

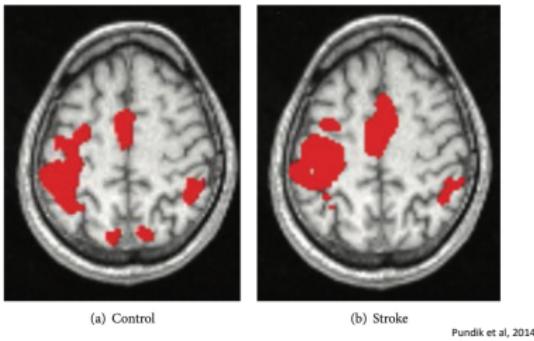
Bayesian Nonparametric approaches for capturing the heterogeneity of neuroimaging experiments

Michele Guindani

Department of Biostatistics
University of California, Los Angeles

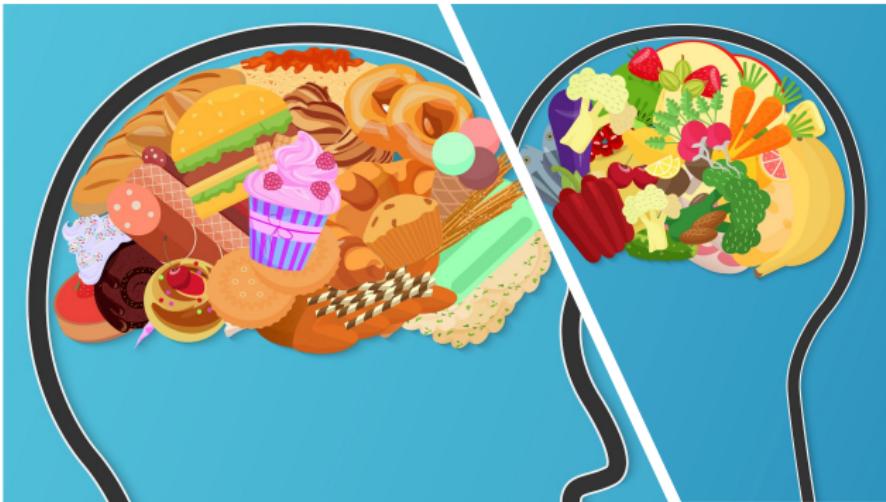
BNP 13 Meeting
October 25th, 2022

- ❑ Standard methods in brain research have long assumed that it is possible to average together the brain maps of all subjects in a study
 - ⇒ **Average maps** are typically used to describe the brain functioning for the individuals in a study.



Pundik et al, 2014

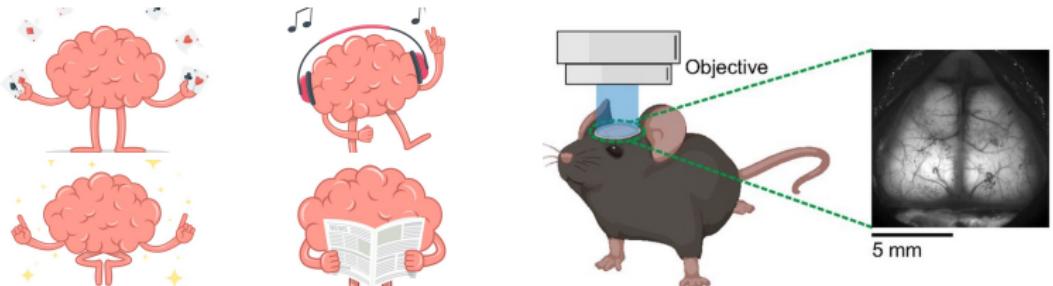
- ❑ Some dimension reduction methods/preprocessing steps (e.g. GICA in fMRI studies) implicitly assume the existence of common patterns across subjects (e.g. by encouraging to match ICA components across subjects)
 - ⇒ Spatial homogeneity of brain patterns; reduced ability to capture inter-subject variability (Michael et al, 2014).



