BARCCSYN
Barcelona Computational, Cognitive and Systems Neuroscience Community

Neural network mechanisms underlying post-decision biases

Miguel Donderis[1] Jose M. Esnaola-Acebes[2] Alex Roxin[2] Klaus Wimmer[2]

[1]Universitat Autònoma de Barcelona [2]Centre de Recerca Matemàtica

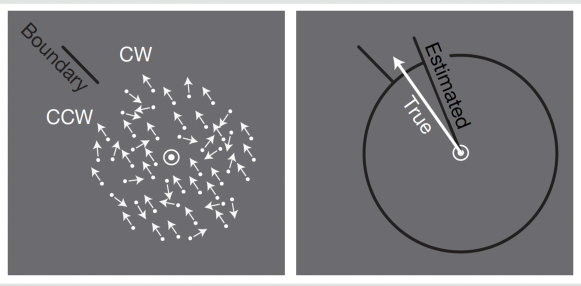
CRM CENTRE DE RECERCA MATEMÀTICA

Introduction. Post-decision biases in humans

Perception is influenced by past choices. In motion discrimination tasks, a **categorical choice biases perceptual reports** about the direction of motion away from the decision boundary. Although explained using neural encoding-decoding models ([1.]) and Bayesian principles, it remains unknown what **neural network mechanisms** give rise to these **post-decision biases**.

Here, we develop a neural network model with the aim of studying the **integration** of spatially modulated inputs in a **bump attractor network**.

Combined discrimination-estimation task.



Perceptual bias in Jazayeri-Movshon experiment

- **Fine discrimination**: Subjects viewed a **field of moving dots** within a circular aperture and reported whether the direction of motion was clockwise (CW) or counterclockwise (CCW).
- **Continuous report**: When subjects were asked to report the direction of motion, their **estimates** were **biased** in register with their discrimination choice.

