Efficient coding provides a direct link between prior and likelihood in perceptual Bayesian inference

Xue-Xin Wei and Alan A. Stocker*
Departments of Psychology and
Electrical and Systems Engineering
University of Pennsylvania
Philadelphia, PA-19104, U.S.A.

ARTICLES

nature neuroscience

A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts

Xue-Xin Wei[1] & Alan A Stocker[1,2]

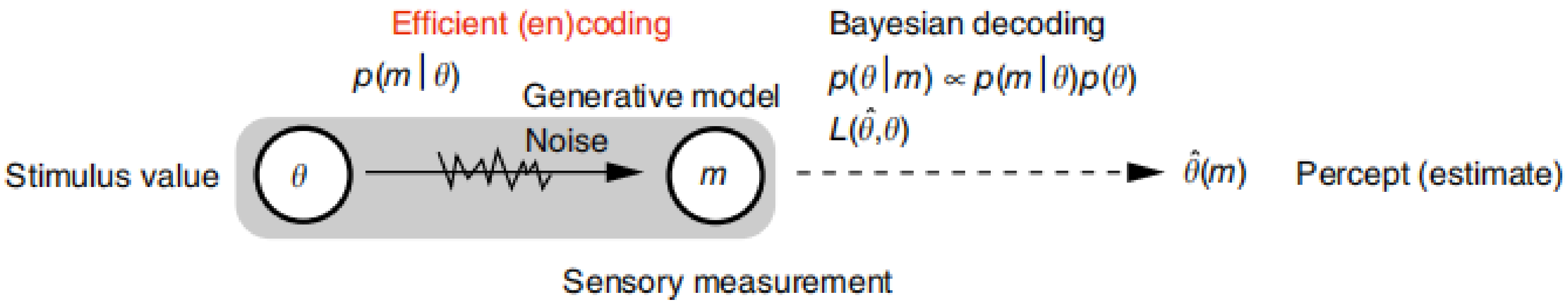
Journal Club

Li, Ang

2019/3/22

Key words

- Efficient coding
- Bayesian inference
- Perceptual bias

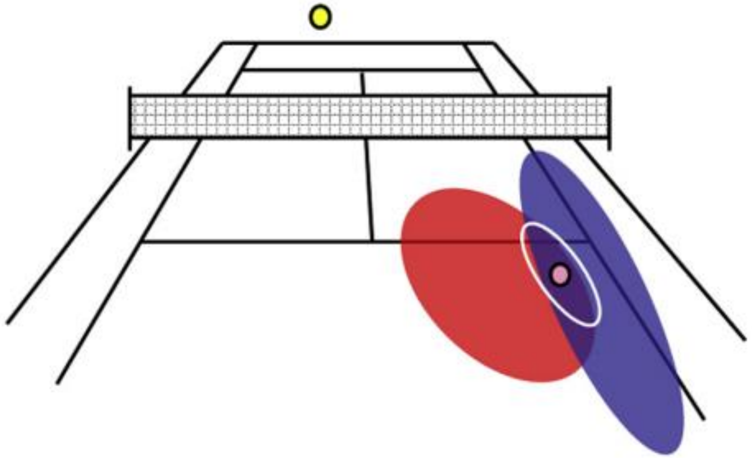


Backgrounds: Efficient coding hypothesis

- neural resource limitations lead to efficient sensory representations that are optimized with regard to the specific stimulus statistics of the natural environment

- For example
  - Max Mutual information (between stimulus and neural response)
  - s.t. limited resource

Backgrounds: Bayesian inference



- Bayes' theorem:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

- Bayesian inference:

likelihood

$$P(A \mid B) \propto P(B \mid A)P(A)$$

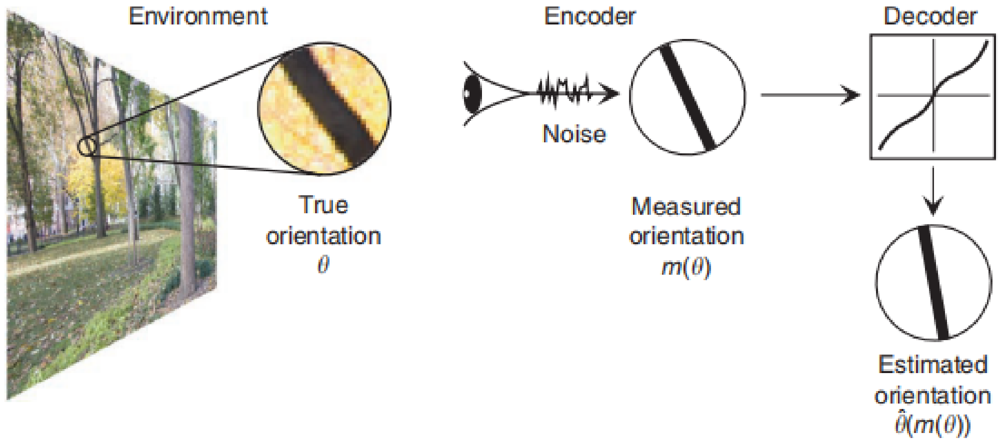
prior (belief)

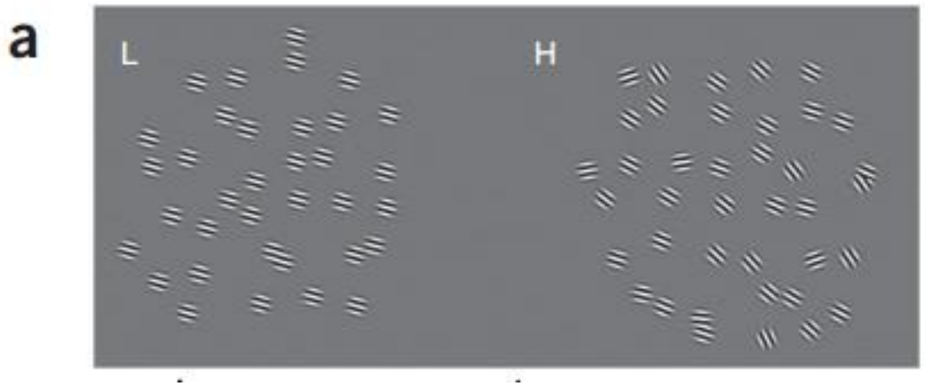
- Bayesian estimate: posterior mean

$$\hat{\theta}(x) = E[\theta|x] = \int \theta\, p(\theta|x)\, d\theta.$$

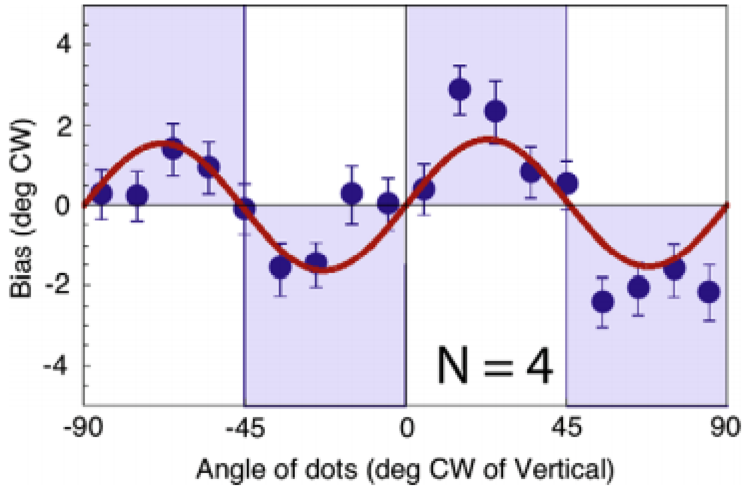
Backgrounds: Perceptual bias

- Systematic errors in perceiving task.



