Fitting Normative Neural Sampling Models to Neuronal Response Data

Suhas Shrinivasan[1,*], Andreas S. Tolias[2], Edgar Y. Walker[3,†], Fabian H. Sinz[1,†]

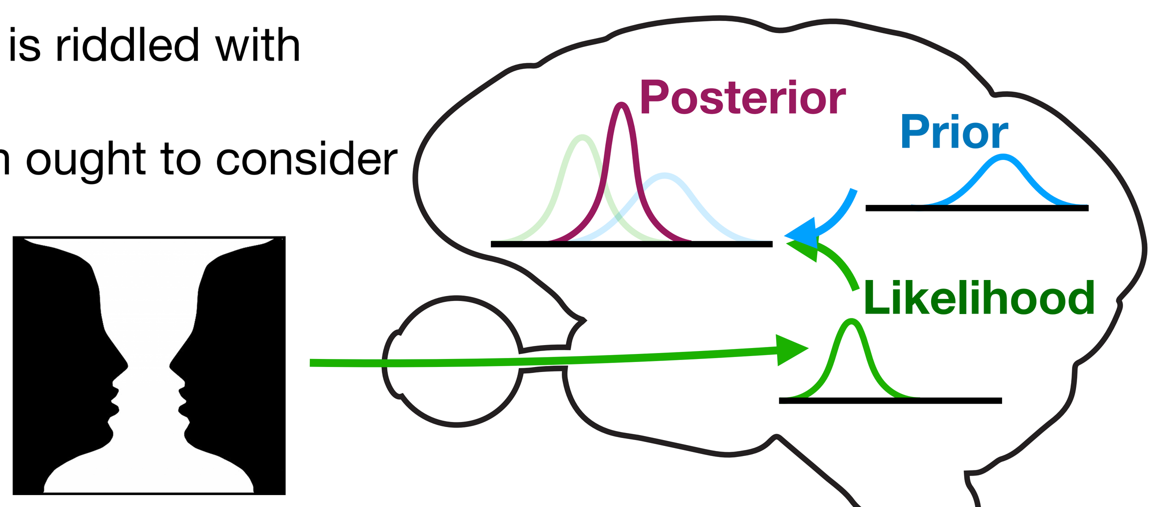
[1]University of Göttingen, Germany; [2]Baylor College of Medicine, Houston, TX, USA; [3]University of Washington, Seattle, USA

*shrinivasan@cs.uni-goettingen.de, †equal contribution

Probabilistic Perception: Bayesian Brain Hypothesis

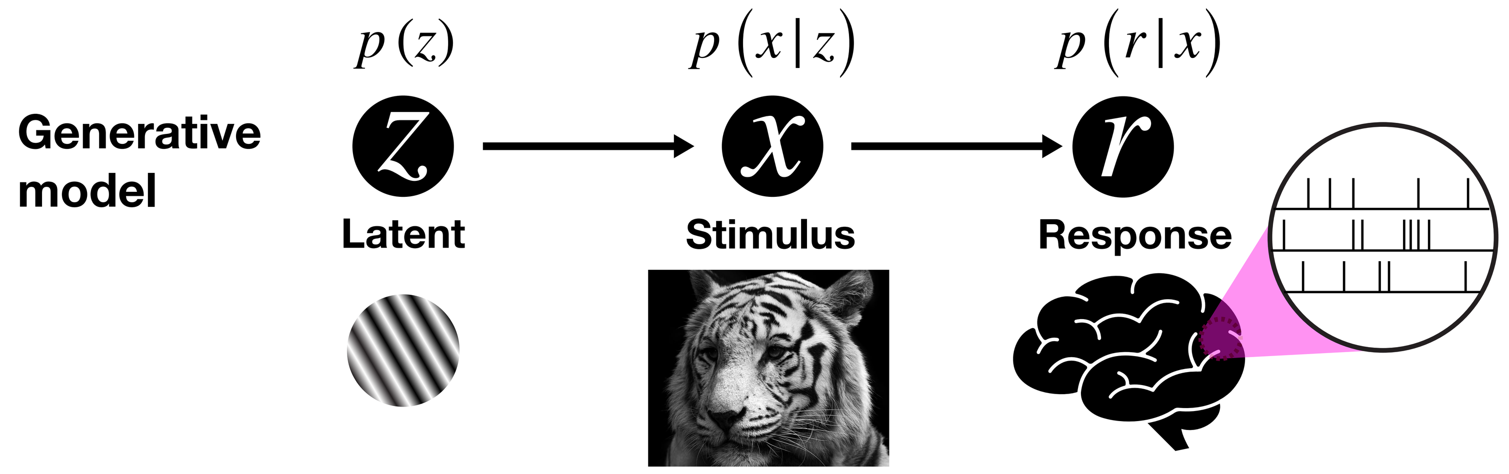
- Our sensory world is riddled with uncertainty
- Optimal perception ought to consider uncertainty