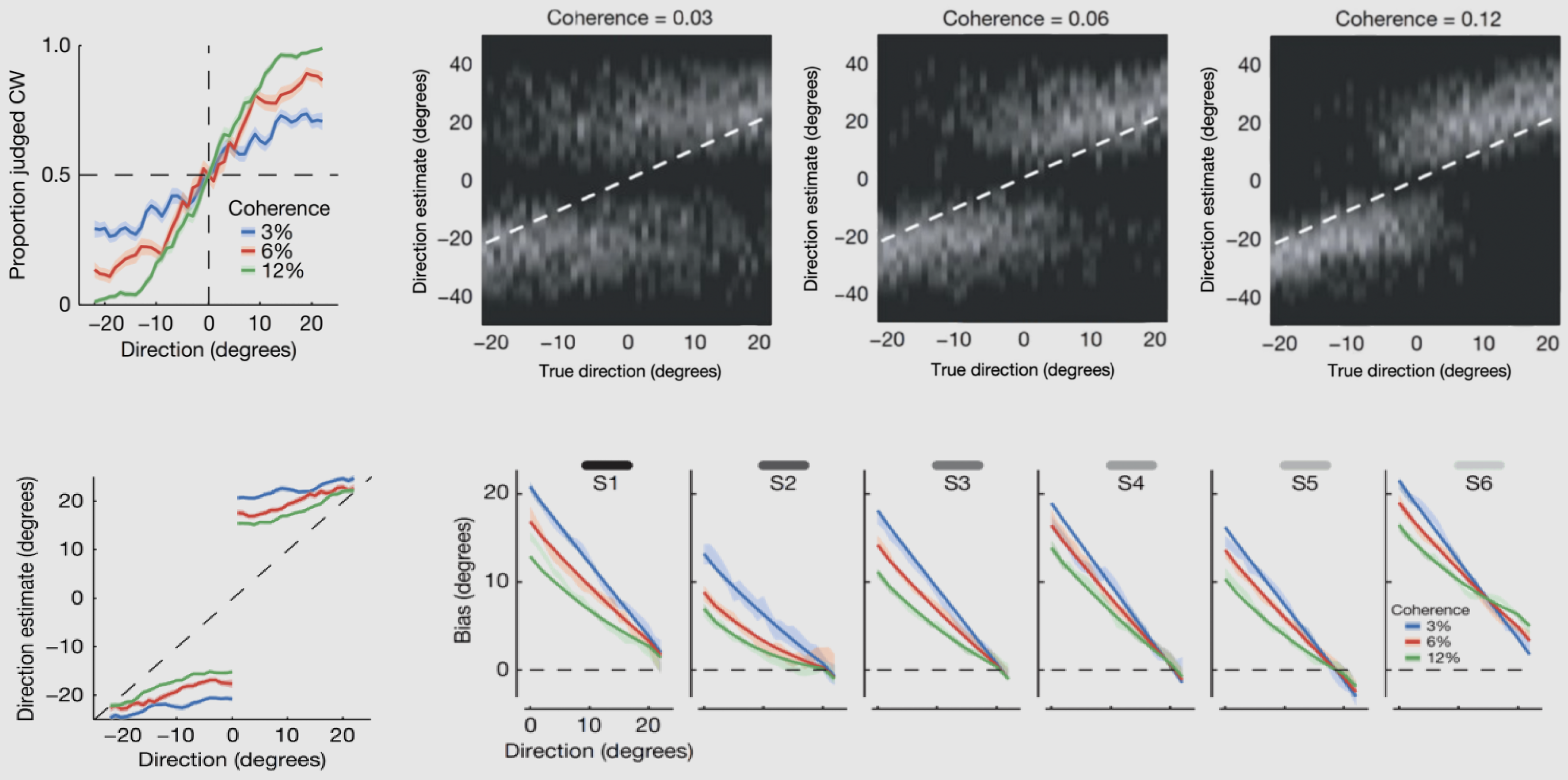
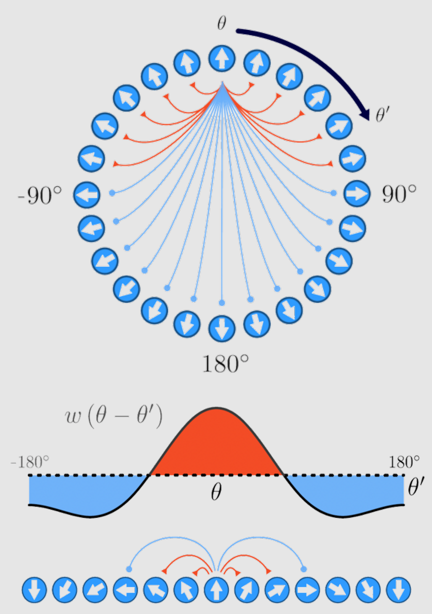


Figure 1. Estimation bias for six different subjects

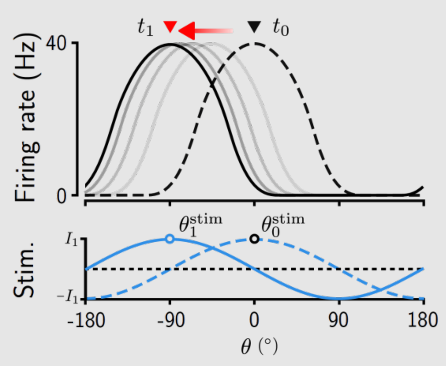
The bump attractor model

Ring network. Population of neurons arranged in a ring structure, $\theta \in [-\pi, \pi)$



Neural field description [2.]

$$\tau\frac{\partial r}{\partial t} = -r + \Phi\left(\frac{\tau}{2\pi}\int_{-\pi}^{\pi}\omega(\theta-\theta')r(\theta',t)d\theta' + I_{\text{exc}} + I_{\text{stim}}(\theta,t) + \xi(\theta,t)\right)$$