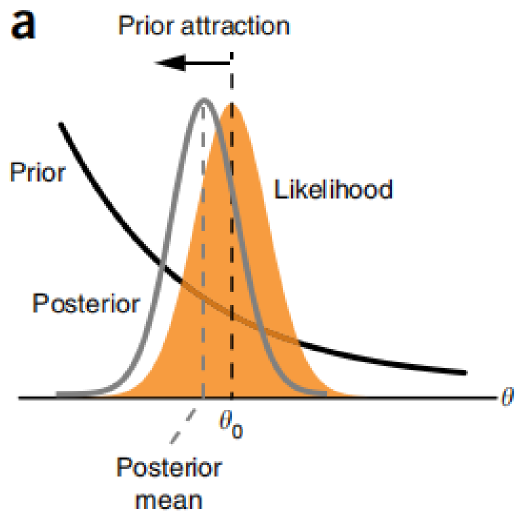


Experiment

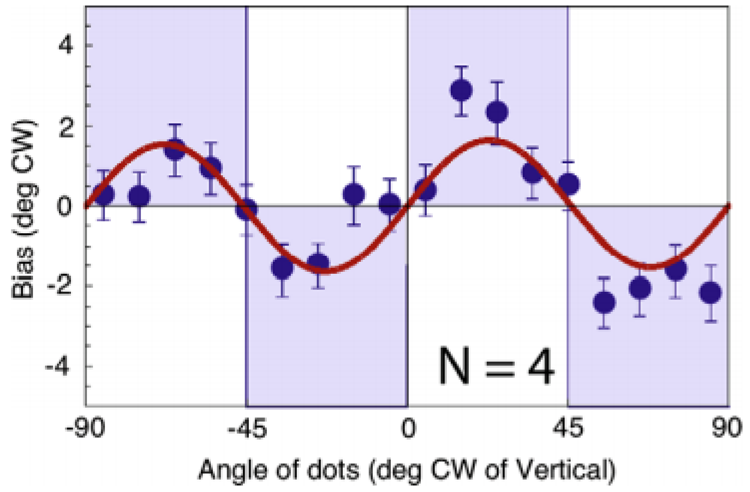


Tomassini et al 2010

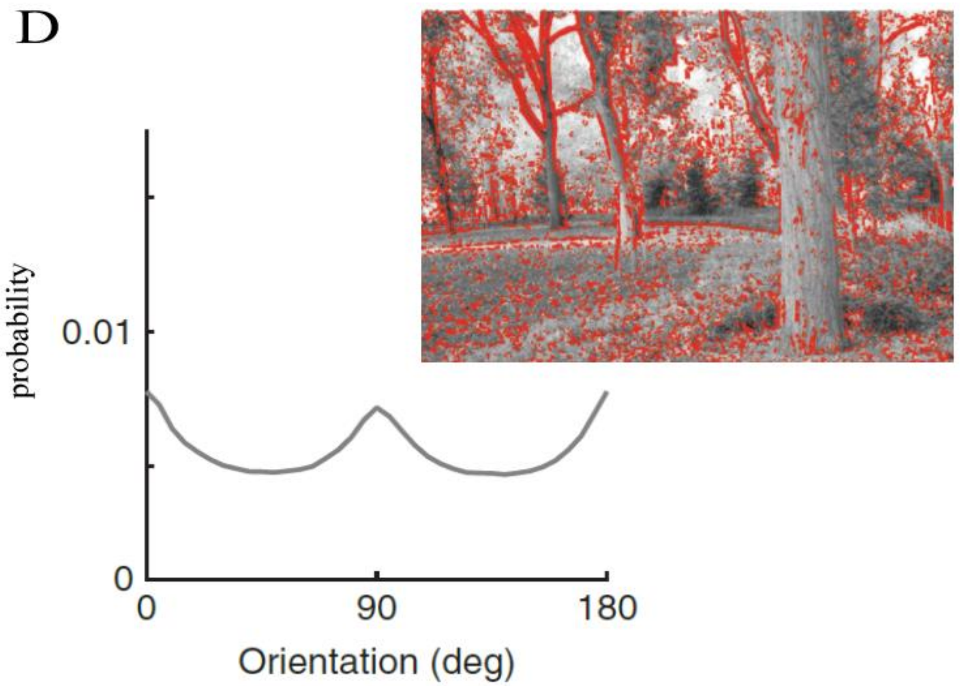
Backgrounds: `Anti-Bayesian` percepts



- Bayesian framework: Prior attraction
- Facts: Prior repulsion



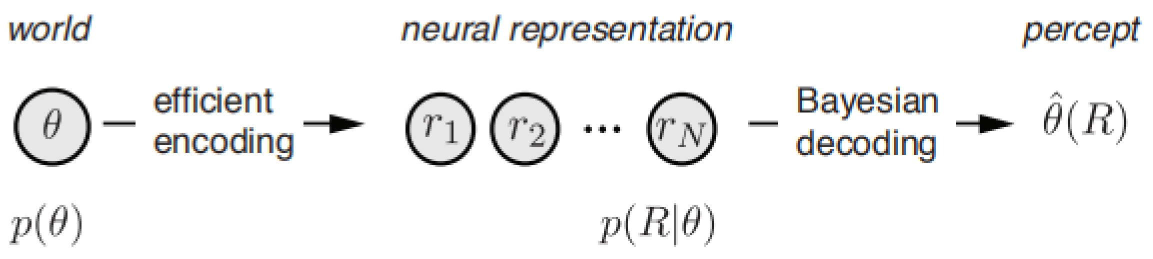
Tomassini et al 2010



Girshick et al 2011

Efficient coding provides a direct link between prior and likelihood in perceptual Bayesian inference

Xue-Xin Wei and Alan A. Stocker*
Departments of Psychology and
Electrical and Systems Engineering
University of Pennsylvania
Philadelphia, PA-19104, U.S.A.



Efficient encoding

- Assumption:
- an efficient coding constraint that maximizes the mutual information between a scalar stimulus variable θ and its sensory representation m

- Quick conclusion:

$$p(\theta) \propto \sqrt{J(\theta)}$$

prior fisher information

Details

Efficient encoding. We assumed an efficient coding constraint that maximizes the mutual information between a scalar stimulus variable θ and its sensory representation m (refs. 2,21). Fisher information $J(\theta)$ defined as

Fisher Information $$J(\theta) = \int \left(\frac{\partial \ln p(m|\theta)}{\partial \theta} \right)^2 p(m|\theta) dm \quad (2)$$

can be used to specify a bound on mutual information in the asymptotic limit of vanishing noise[22]. Assuming the bound is tight, mutual information can be expressed as[23]

Lower bound of Mutual information $$I[\theta, m] = \frac{1}{2} \ln \left(\frac{S^2}{2\pi e} \right) - KL \left(p(\theta) \,||\, \frac{\sqrt{J(\theta)}}{S} \right) \quad (3)$$

where $S = \int_\theta \sqrt{J(\theta)} d\theta$. S can be intuitively understood as the total amount of coding resource available. $KL(||)$ represents the Kullback-Leibler (KL) divergence[51], which is always non-negative.

The goal is to choose $J(\theta)$ to maximize $I[\theta, m]$ for a fixed $p(\theta)$. Technically, this requires us to impose an additional constraint because $I[\theta, m]$ is not limited otherwise. To bound $I[\theta, m]$ and create a well-posed optimization problem, we assume

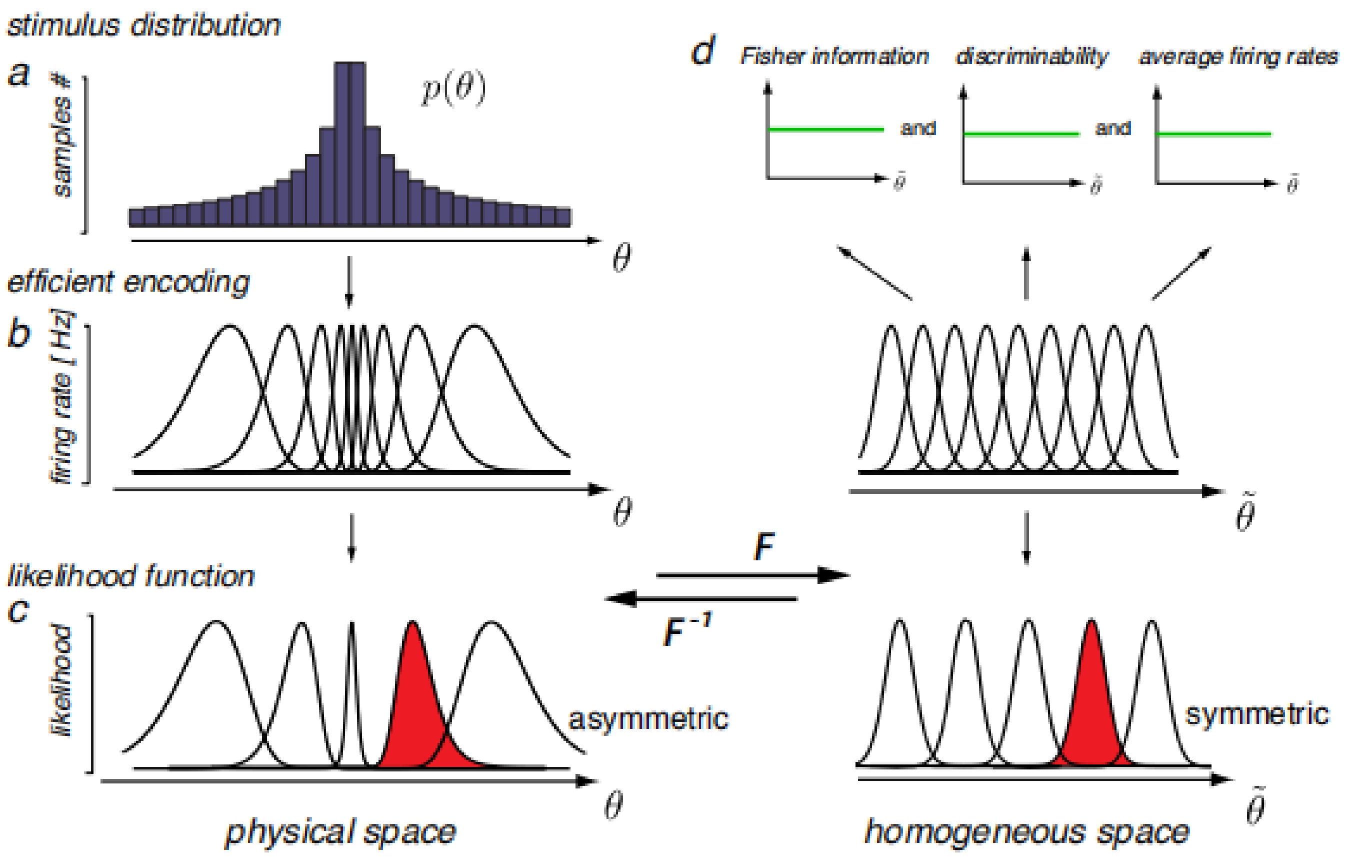
Addition assumption $$S = \int_\theta \sqrt{J(\theta)} d\theta \le C \quad (4)$$