The Rubin Vase illusion: Two faces or a vase?

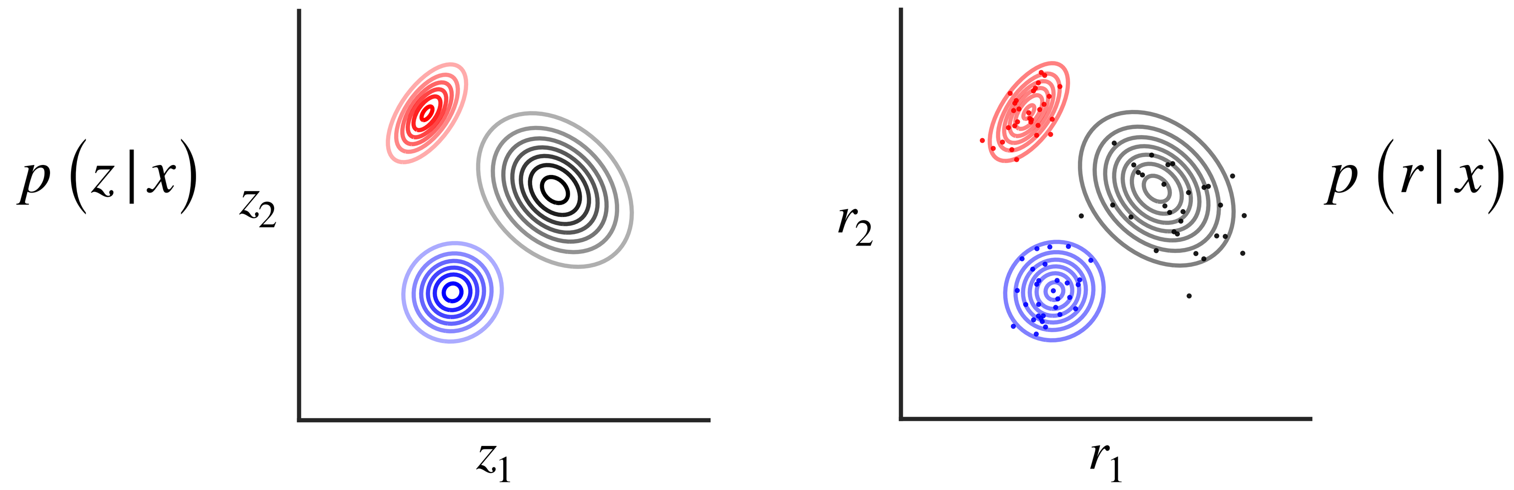


Bayesian Brain Hypothesis: Brains perform perception **probabilistically** by combining stimulus-based **likelihood** and **prior** knowledge to obtain the **posterior** distribution