- I_{exc} : global net **excitatory drive**
- $\omega(\theta)$: connectivity function
- $r(\theta,t)$: firing rate of the network
- τ : neural time constant
- Φ : transfer function
- Stimulus: **spatially modulated** input $I_{\text{stim}}(\theta,t) = I_1\cos(\theta - \theta^{\text{stim}}(t))$

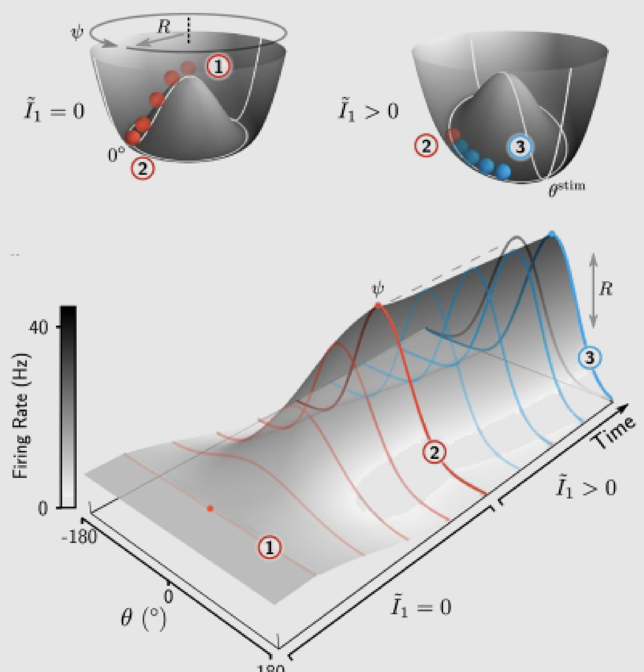
A persistent bump of activity emerges at a position determined by the stimulus.
For time-varying inputs, the bump moves towards the new position.

Neural mechanisms underlying stimulus integration

The dynamics of the bump close to the Turing bifurcation are described by the amplitude equation ([3.])

$$\tau\frac{\partial R}{\partial t} = \tilde{I}_1\cos(\psi - \theta^{\text{stim}}) + \tilde{I}_0 R - cR^3 + \xi_1(t)$$
$$\tau\frac{\partial \psi}{\partial t} = -\frac{\tilde{I}_1}{R}\sin(\psi - \theta^{\text{stim}}) + \frac{\xi_2(t)}{R}$$

- $R(t)$: amplitude of the bump
- $\psi(t)$: phase of the bump



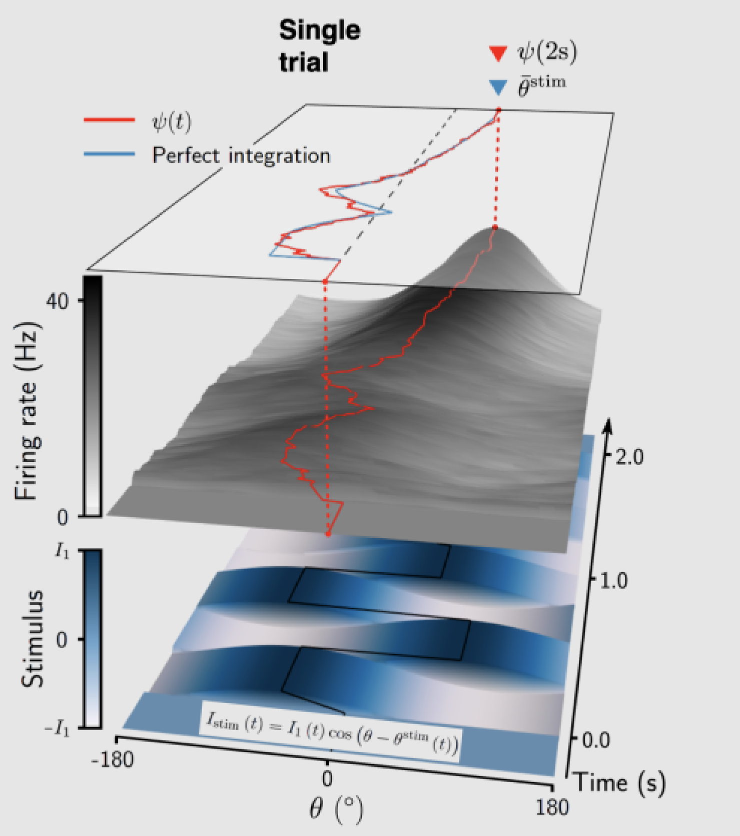
The movement of the phase is inversely proportional to the amplitude of the bump