With this constraint, maximizing mutual information requires the above KL divergence term to be zero. This is equivalent to

Conclusion $$p(\theta) \propto \sqrt{J(\theta)} \quad (5)$$

Because the mutual information $I[\theta, m]$ is invariant with respect to any re-parameterization of θ, it is desirable that the constraint also shares this property. The chosen constraint (equation (4)) is invariant whereas constraints using other power functions of $J(\theta)$, for example, $\int_\theta J(\theta) d\theta$, are not.

Fisher information & likelihood function



Key: Asymmetric likelihood function

The mapping function: F(theta)

In order to constrain the tuning curves of individual neurons in the population we also impose a homogeneity constraint, requiring that there exists a one-to-one mapping $F(\theta)$ that transforms the physical space with units θ to a *homogeneous space* with units $\tilde{\theta} = F(\theta)$ in which the stimulus distribution becomes uniform. This defines the mapping as

$$F(\theta) = \int_{-\infty}^{\theta} p(\chi) d\chi \ , \tag{1}$$

which is the cumulative of the prior distribution $p(\theta)$. We then assume a neural population with identical tuning curves that evenly tiles the stimulus range in this homogeneous space. The population provides an efficient representation of the sensory variable θ according to the above constraints [11]. The tuning curves in the physical space are obtained by applying the inverse mapping $F^{-1}(\tilde{\theta})$. Fig. 2

Bayesian decoding

Let us consider a population of N sensory neurons that efficiently represents a stimulus variable θ as described above. A stimulus θ_0 elicits a specific population response that is characterized by the vector $R = [r_1, r_2, ..., r_N]$ where r_i is the spike-count of the i_{th} neuron over a given time-window τ. Under the assumption that the variability in the individual firing rates is governed by a Poisson process, we can write the likelihood function over θ as

$$p(R|\theta) = \prod_{i=1}^{N} \frac{(\tau f_i(\theta))^{r_i}}{r_i!} e^{-\tau f_i(\theta)} , \qquad (2)$$

Likelihood function

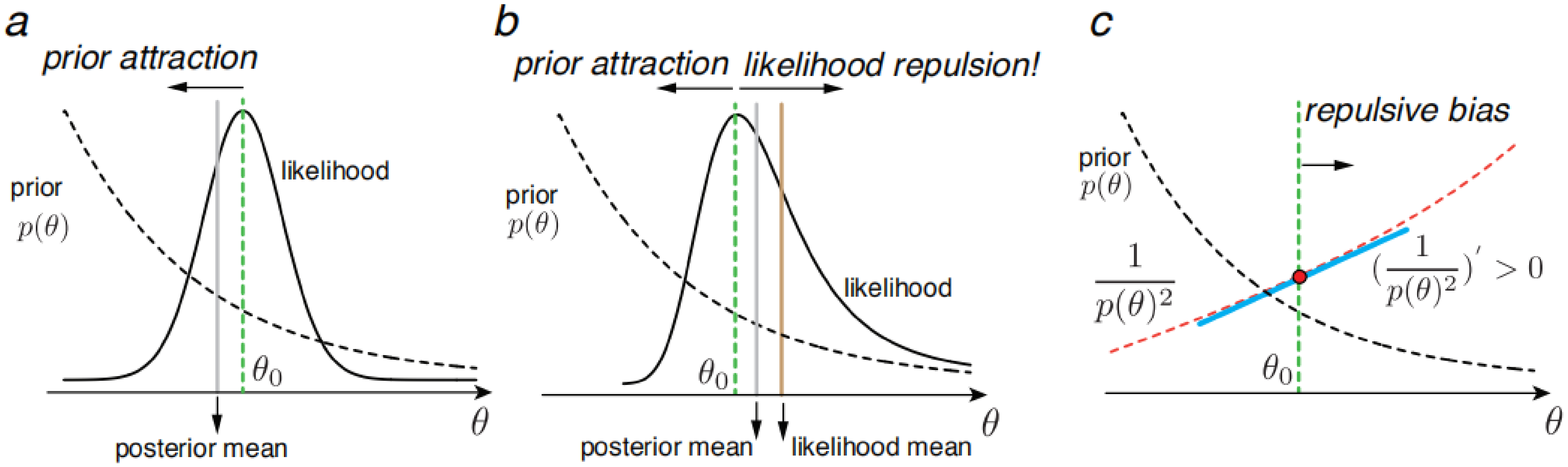
with $f_i(\theta)$ describing the tuning curve of neuron i. We then define a Bayesian decoder $\hat{\theta}_{\mathrm{LSE}}$ as the estimator that minimizes the expected squared-error between the estimate and the true stimulus value, thus

$$\hat{\theta}_{\mathrm{LSE}}(R) = \frac{\int \theta p(R|\theta)p(\theta)d\theta}{\int p(R|\theta)p(\theta)d\theta} , \qquad (3)$$

Bayesian estimator

where we use Bayes' rule to appropriately combine the sensory evidence with the stimulus prior $p(\theta)$.

Bayesian estimates can be biased away from prior peaks



Symmetric likelihood function

Asymmetric likelihood function

Estimation bias

$$b(\theta_0) \approx C(\frac{1}{p(\theta_0)^2})' ,$$

Perception of visual orientation

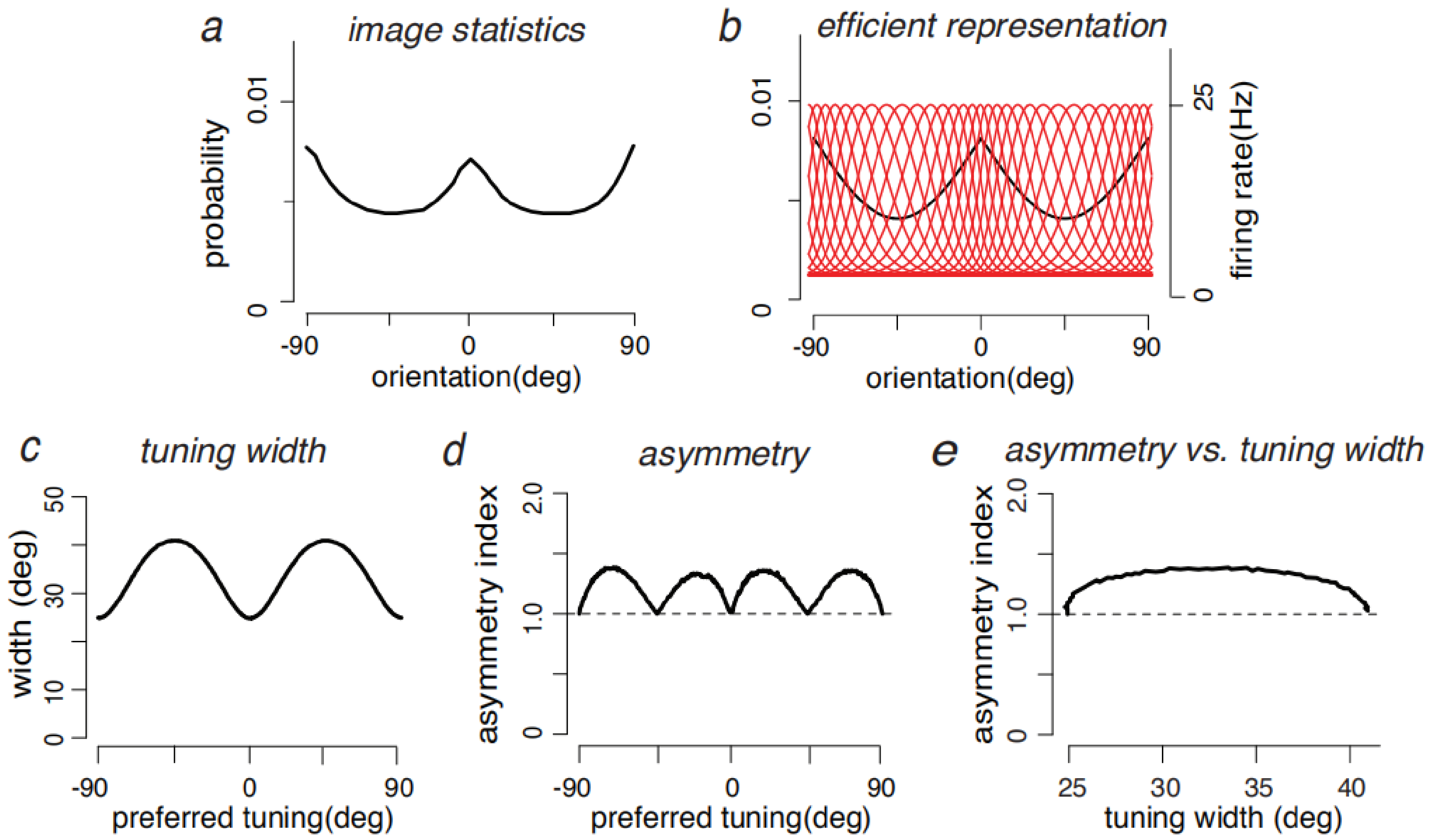
4.1 Efficient neural model population for visual orientation

