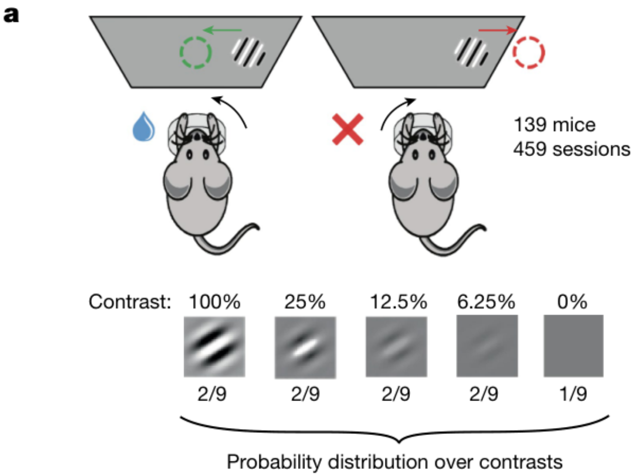
Brain-wide representations of prior information in mouse decision-making

Luis Luarte

UDD

October 28, 2025

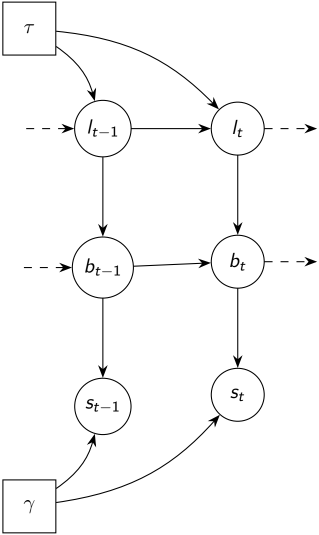
Decision making task



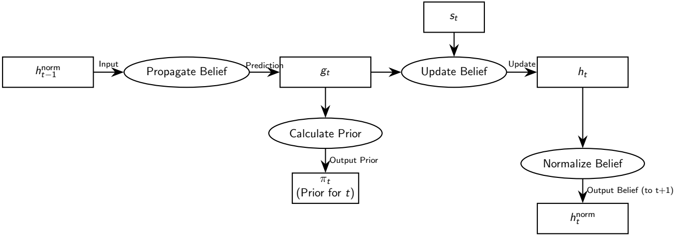
Markov transition matrices represent task mechanics

- Length transition matrix $P(l_t = n+1 | l_{t-1} = n) = H_\tau(n)$
- Block type transition matrix (continue) $P(b_t | b_{t-1}, continue)$
- Block type transition matrix (switch) $P(b_t | b_{t-1}, switch)$

Markov transition matrices represent task mechanics



Hidden Markov Model forward pass



- Propagate belief:

$$g_t(l_t, b_t) = \sum_{l_{t-1}, b_{t-1}} h_{t-1}(l_{t-1}, b_{t-1}) \cdot P(l_t | l_{t-1}) \cdot P(b_t | b_{t-1}, l_t) \tag{1}$$

- Calculate prior:

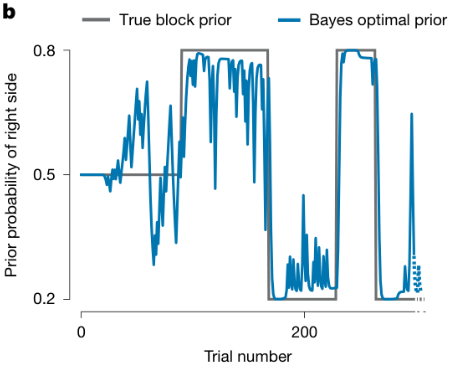
$$\pi_t = 0.5 \cdot P(b_t = 0 | s_{1:(t-1)}) + \gamma \cdot P(b_t = 1 | s_{1:(t-1)}) + (1 - \gamma) \cdot P(b_t = -1 | s_{1:(t-1)}) \tag{2}$$

$$P(b_t = k | s_{1:(t-1)}) = \frac{\sum_{l_t} g_t(l_t, b_t = k)}{\sum_{l_t, b_t} g_t(l_t, b_t)} \tag{3}$$

- Update belief:

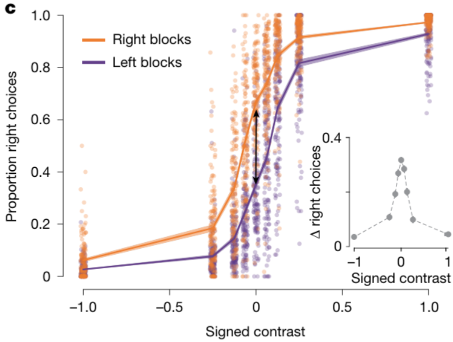
$$h_t(l_t, b_t) = P(s_t | b_t) \cdot g_t(l_t, b_t) \tag{4}$$