Neural Sampling Hypothesis: A neural basis of probabilistic perception



- NSH posits that neuronal responses r are samples from posterior distribution over latent z given stimulus x: $p(z|x)$
- It follows that $p(r|x)$ ought to match $p(z|x)$



Central challenge: identifying the generative model over stimulus $p(x,z)$, specifically, identifying $p(z)$ and $p(x|z)$ employed by the brain

Our goal: Fitting NSH in a data-driven manner

- Assuming NSH, **learn** the generative model directly from a dataset of the neuronal responses to naturalistic visual stimuli
- Use **flexible, deep learning-based** generative models to fit data best while minimizing model bias prevalent in classical NSH works
- Establish **quantitative** evaluation and comparison of different NSH generative models, and to SOTA models in system identification

Formalization of NSH

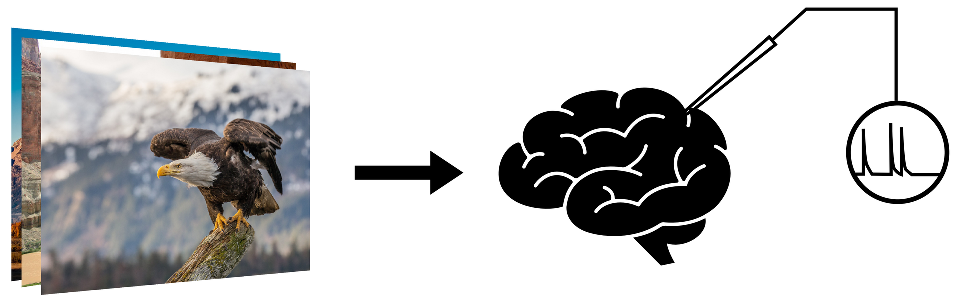
- We formulate NSH as positing the density equivalence $p_{r|x} \overset{d}{=} p_{z|x}$
- We can thereby derive the equivalence between marginals $p_r \overset{d}{=} p_z$
- NSH establishes a one-to-one correspondence between z and r, yielding:

$$p(r) \equiv p(z) \quad r \longrightarrow x \quad p(x|r) \equiv p(x|z)$$

- It is assumed then that each neuron encodes a latent variable underlying the stimulus

Learning the generative model under NSH

- This equivalence allows us to hence learn $p(x,z)$ by learning $p(x,r)$



- Dataset of natural images and responses: $\{x_i, r_i\}_{i=1}^N$

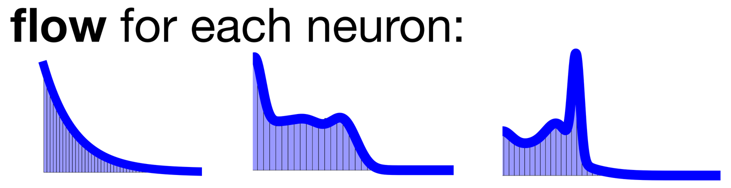
- We learn $p(x,r)$ on the recorded dataset by maximizing the likelihood of data

$$\theta^* = \arg\max_\theta \prod_i^N \left\{ p\left(x_i | r_i; \theta\right) \cdot p\left(r_i; \theta\right) \right\}$$

where θ is the parameter of the generative model

- We model $p(r)$ using a **deep normalizing flow** for each neuron:

$$p(r) = \mathcal{N}\left(\mathcal{F}_\phi^\dagger(r) | 0, \mathbf{I}\right) \cdot \left| \det \frac{\partial \mathcal{F}_\phi^\dagger(r)}{\partial r} \right|$$



- We model $p(x|r)$ as $p(x|r) = \mathcal{N}\left(x | \mu = g_\theta^\dagger(r), \sigma^2 = h_\theta^\dagger(r)\right)$

† $\mathcal{F}_\phi$ is a series of invertible transformations † g_θ and h_θ are multilayer perceptrons (MLP)