Previous studies measured the statistics of the local orientation in large sets of natural images and consistently found that the orientation distribution is multimodal, peaking at the two cardinal orientations as shown in Fig. 4a [16, 20]. We assumed that the visual system's prior belief over orientation $p(\theta)$ follows this distribution and approximate it formally as

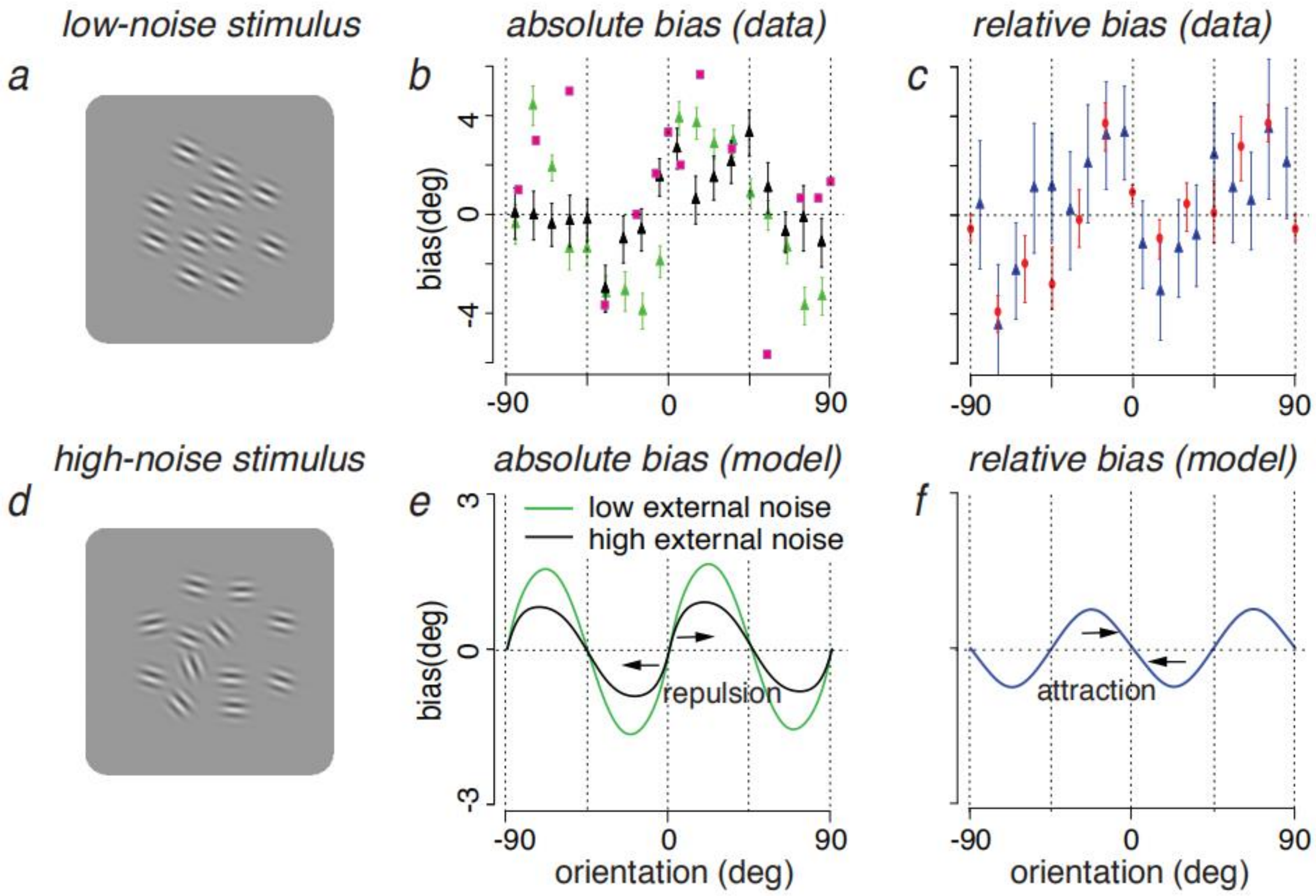
$$p(\theta) \propto 2 - |\sin(\theta)| \text{ (black line in Fig. 4b) .} \tag{7}$$

Based on this prior distribution we defined an efficient neural representation for orientation. We assumed a population of model neurons ($N = 30$) with tuning curves that follow a von-Mises distribution in the homogeneous space on top of a constant spontaneous firing rate (5 Hz). We then applied the inverse transformation $F^{-1}(\tilde{\theta})$ to all these tuning curves to get the corresponding tuning curves in the physical space (Fig. 4b - red curves), where $F(\theta)$ is the cumulative of the prior (7). The concentration parameter for the von-Mises tuning curves was set to $\kappa \approx 1.6$ in the homogeneous space in order to match the measured average tuning width (~ 32 deg) of neurons in area V1 of the macaque [9].

Tuning characteristics of model neurons



Bias in perceived orientation



Takeaways

- A framework for perception that combines efficient coding & Bayesian decoding.
- Efficient coding imposes tuning characteristics of a population of neurons according to the prior.
- The resulted likelihoods are in general asymmetric, with heavier tails away from prior peaks, can explain perceptual bias away from prior.

Discussion & question

- How to understand the mapping F(theta)
- Other efficient coding objectives?
- Can the model be applied to other modalities?
- Other `anti-Bayesian` percepts?

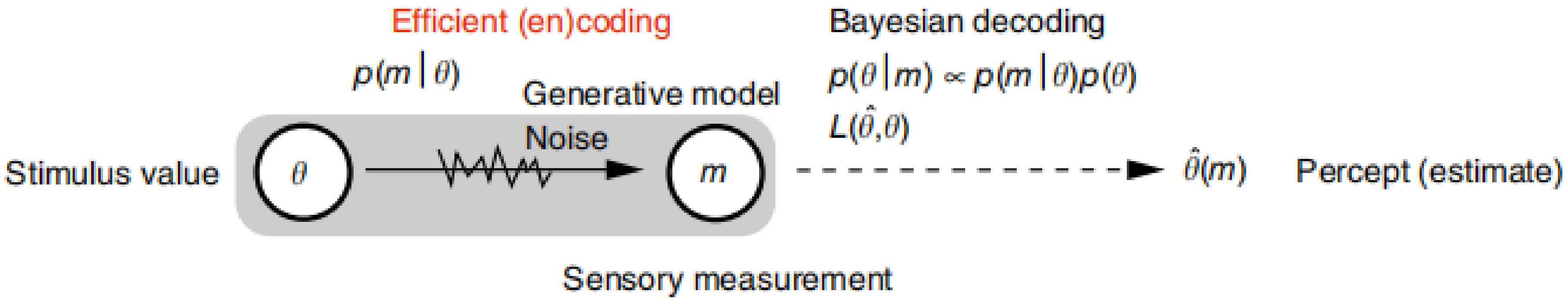
Follow-up

ARTICLES

nature neuroscience

A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts

Xue-Xin Wei[1] & Alan A Stocker[1,2]