$$Rd\psi \propto \tilde{I}_1 \implies d\psi \propto \frac{\tilde{I}_1}{\tilde{I}_0^{\frac{1}{2}}}$$

Intuitive representation: motion of a particle in a potential of the form

$$\Theta(R,\psi) = -R\tilde{I}_1\cos(\psi - \theta^{\text{stim}}) - \frac{\tilde{I}_0}{2}R^2 + c\frac{R^4}{4}$$

The phase of the bump as the estimate of the stimulus direction



- $\tilde{I}_0$ determines the depth of the potential and therefore the amplitude of the bump ($R \propto \tilde{I}_0^{\frac{1}{2}}$).
- $\tilde{I}_1$ brakes the radial simmetry of the potential and forces the movement of the particle towards the location of the stimulus θ^{stim}.

Post-decision bias in the model for a $\cos(2(\theta - \frac{\pi}{2}))$ modulation signal

To force a categorical decision, we introduce a spatially structured inputs (i.e. attention or urgency signals) after the integration phase.

We first consider an input $I = I_0 + 2I_2\cos(2(\theta - \theta^d))$, for which the dynamics of the amplitude and phase of the bump are

$$\tau\dot{R} = I_0R + I_2R\cos(2(\psi-\theta^d)) - cR^3,$$
$$\tau\dot{\psi} = -I_2\sin(2(\psi-\theta^d)).$$

For the case $\theta^d = \frac{\pi}{2}$, the dynamical system has two **stable** solutions at $\psi = \pm\frac{\pi}{2}$, and two **unstable** solutions in between, at $\psi = 0, \frac{3\pi}{2}$, so the bump will be attracted to the **nearest** stable solution.

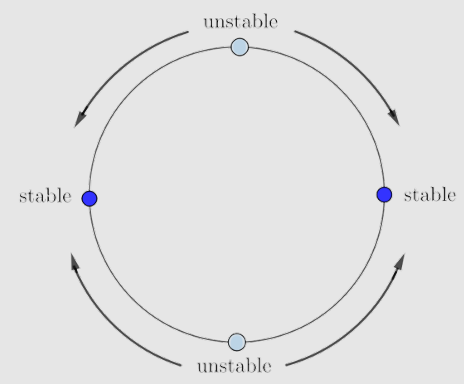
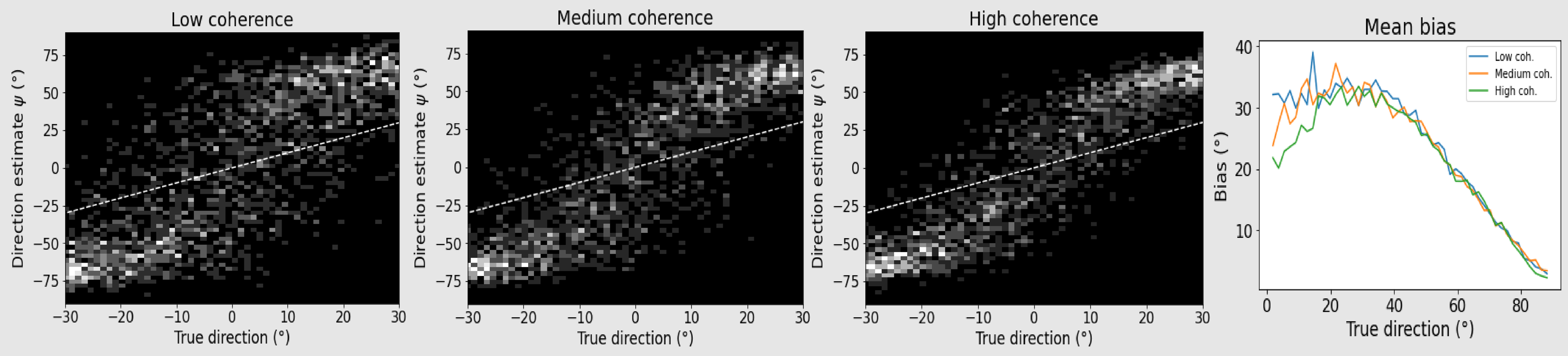