Task mechanics and Bayesian prior predictions



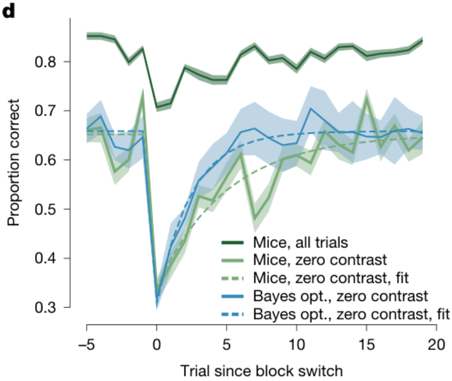
- $P(s_t = 1) = 0.2 or 0.8$
- For 90 trials $s_t = 0$
- Block length $H_\tau(n), n \geq 20, n \leq 100, \tau = 60$
- Positive feedback = water, Negative feedback = white noise + timeout
- Next trial after delay + wheel fixed

Prior-based mechanism > action bias



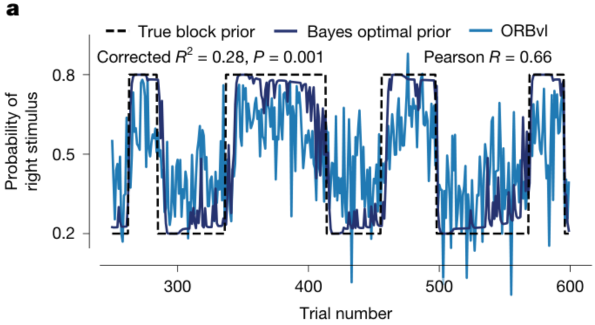
- Negative values = stimulus on left
- Positive values = stimulus on right
- Zero = no contrast stimulus

Performance recovers slower than predicted by agent sampling prior



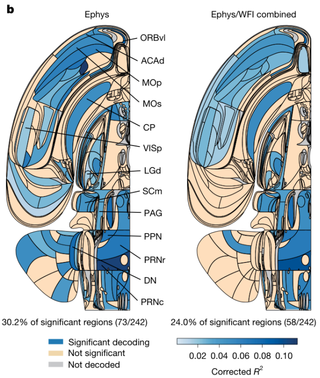
- $p(\text{correct at trial } t) = \begin{cases} B + (A - B) \cdot e^{-t/\tau} & \text{if } t \geq 0 \\ B & \text{if } t < 0 \end{cases}$
- A and B are free parameters, A = performance drop, B = asymptotic performance

Prior decoding during the ITI



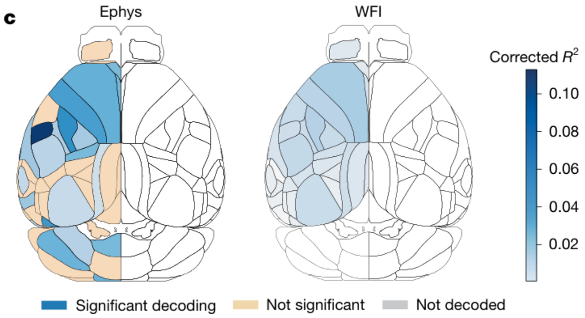
- Single cell recording (almost, spike-sorting) using neuropixels
- Decoding is a LASSO linear regression $\hat{y}_i = x_i w + b$ over Bayes-optimal prior π_t
- LASSO shrinks uncorrelated neuron activity, and is robust to outlier (drift)
- ORB_{vl} value representation, cross validation reporting median R^2

Prior decoding during the ITI



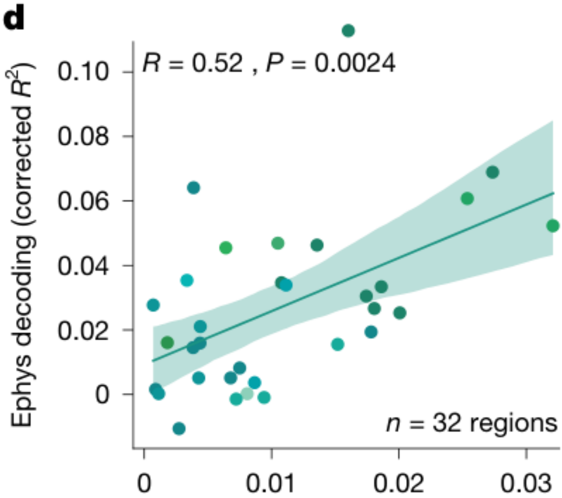
- Wide Field Calcium Imaging, aggregate cortical activity
- *VGCCs* → *Calcium influx* → *Vesicle fusion* → *GECI* → *Bright*
- Merge maps with Fisher's combined probability test
- Sensory, Associative, Motor, Sub-cortical