- ❑ There is an increasing recognition that brain functioning is **heterogeneous** and varies greatly **both within and between** individuals:
 - differences in activation to different stimuli
 - differences in connectivity to different stimuli
 - the different brain activity patterns may be associated to a clinical outcome or different behaviors (e.g. large brain responses to food-related cues predict cue-induced eating, Versace et al, 2019)

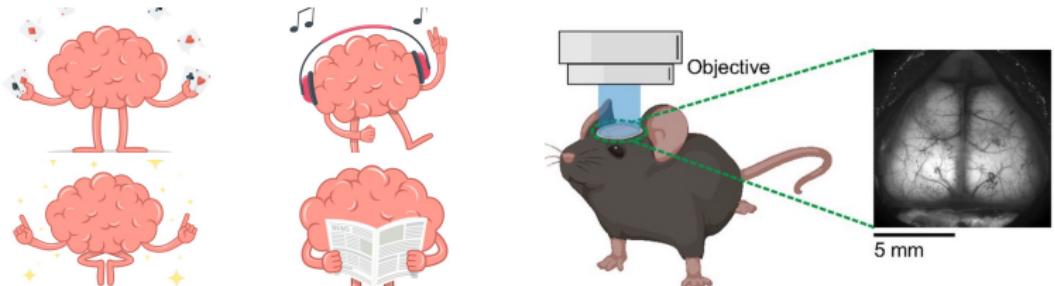


- ❑ In this talk, we will discuss a few frameworks for describing the heterogeneity of brain patterns in some specific data:



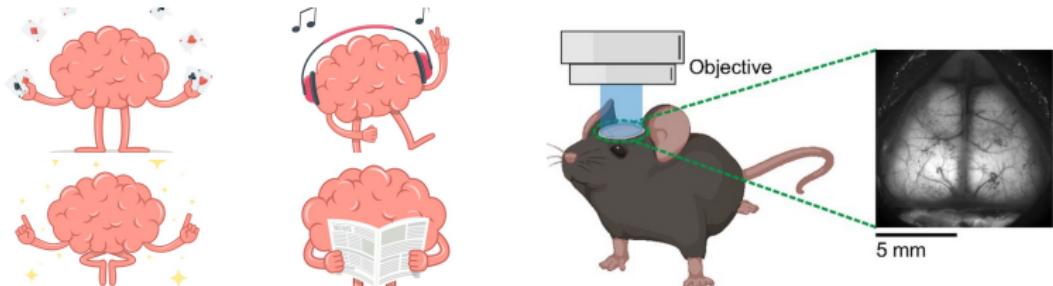
Ren & Komiyama (2021), Journal of Neurosciences

- ❑ In this talk, we will discuss a few frameworks for describing the heterogeneity of brain patterns in some specific data:
- ① Capturing **activity spikes** in *in-vivo* experiments in animals



Ren & Komiyama (2021), Journal of Neurosciences

- ❑ In this talk, we will discuss a few frameworks for describing the heterogeneity of brain patterns in some specific data:
 - ① Capturing **activity spikes** in *in-vivo* experiments in animals
 - ② Capturing the **heterogeneity of signals** in large-scale hypothesis testing



Ren & Komiyama (2021), Journal of Neurosciences

- ❑ In this talk, we will discuss a few frameworks for describing the heterogeneity of brain patterns in some specific data:
 - ① Capturing **activity spikes** in *in-vivo* experiments in animals
 - ② Capturing the **heterogeneity of signals** in large-scale hypothesis testing
- **Ultimate Objective:** Association with clinical/behavioral outcomes - more, in general - with observable phenotypes.