Figure 2. Stability diagram for the case of $\cos(2(\theta - \frac{\pi}{2}))$.



- The modulation signal leads to a repulsive bias as observed in the experiment.
- However, this particular modulation signal cannot explain the bias curve quantitatively.

Post-decision bias for higher Fourier modes

To investigate whether our model can provide a more accurate fit to the psychophysical data, we now consider modulated inputs of the form $I = I_0 + 2I_k\cos(k(\theta - \theta^d))$, i.e. different spatial Fourier modes.

We find that the **spatial structure** of the input determines the **fixed points** governing the temporal evolution of the phase of the bump: an input of Fourier mode k leads to k stable (and unstable) fixed points.

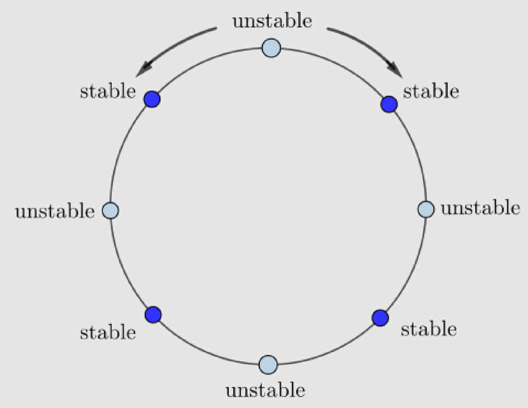
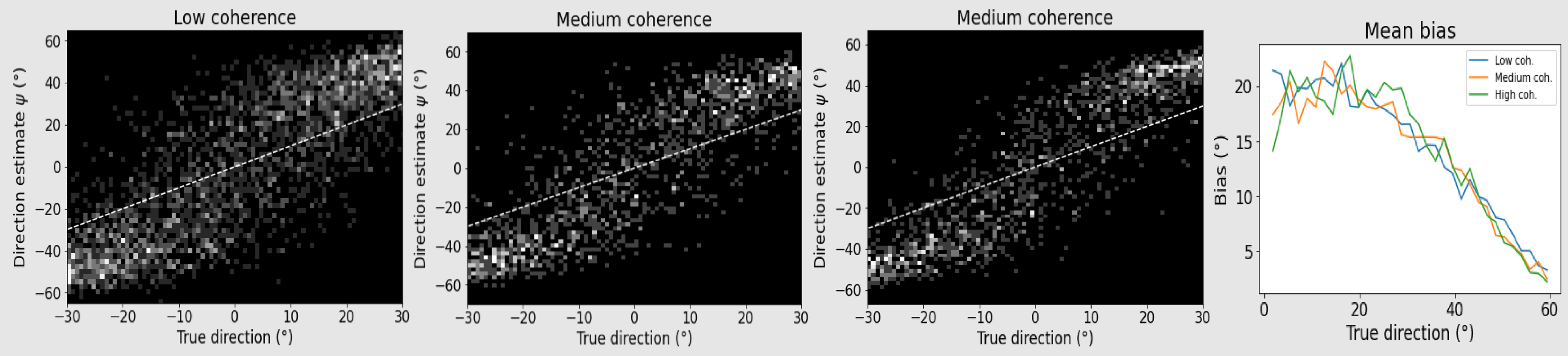


Figure 3. Stability diagram for $k = 4$.