Prior decoding during the ITI



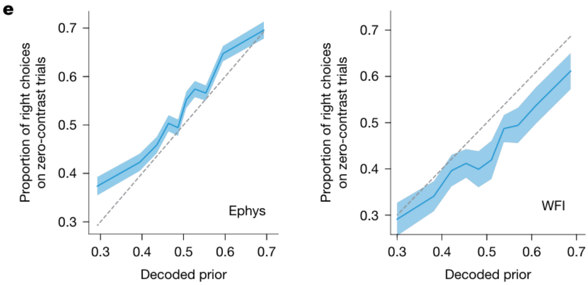
- Cross-modal comparison

Both modalities are correlated in decoding



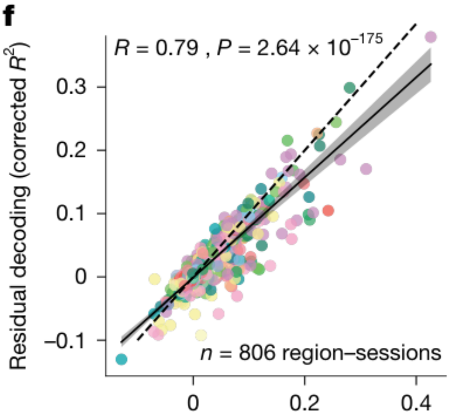
- Cross-modal comparison

Decoded signal strength coincides with theoretical importance of prior



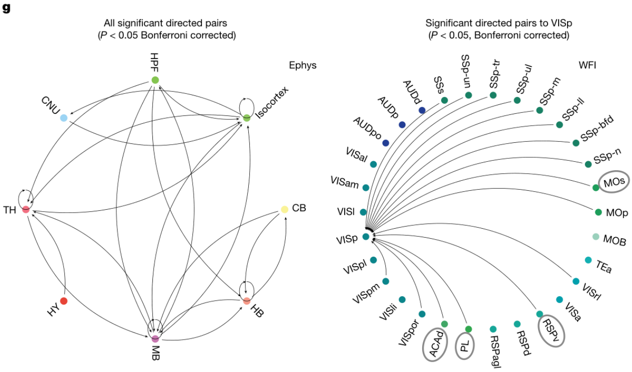
- Zero-contrast trials represent prior-guided action (psychometric curves)
- $P(s_t = Right|History, Evidence \to 0) = \pi_t$

Neuron prior model > Embodied model



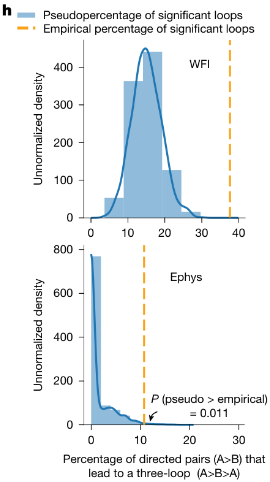
- Better region prior decoding predicts increased $residuals_{Bayes\ prior-Embodied\ features}$
- Prior decoding regions decode Emobodied residuals
- $Residuals_{Embodied\ model} = \beta_0 + \beta_{Neuron\ activity}$

Evidence for the 'Bayesian network'



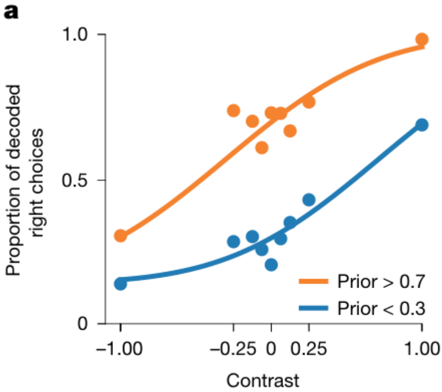
- Brain-wide bi-directional network encode prior
- Non-hierarchical modulation of sensory areas

Evidence for the 'Bayesian network'