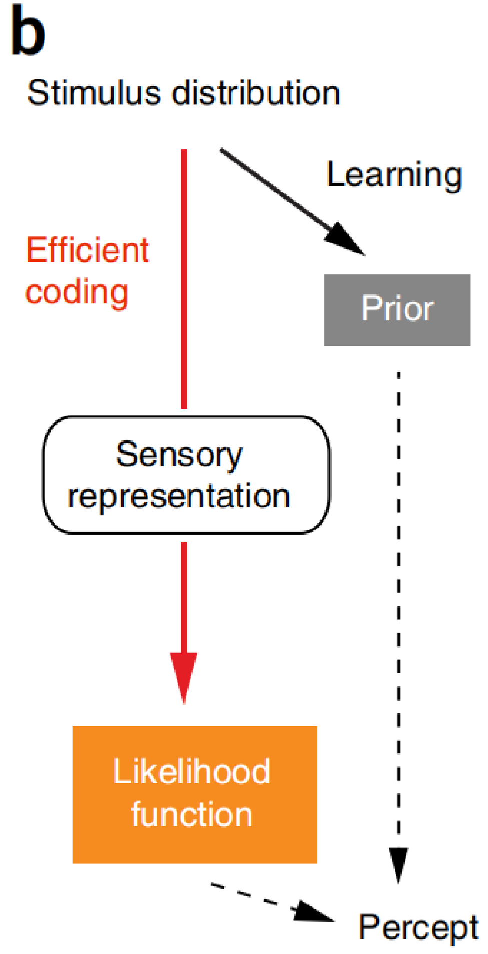


Framework

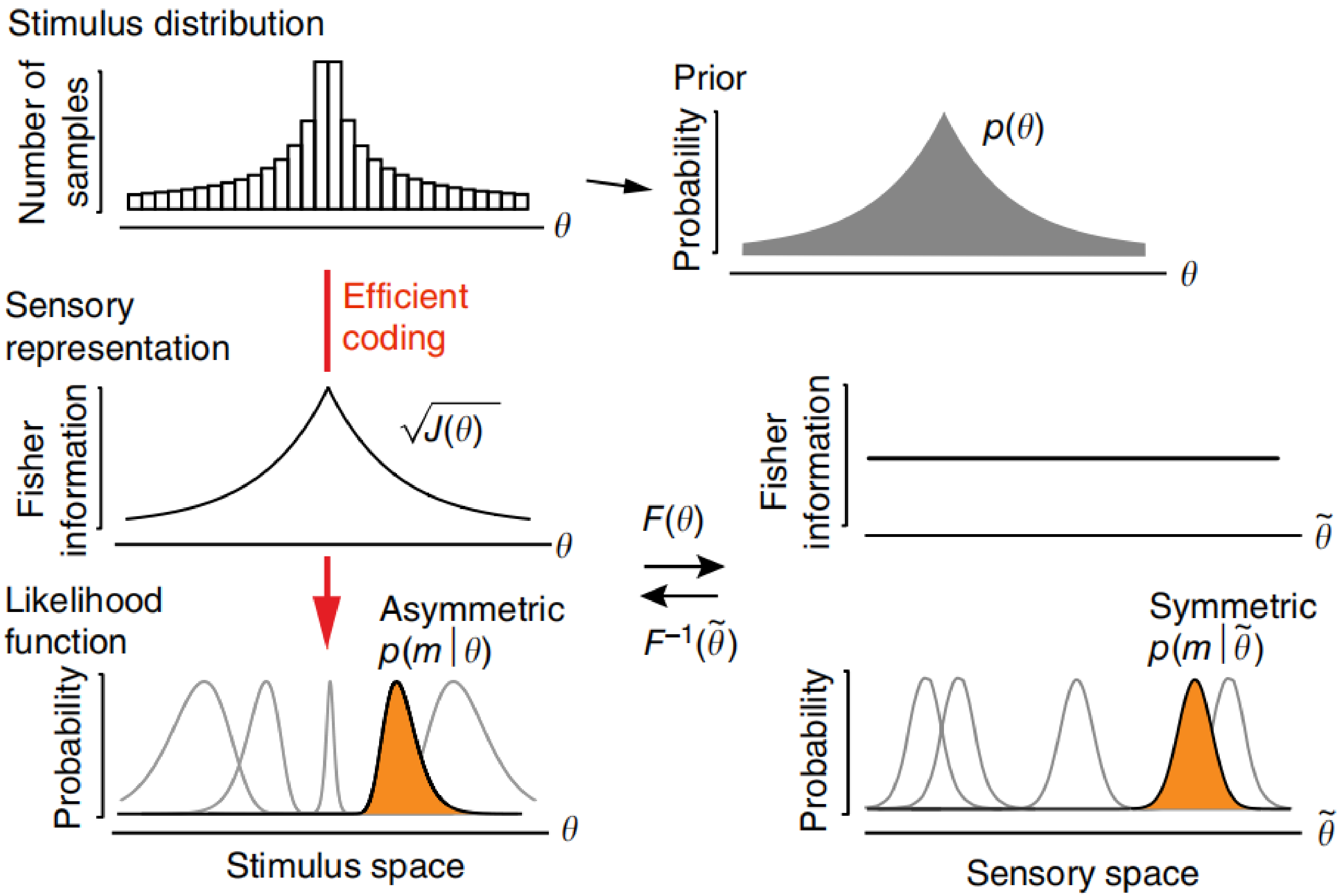
- Efficient coding:
  - Sensory representation (likelihood) is constrained by the natural stimulus distribution: the prior

- Bayesian decoding:

$$\hat{\theta}_{\text{LSE}}(R) = \frac{\int \theta p(R|\theta)p(\theta)d\theta}{\int p(R|\theta)p(\theta)d\theta},$$



Framework



Stimulus distribution
Number of samples
θ
Prior
Probability
p(θ)
θ
Sensory representation
Efficient coding
Fisher information
√J(θ)
θ
F(θ)
F⁻¹(θ̃)
Fisher information
θ̃
Likelihood function
Asymmetric
p(m | θ)
Probability
θ
Stimulus space
Symmetric
p(m | θ̃)
Probability
θ̃
Sensory space

The mapping function: F(theta)F(theta)

Although Fisher information constrains the likelihood function, it is not sufficient to fully specify its shape. An additional assumption about the noise structure is required. Let us consider a function $F(\theta)$ that maps the stimulus space to a new space in which Fisher information is uniform (**Fig. 1c**). We refer to this space as the 'sensory space' in reference to Gustav Fechner because discriminability, when measured in units of this space, is uniform[25]. The mapping $F(\theta)$ is defined as the cumulative of the stimulus distribution (equation (8), Online Methods). Uniform Fisher information implies that the noise and thus the likelihood function is homogeneous in this space. We make the additional assumption that the noise is such that the expected likelihood function (that is, averaged out over many trials) is symmetric around the stimulus value in the sensory space. Additive and symmetric noise is the simplest condition for which this assumption is true (for example, Gaussian as illustrated in **Fig. 1c**). Although the assumption seems parsimonious given the homogeneity of the space, the degree to which it is a valid assumption for real neural populations is unclear. In simulations of reasonably realistic neural population models, we found the assumption to approximately hold under a fairly large range of conditions (see Discussion).

With the likelihood defined in the sensory space, the likelihood function in the stimulus space can then be obtained by simply applying the inverse mapping $F^{-1}(\tilde{\theta})$. As a result, the likelihood functions when formulated in stimulus space are typically asymmetric, with a long tail away from the peak of the prior distribution.