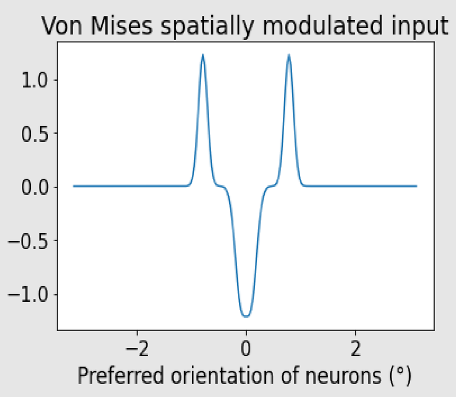


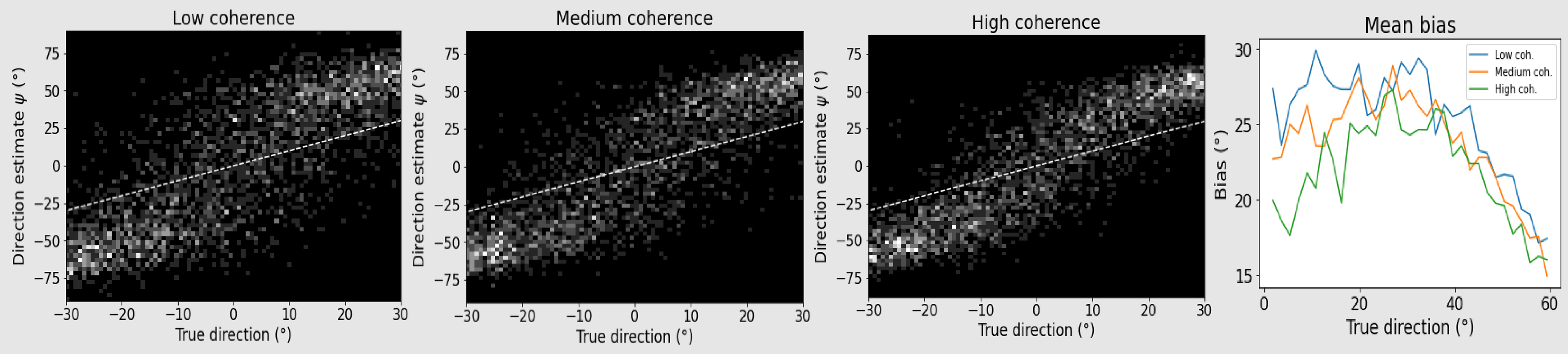
We obtain an estimation bias that is largest close to the decision boundary and reaches zero at the fixed points.

Choice commitment signal with excitation and inhibition

Repulsive bias effects can also be obtained through a combination of positive (excitatory) and negative (inhibitory) inputs.

As an example, similarly to the previous cosinus modulation signals, we used a modulation signal with two narrow positive peaks at $\pm\frac{\pi}{4}$ and a negative peak at 0 (realized as sum of von Mises functions).





Summary

- Ring attractor dynamics provides a potential mechanism for both stimulus integration and perceptual categorization.
- The phase of the activity bump tracks the running circular average of the stimulus orientation.
- Including a spatially-modulated choice-commitment signal forces the network to a categorical choice.
- This explains the emergence of a repulsive post-decision estimation bias, with a bias curve determined by the shape of the modulation signal.

References

1. M. Jazayeri, J.A. Movshon (2007), *A new perceptual illusion reveals mechanisms of sensory decoding*, Nature.
2. R. Ben-Yishai, R. Lev Bar-Or, H. Sompolinsky (1995), *Theory of orientation tuning in visual cortex*, Proc. Natl. Acad. Sci.
3. J.M Esnaola-Acebes, A. Roxin, K. Wimmer (2021), *Bump attractor dynamics underlying stimulus integration in perceptual estimation tasks*, Biorχiv

 miguel.donderis@autonoma.cat