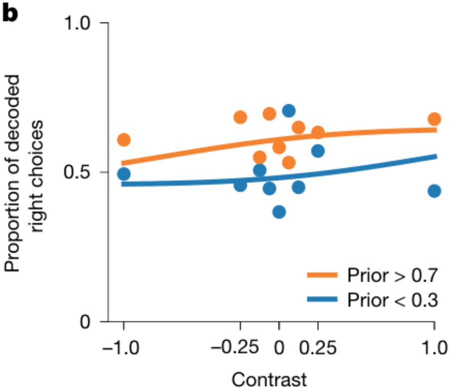
- Pseudosession: *Sample from* $G_{mechanism} \rightarrow$ *build M series* $\rightarrow$ *compute* $\pi_{synthetic} \rightarrow GC(\pi_{synthetic})$

Neurometric curves



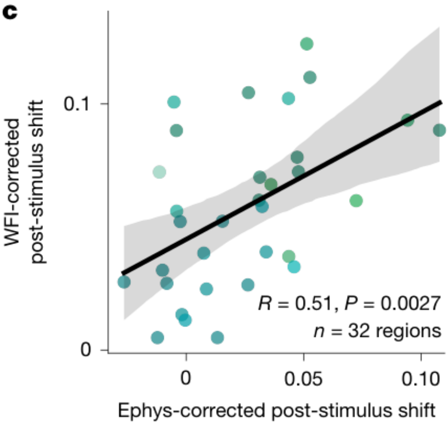
- Post-stimulus measurement
- Determine if the prior persists and biases the integration of stimulus
- Action is greedy over decoded prior, represent 'update belief'

Neurometric curves



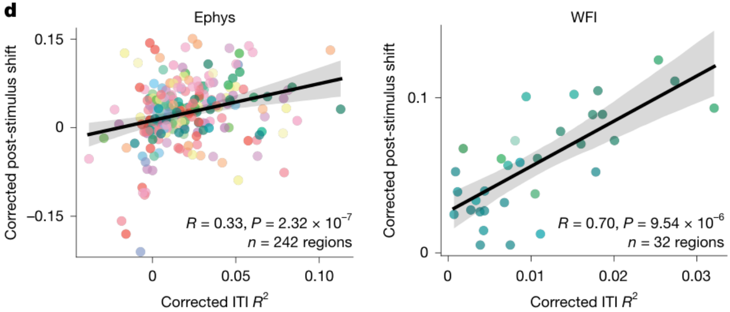
- Same but for ITI data
- Separation is significant but significantly reduced
- This is just the prior

Neurometric curves



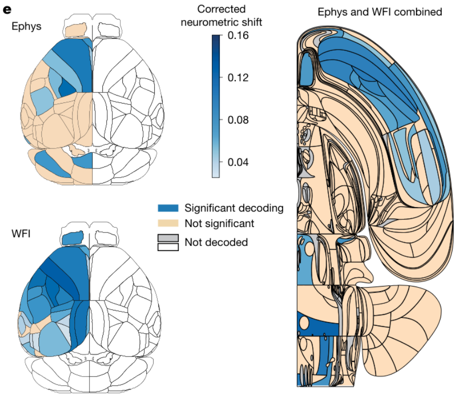
- The shift is similar for cell-level and aggregate data

Neurometric curves



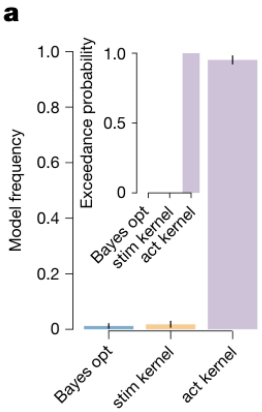
- Areas that strongly decode the prior, also show higher level of bias post stimulus