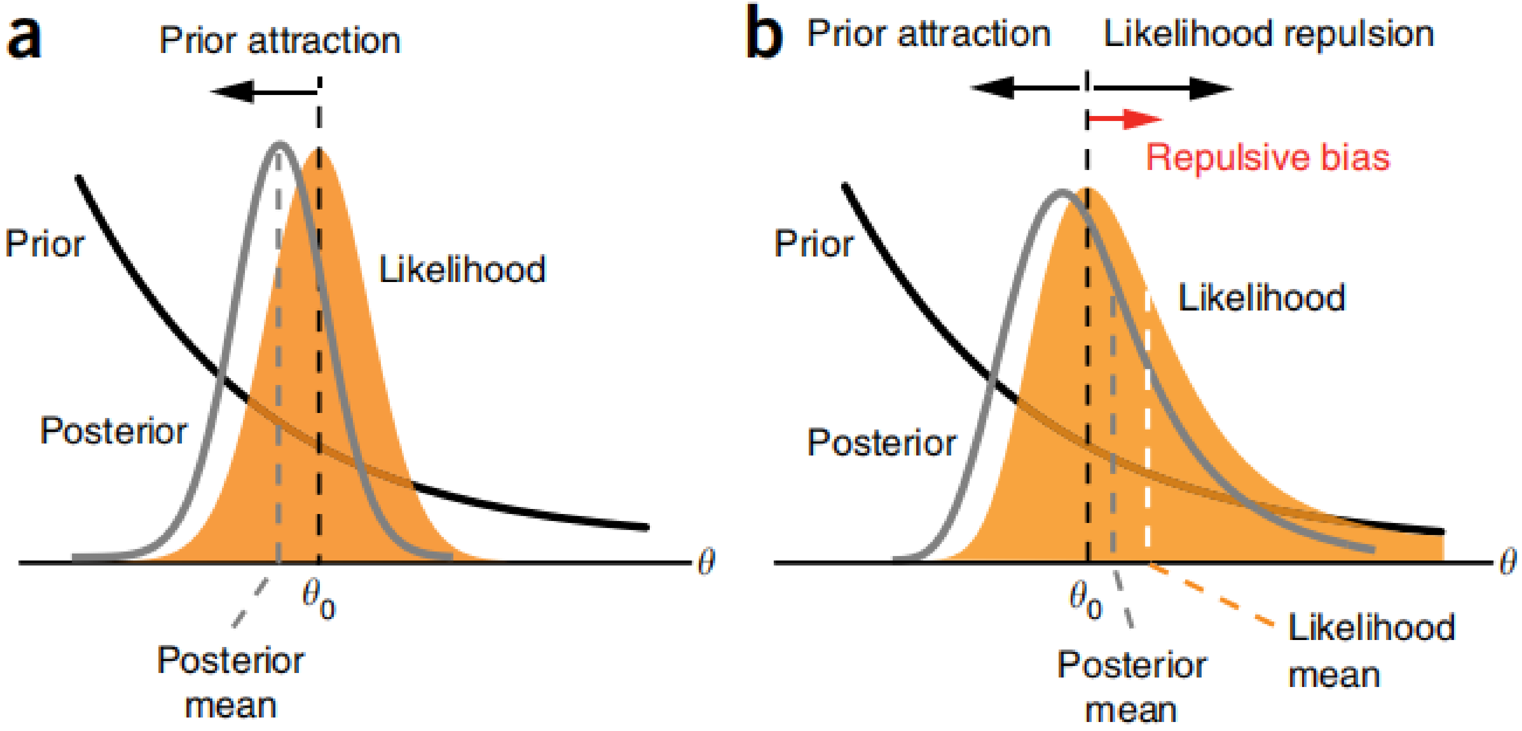
With the efficient coding assumption above, equation (5), we can now express the bias as a function of the prior belief. First, we define a one-to-one mapping $F(\theta)$ that transforms the stimulus space to a *sensory space* with units $\tilde{\theta} = F(\theta)$ for which the Fisher information (as well as the stimulus distribution) is uniform[25,42]. The mapping is defined as

$$F(\theta) = \int_{-\infty}^{\theta} p(\chi)\, d\chi \qquad (8)$$

which is the cumulative of the prior distribution $p(\theta)$.

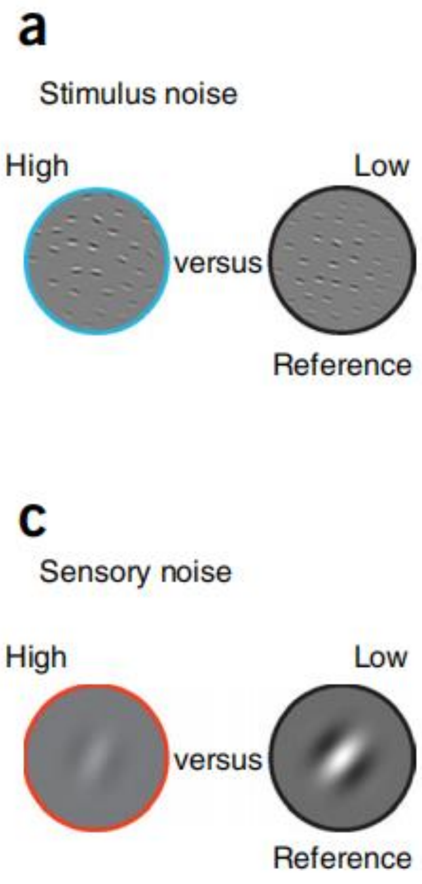
Prediction 1:

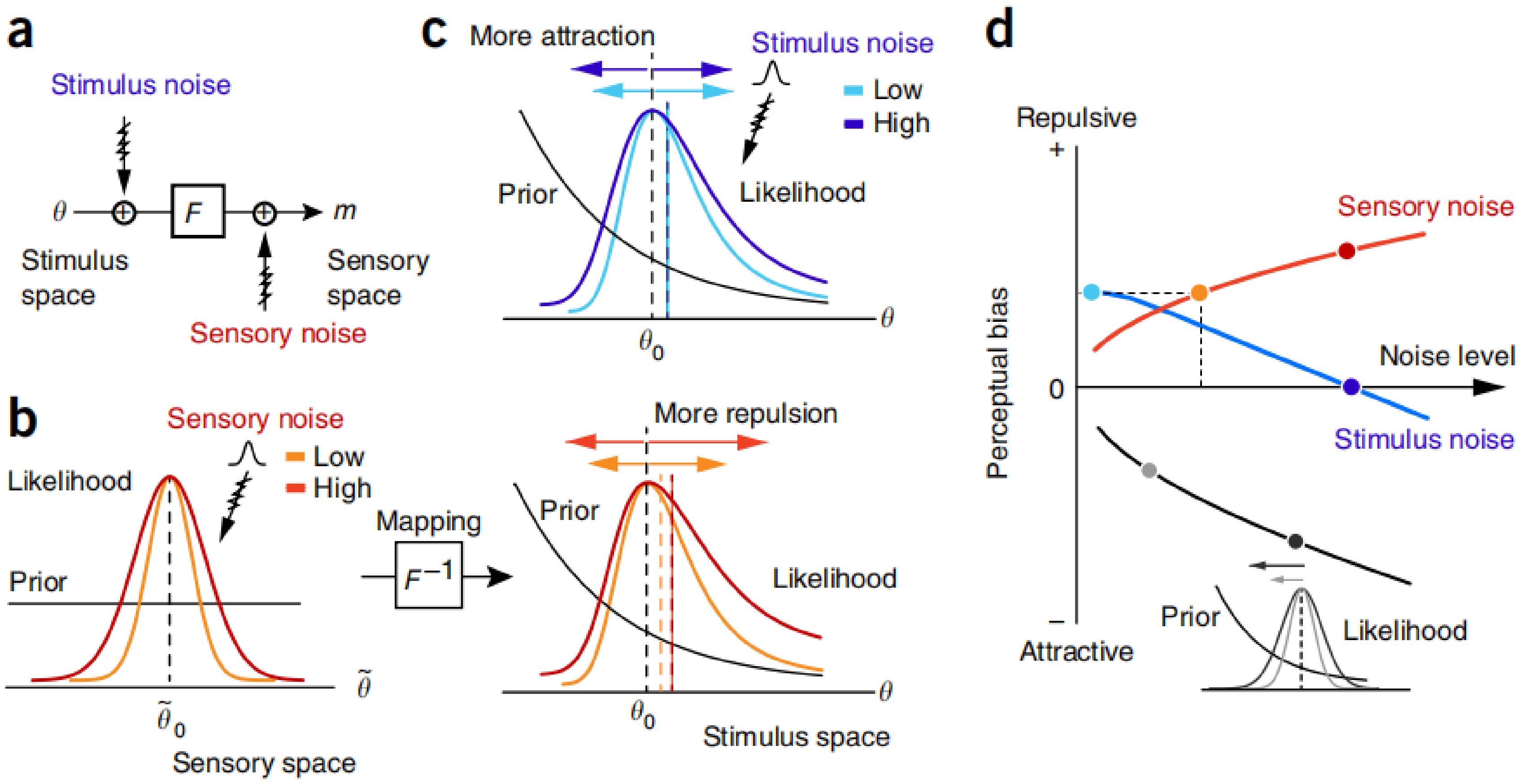
- Bayesian perception can be biased away from the prior peak



