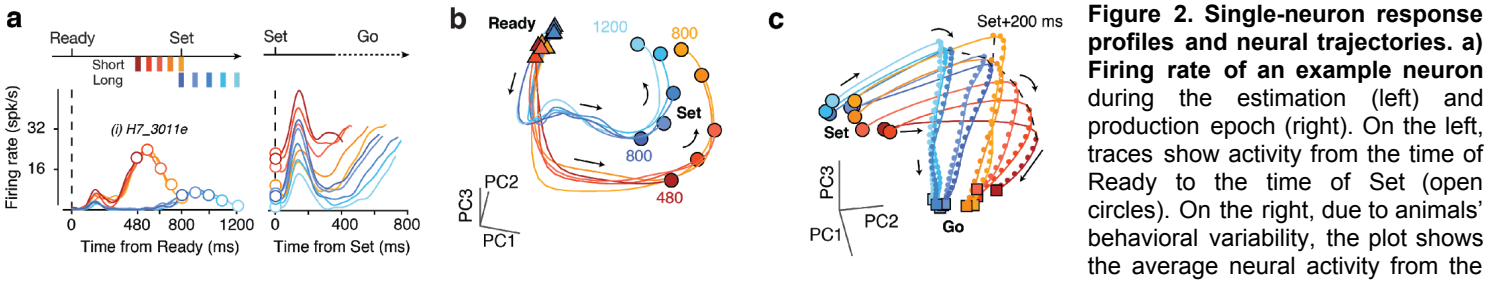

Figure 2. Single-neuron response profiles and neural trajectories. a) Firing rate of an example neuron during the estimation (left) and production epoch (right). On the left, traces show activity from the time of Ready to the time of Set (open circles). On the right, due to animals' behavioral variability, the plot shows the average neural activity from the time of Set to the minimum t_p for each t_s. **b) Neural trajectories during the estimation epoch**. A representative dataset is shown for a plot of the first three principal components (PCs) of neural activity (Monkey H, Eye Left condition) during the estimation epoch. Arrows illustrate the direction along which the trajectories evolve with time (triangles for Ready and circles for Set). **c) Neural trajectories in the production epoch** (circles for Set and squares for Go). For each prior condition, the dashed line connects the neural states along the different trajectories 200 ms after Set. The small dots in the trajectories show neural states at 20-ms increments. The distance between consecutive dots is proportional to the speed of the neural trajectory (e.g., higher speed for dark red compared to light blue).

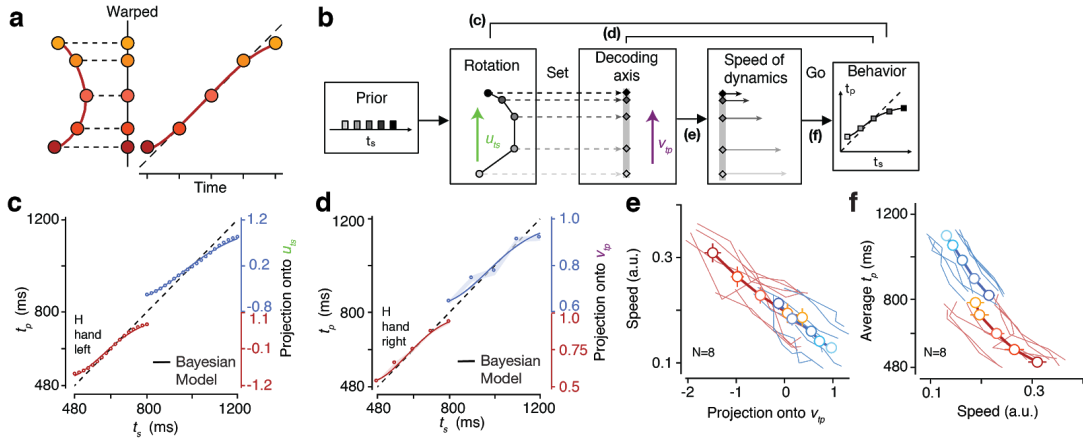

Figure 3. Neural signatures of Bayesian integration. a) A geometric illustration of why linear projection of points along a 2D curve onto a 1D line could cause warping mimicking the 'regression to the mean' effect caused by Bayesian integration. **b) The cascade of computations** during estimation and production epoch. The prior distribution of t_s (leftmost panel) establishes rotational dynamics during the estimation epoch (second leftmost panel). Geometrically, projection of the points along the rotating trajectory onto an encoding axis (green vector, u_{ts}) creates a warped 1D representation of time that exhibits prior-dependent biases. Presentation of Set maps neural states onto a decoding axis (middle panel; purple vector, v_{tp}). Neural states along v_{tp} serve as the system's initial conditions during the production epoch, dictating the speed of neural trajectories (second rightmost panel) and allowing the system to generate Bayes-optimal behavior (rightmost panel). The parenthetical labels (c) to (f) are evaluated quantitatively in the corresponding panels. **c) Projection of the prior-dependent rotating trajectory on u_{ts}**. As u_{ts}, we chose the vector pointing from the states associated with the shortest to the longest t_s for each prior condition. Circles show projections every 20 ms for Short (red) and Long (blue) prior conditions. **d) Projection of neural states 200 ms after Set onto v_{tp}**. Results of analyses on the decoding axis in the same format as in c) for the encoding axis. **e) Speed at which neural states evolved during the production epoch** (from Set + 200 ms to Go) as a function of the neural state along v_{tp}. The speed was estimated by averaging distances between successive bins in the state space (thin lines for individual datasets, 2 animals x 2 effectors x 2 directions, and thick line for the averages with error bars for s.e.m.). **f) Average produced interval (t_p) as a function of speed** at which neural states evolved during the production epoch. Results are shown with the same format as in d).