Detection of spikes in noisy calcium imaging data

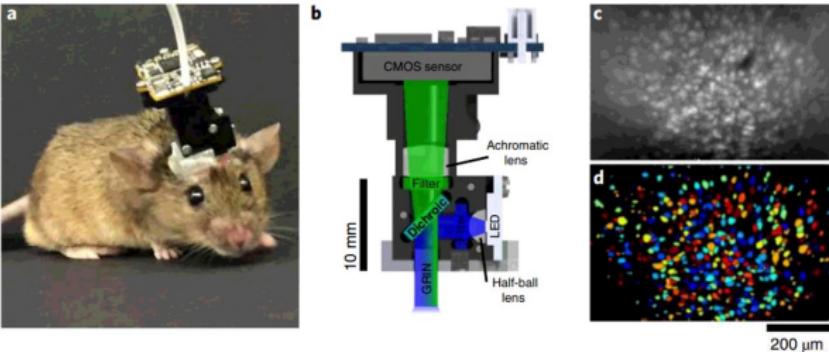
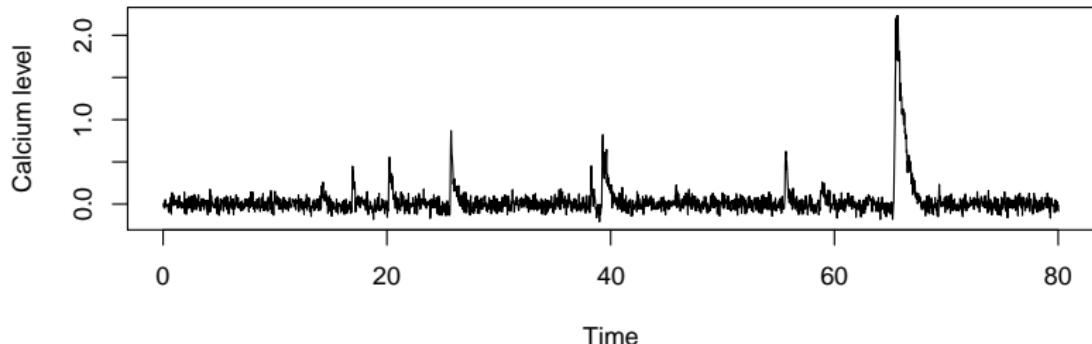


Fig. 1 | Open-source UCLA Miniscope. **a**, A mouse with a head-mounted Miniscope. **b**, Cross-sectional rendering of the Miniscope optical path. Blue, excitation path; green, emission optical path; GRIN, gradient-index lens. **c**, Maximum projection of a 10-minute motion-corrected Miniscope recording of hippocampal CA1 pyramidal neurons labeled with GCaMP6f. **d**, Spatial footprints of identified neurons from the recording in **c**. Scale bar in **d** applies to **c**.

NATURE METHODS | VOL 16 | JANUARY 2019 | 9–15 | www.nature.com/naturemethods

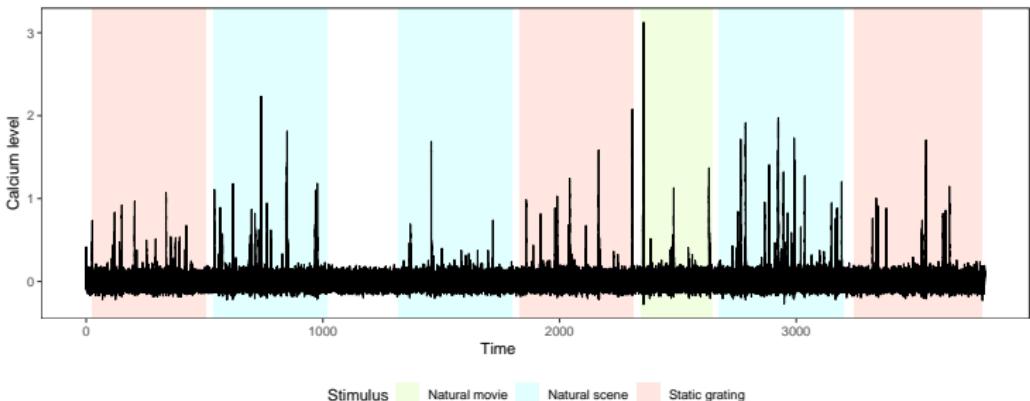
- ✓ Calcium imaging is a microscopy technique to optically measure the intra-cellular calcium concentration of neurons in awake animals.
- ✓ The mechanism at the basis of calcium imaging is a physiological process of the cells: when a neuron fires, **calcium floods the cell** and produces a transient spike in its concentration
- ☞ Outcome: **movie** of time-varying fluorescence intensities for each observable neuron in a targeted area.