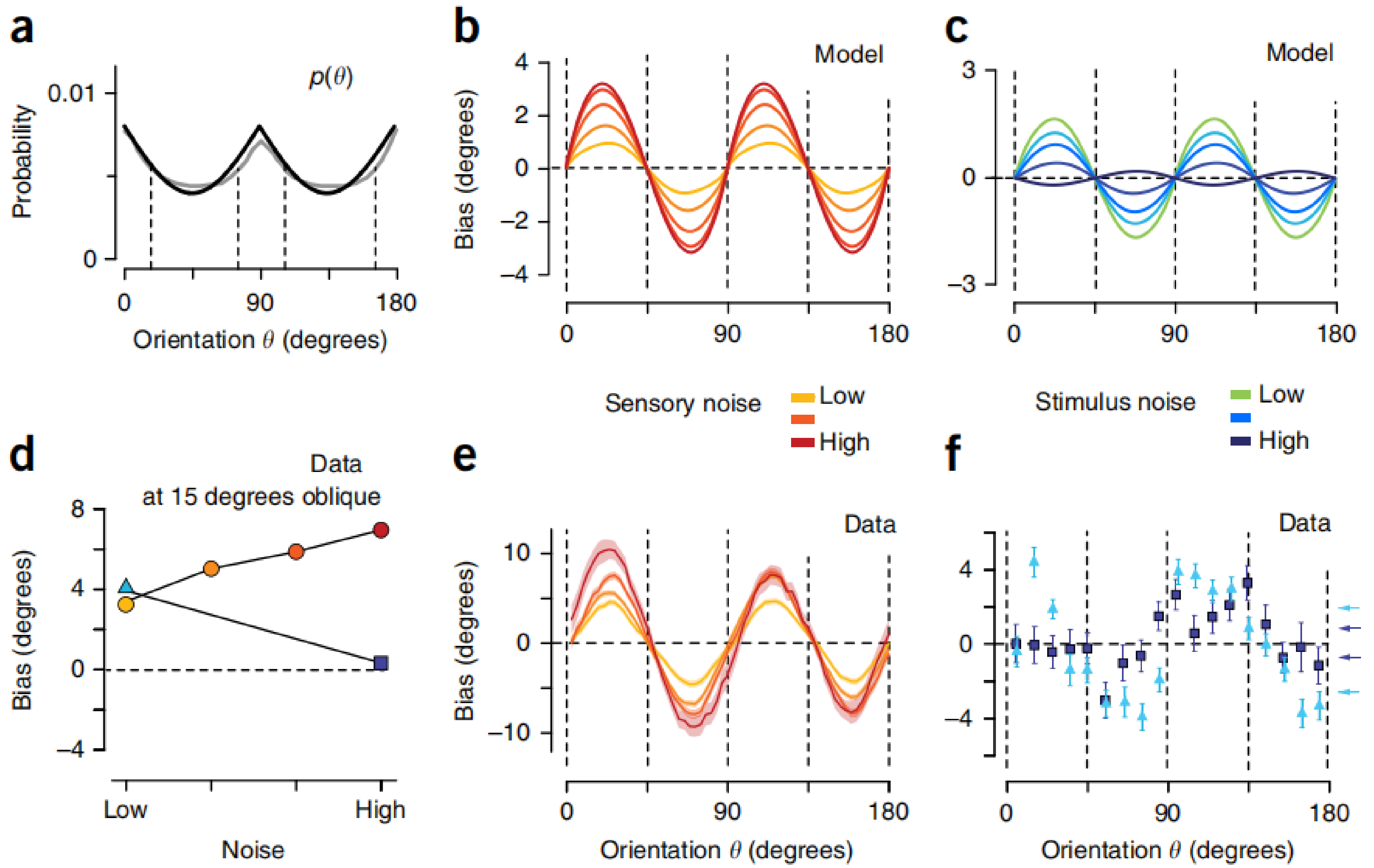
Prediction 2:

- Stimulus (external) and sensory (internal) noise differentially affect perceptual bias.