Simulations on model images and neurons

We simulated pairs of images and responses under

- Hoyer & Hyvärinen model (HNH)*
- Olshausen & Field model (ONF)*
- baseline full Gaussian model (Gauss)*
- Flexible model (Flex)

and then fit each model to get log likelihood scores

Flex **outperforms** the fit of mismatched generative models

Models \ Data	HNH	ONF	Gauss	Flex
HNH	−92.55			−449.81
ONF	−122.73	−381.58	−360.19	−460.91
Gauss	−131.58	−388.44	−355.80	−523.58
Flex	−93.62	−386.00	−358.54	−217.94

Colour scale: Best – Worst

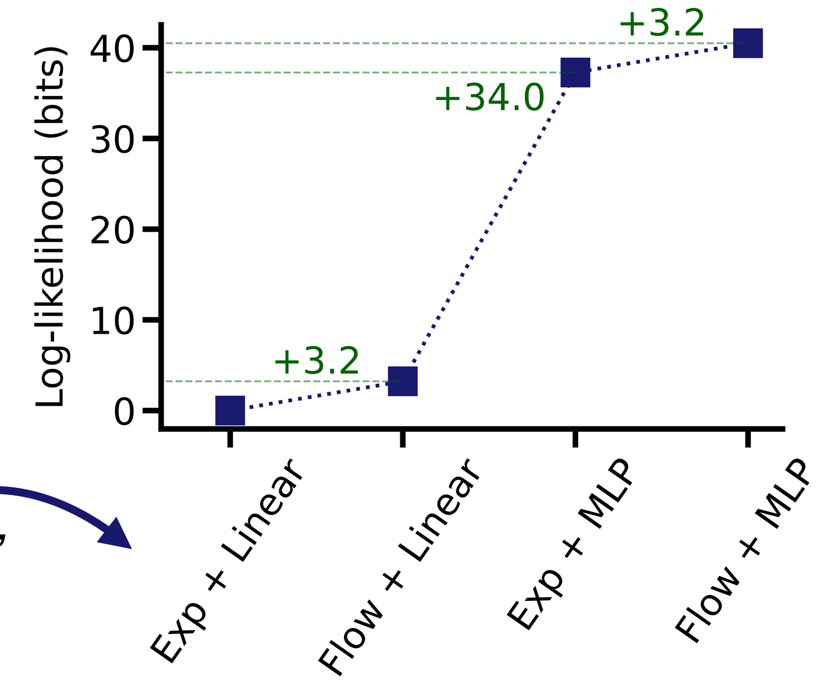
* Same Likelihood $p(x|r) = \mathcal{N}\left(x | \mathrm{A}r, \sigma^2 \cdot I\right)$

$$p_{\mathrm{HNH}}(r) = \frac{1}{\lambda} e^{\frac{-r}{\lambda}} \cdot \mathbf{H}(r), \quad p_{ONF}(r) = \frac{1}{2\mathrm{b}} e^{-\frac{|r-a|}{b}},$$
$$p_{\mathrm{Gauss}}(r) = \mathcal{N}\left(r | \mu, \sigma^2 \cdot I\right)$$

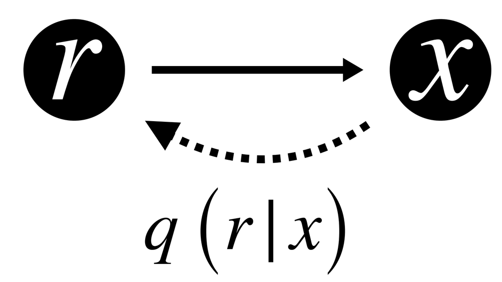
Fitting generative model $p(x,r)$ on V1 spike counts from population recordings

- We obtained V1 population spike counts to **natural images** (ImageNet dataset) recorded from awake Macaques using 32-channel (NeuroNexus) arrays.
- Each image was presented for 120 ms and we extracted spike counts from 40 ms to 160 ms after the image onset.
- We fit the classical as well as our flexible models
- The fits let us compute exact log likelihood scores, that allows us to rigorously compare normative hypotheses