- ✓ Calcium concentration can be used as a **proxy of the neuronal activity**.
- ☞ The goal is to investigate **how individual neurons react to stimulation** and how they encode information.
- ☞ It's important **to deconvolve the calcium traces** and identify the **precise spike times** of the observable neurons

Usually, the experiment involves multiple stimuli (e.g. visual stimuli, or odors):

- the interest is to understand how the **different types of stimuli** affect the neuronal activity \Rightarrow investigate **similarities and differences** in the distribution of spikes over time and conditions.



- Calcium traces are not the target of the analysis.
- ✓ The interest is to study the **spikes' times and amplitudes**, which constitute the neuronal activity.
- ☞ Limits of calcium imaging:
 - presence of measurement noise
 - slow decay of the calcium concentration compared to the spiking activity.
 - long time series

- Calcium traces are not the target of the analysis.
- ✓ The interest is to study the **spikes' times and amplitudes**, which constitute the neuronal activity.
- ☞ Limits of calcium imaging:
 - presence of measurement noise
 - slow decay of the calcium concentration compared to the spiking activity.
 - long time series
- **Objectives:**
 - identify the spikes' times → discriminate between signal and noise
 - estimate the **distribution** of the spikes' amplitudes
 - model **the effect of different stimuli on the distribution** of the spikes' amplitudes.

Model for the calcium dynamics

A popular model¹ relates the observed trace y_t to the underlying true calcium concentration c_t , and the neuronal activity A_t :

$$y_t = b + c_t + \epsilon_t \quad \epsilon_t \sim N(0, \sigma^2)$$

$$c_t = \gamma c_{t-1} + A_t + \omega_t \quad \omega_t \sim N(0, \tau^2)$$

for $t = 1, \dots, T$; with b baseline level, ϵ_t measurement error.

¹ Vogelstein et al. (2010). Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology* **104**, 3691–3704

Model for the calcium dynamics

A popular model¹ relates the observed trace y_t to the underlying true calcium concentration c_t , and the neuronal activity A_t :

$$y_t = b + c_t + \epsilon_t \quad \epsilon_t \sim N(0, \sigma^2)$$

$$c_t = \gamma c_{t-1} + A_t + \omega_t \quad \omega_t \sim N(0, \tau^2)$$

for $t = 1, \dots, T$; with b baseline level, ϵ_t measurement error.

- In absence of neuronal activity: $A_t = 0$ and the calcium level follows a AR(1) process controlled by the parameter γ ;

¹ Vogelstein et al. (2010). Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology* **104**, 3691–3704

Model for the calcium dynamics

A popular model¹ relates the observed trace y_t to the underlying true calcium concentration c_t , and the neuronal activity A_t :

$$y_t = b + c_t + \epsilon_t \quad \epsilon_t \sim N(0, \sigma^2)$$

$$c_t = \gamma c_{t-1} + A_t + \omega_t \quad \omega_t \sim N(0, \tau^2)$$

for $t = 1, \dots, T$; with b baseline level, ϵ_t measurement error.

- In absence of neuronal activity: $A_t = 0$ and the calcium level follows a AR(1) process controlled by the parameter γ ;
- when a spike occurs: $A_t > 0$ and the concentration increases instantaneously with the spike amplitude A_t .

¹ Vogelstein et al. (2010). Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology* **104**, 3691–3704

To allow the response to vary according to the condition, we assume that the spikes A_t come from **stimulus-specific distributions**: for $j = 1, \dots, J$

$$A_t \mid g_t = j, G_j \sim G_j.$$

To allow the response to vary according to the condition, we assume that the spikes A_t come from **stimulus-specific distributions**: for $j = 1, \dots, J$

$$A_t | g_t = j, G_j \sim G_j.$$

To model the G_j 's we could adopt a Bayesian **nested Dirichlet Process**:

nested structure → reconstruct the distribution within each experimental condition + borrow information between groups (distributional clustering)

mixture formulation → cluster the A_t across and within distributions
⇒ discover similarities in the activation response to different stimuli.