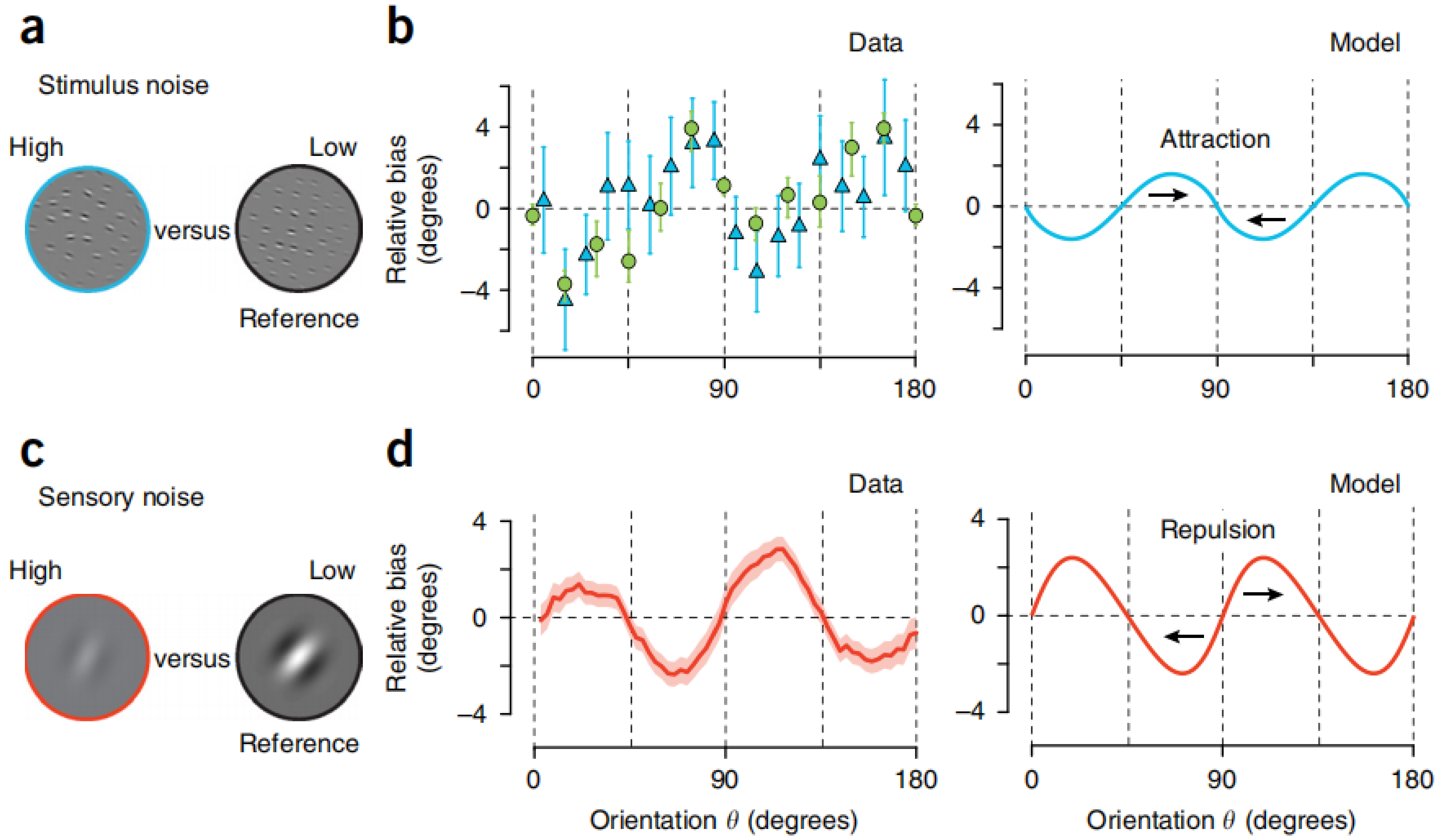




Bias in perceived orientation



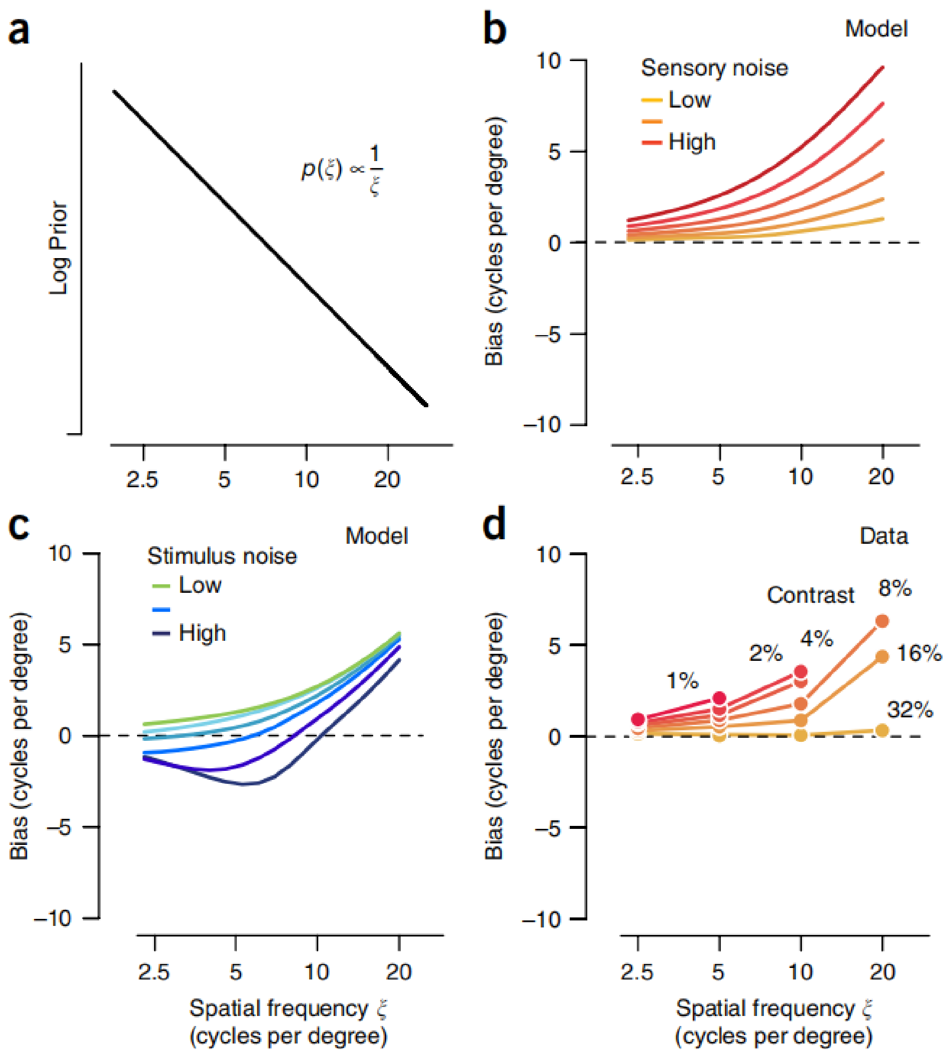
Relative bias in perceived orientation



Bias in perceived spatial frequency

Prior

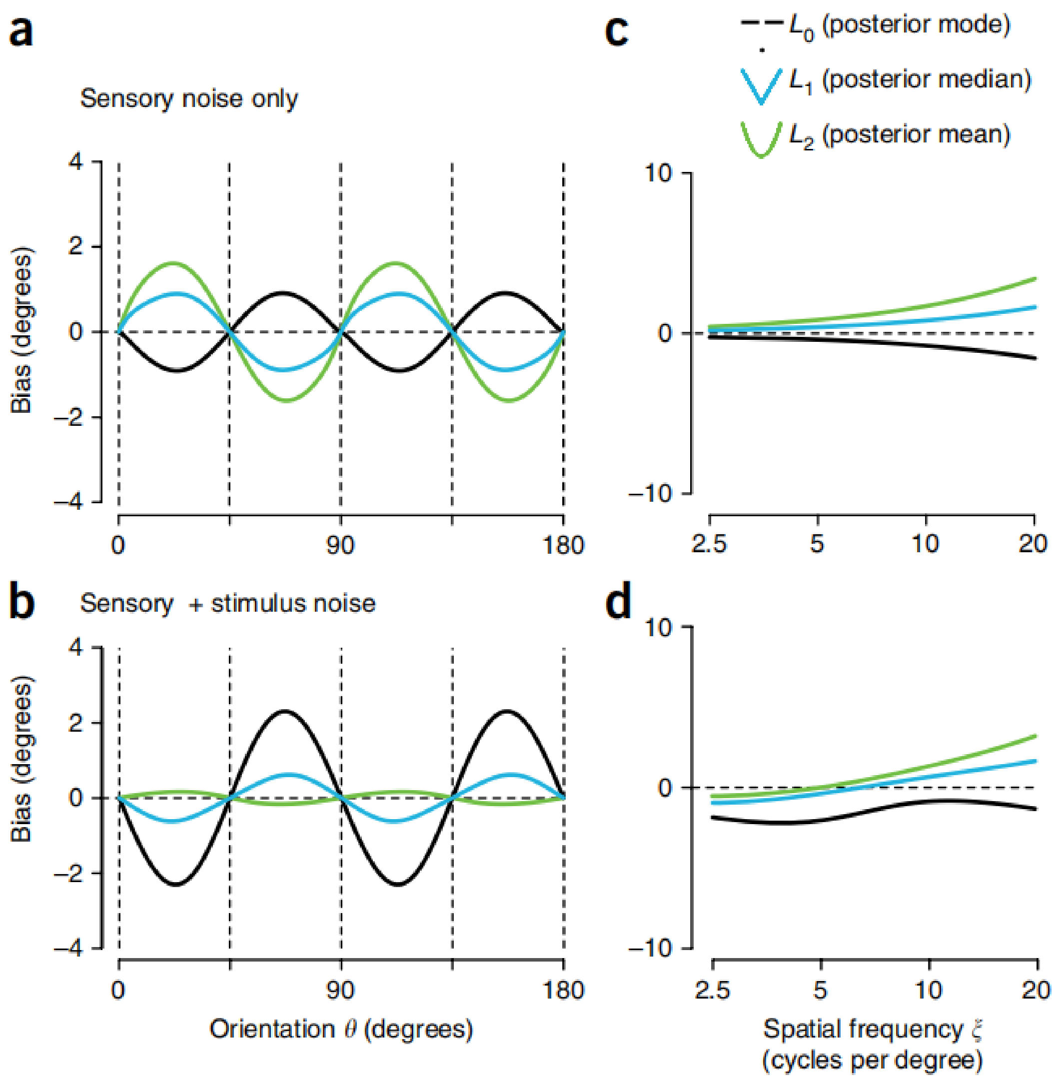
Bias with stimulus noise



Bias with sensory noise

Experiment data

Predicted biases for different loss functions

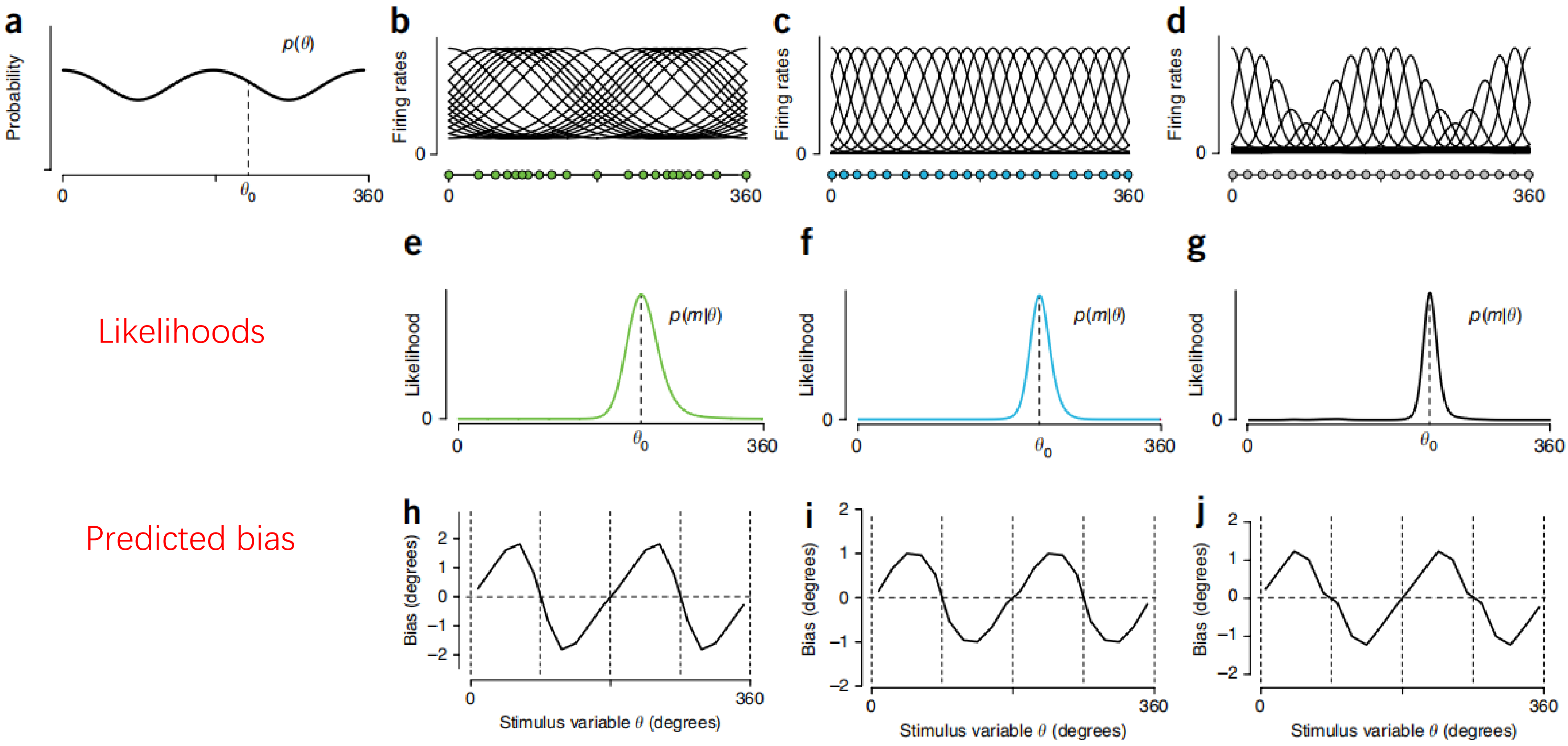


Equivalent efficient neural representation

Tuning characteristics

Likelihoods

Predicted bias



Takeaways

- The framework can be applied to spatial frequency perception
- More results on different noise type
- Can support different types of neural representation

Discussion & question

- How to understand the mapping F(theta)
- Other efficient coding objectives?
- Other `anti-Bayesian` percepts?

Have a great weekend!