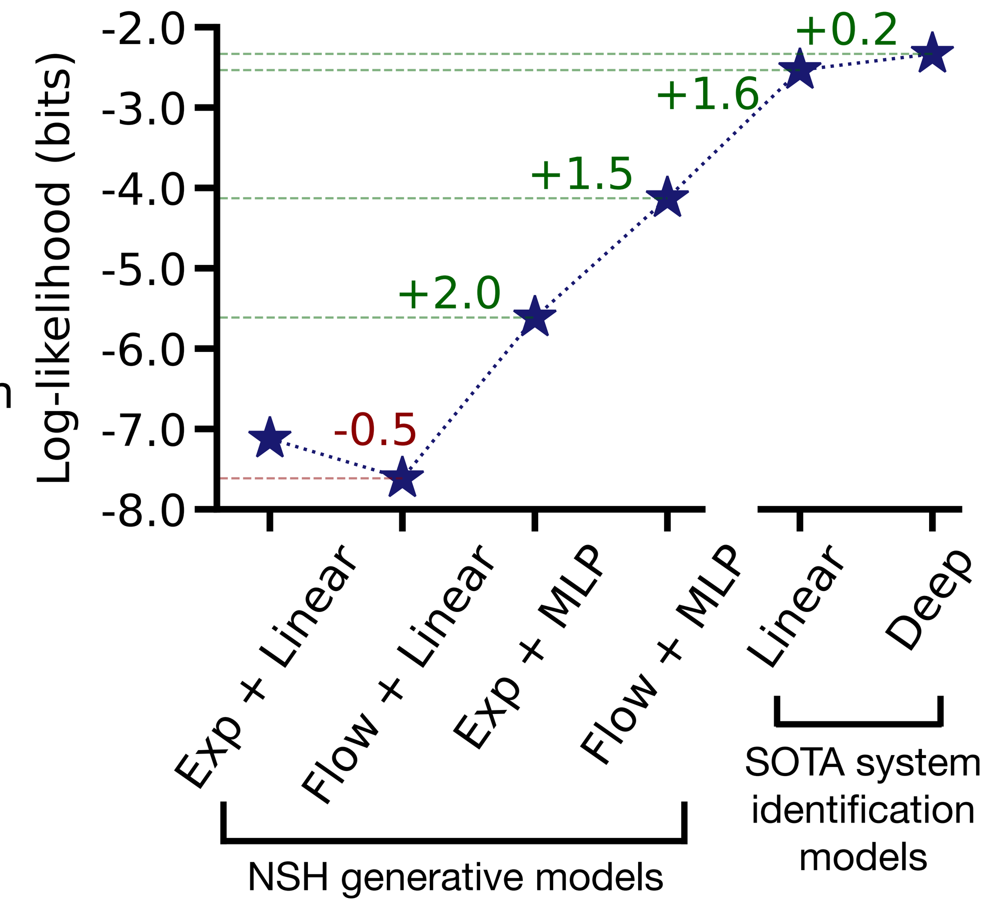
Each is a generative model consisting of "Prior + Likelihood"



Getting image conditioned neuron-specific predictions via variational inference, and system identification



- Approximate posterior using a Gamma distribution and an MLP amortized inference function
- Enables comparison to SOTA system identification
- Enables neuron-specific predictions from normative theory



Ongoing and future work

- Extend approximate posterior to be more flexible (flow-based)
- Fit models on data from different areas (V4) and different animals (mouse V1)

References

Olshausen & Field. "Emergence of simple-cell receptive field properties by learning a sparse code for natural images." *Nature* (1996)

Hoyer & Hyvärinen. "Interpreting neural response variability as Monte Carlo sampling of the posterior." *NeurIPS* 15 (2002)

Haefner et al. "Perceptual decision-making as probabilistic inference by neural sampling." *Neuron* (2016)