Neurometric curves



- From belief propagation to 'update belief' the prior is represented brain-wide

Heuristic models comparison

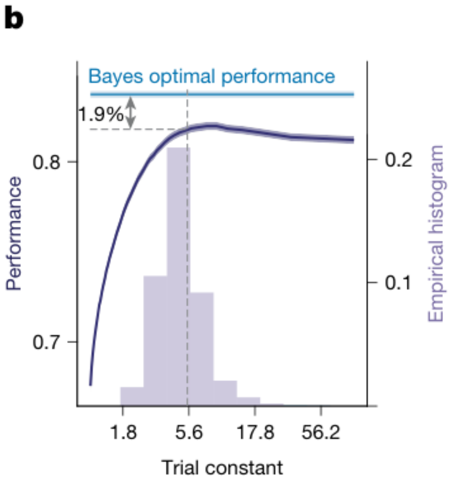


- Model selection favored the action kernel heuristic

$$\pi_t = (1-\alpha)\pi_{t-1} + \alpha \cdot [s_{t-1} = Right] \quad (5)$$

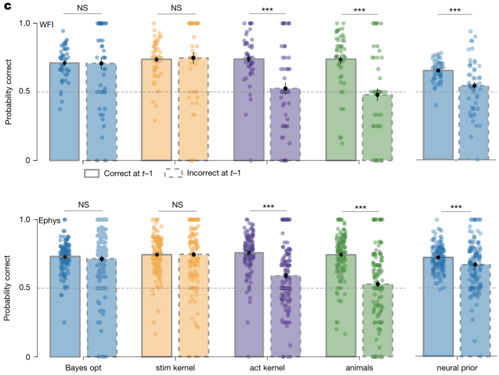
$$\pi_t = (1-\alpha)\pi_{t-1} + \alpha \cdot [a_{t-1} = Right] \quad (6)$$

Heuristic models comparison



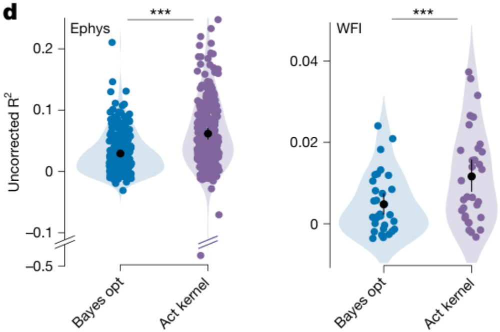
- Action kernel heuristic is near optimal
- $\tau_{optimal} \approx \tau_{empirical}$

Heuristic models comparison



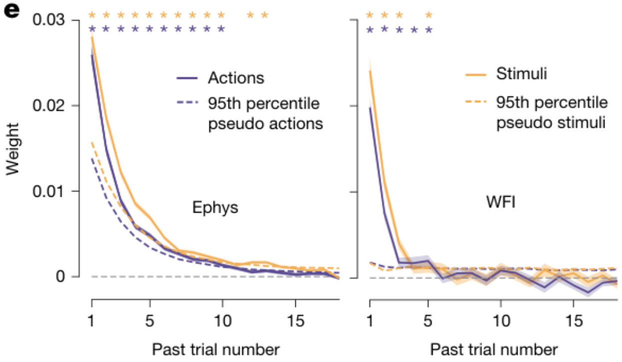
- Action kernel better captures effects of being wrong at $t - 1$
- This is consistent with neural prior

Heuristic models comparison



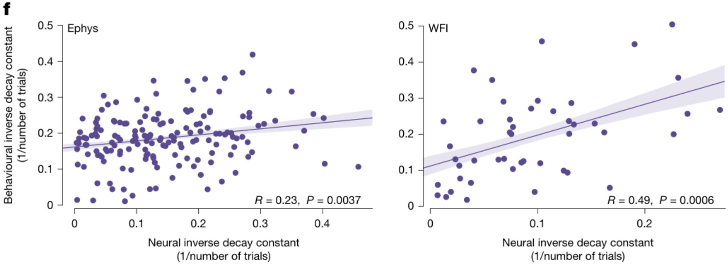
- Decoding the action kernel provides better fit with neural activity

Moving average models evidence



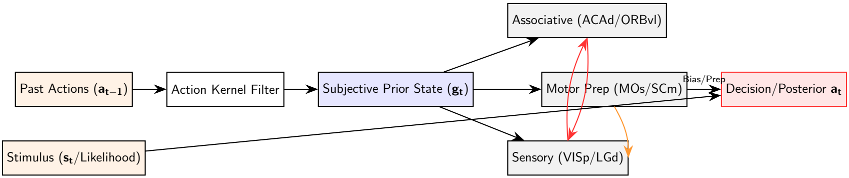
- Neural prior considers past 5 trials of information
- Normalized coefficients
- Stepwise regression

Moving average models evidence



- 'Memory' is present in behavioral and neural fits
- Action kernel strategy is likely implemented by neural activity

Conclusion



Associative (ACAd/ORBvl)
Past Actions ($\mathbf{a_{t-1}}$)
Action Kernel Filter
Subjective Prior State ($\mathbf{g_t}$)
Motor Prep (MOs/SCm)
Bias/Prep
Decision/Posterior $\mathbf{a_t}$
Stimulus ($\mathbf{s_t}$/Likelihood)
Sensory (VISp/LGd)