To allow the response to vary according to the condition, we assume that the spikes A_t come from **stimulus-specific distributions**: for $j = 1, \dots, J$

$$A_t | g_t = j, G_j \sim G_j.$$

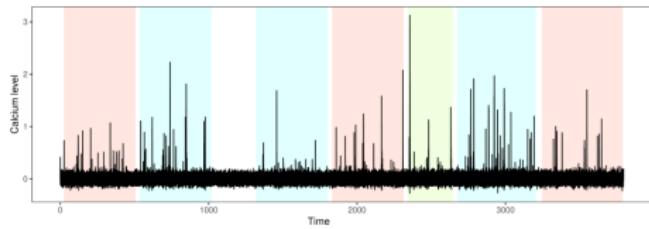
To model the G_j 's we could adopt a Bayesian **nested Dirichlet Process**:

nested structure → reconstruct the distribution within each experimental condition + borrow information between groups (distributional clustering)

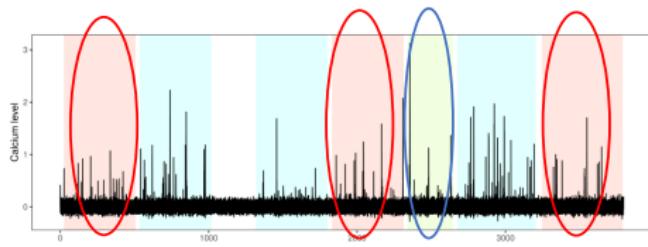
mixture formulation → cluster the A_t across and within distributions
⇒ discover similarities in the activation response to different stimuli.

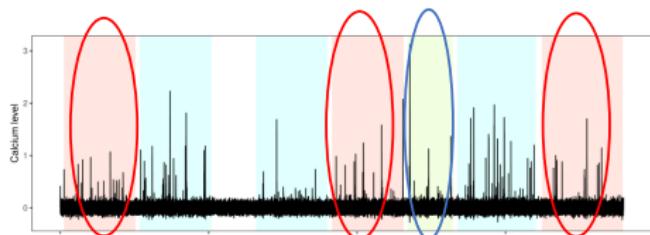
The model allows to represent the data through two-layers: at the first level clusters of distributions across conditions, and at the second level a convenient representation of the group distributions

Visual idea of the NDP



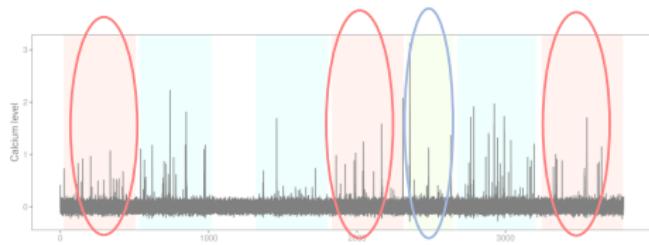
Visual idea of the NDP





$$A_t \mid g_t = 1 \sim G_1.$$

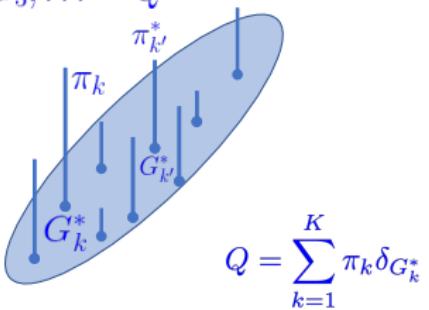
$$A_t \mid g_t = 3 \sim G_3$$

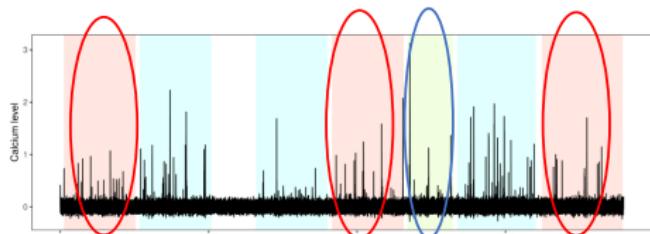


$$A_t \mid g_t = 1 \sim G_1.$$

$$A_t \mid g_t = 3 \sim G_3$$

$$G_1, G_2, G_3, \dots \sim Q$$

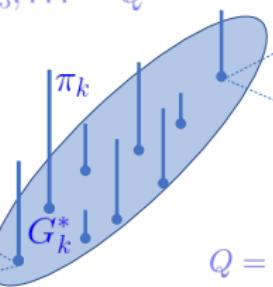
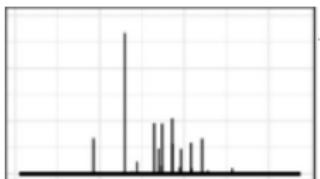




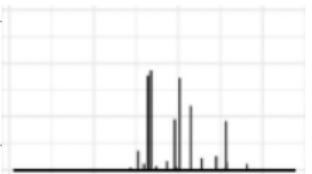
$$A_t \mid g_t = 1 \sim G_1.$$

$$A_t \mid g_t = 3 \sim G_3$$

$$G_1, G_2, G_3, \dots \sim Q$$



$$Q = \sum_{k=1}^K \pi_k \delta_{G_k^*} \quad G_k^*(\cdot) = \sum_{l=1}^{\infty} \omega_{lk} \delta_{\theta_{lk}}(\cdot)$$



- Camerlenghi et al (2019) have recently proved that the inference obtained using the nDP may be affected by a *degeneracy* property:
 - ⇒ If two distributions share **even only one atom in their support**, the two distributions are automatically assigned to the same cluster.

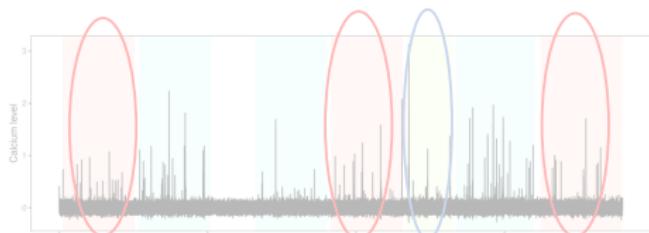
More precisely, the partially exchangeable partition probability function (pEPPF), i.e. the function which describes the probability of each clustering allocation for partially exchangeable data modeled with a nDP, collapses to a fully exchangeable case when ties are present among the observational atoms.

The problem persists with nDP mixture model formulations

- Camerlenghi et al (2019) propose a class of latent nested processes, which relies on estimating a latent **mixture of shared and idiosyncratic processes** \Leftrightarrow computationally complex, only small datasets with few groups.

- Camerlenghi et al (2019) propose a class of latent nested processes, which relies on estimating a latent **mixture of shared and idiosyncratic processes** \Leftrightarrow computationally complex, only small datasets with few groups.
- Beraha, Guglielmi & Quintana (2021) propose a variation of the hierarchical DP, where the **baseline distribution** is itself a **mixture of a DP and a non-atomic measure** (semi-HDP). They further combine the semi-HDP prior with a random partition model that allows different populations to be grouped in clusters that are internally homogeneous, i.e. arising from the same distribution.
- Denti, Camerlenghi, Guindani & Mira (2022+) show that the degeneracy is avoided if the prior explicitly models **commonality of atoms** between groups.
- Lijoi, Pruenster, Rebaudo (2022+) move this idea further along by essentially combining the NDP and the HDP into a *hidden hierarchical Dirichlet Process*.

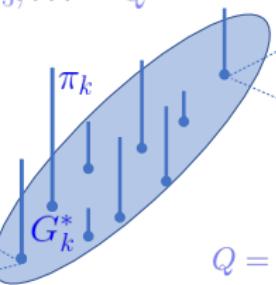
Visual idea of the Common Atom Model



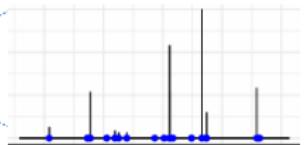
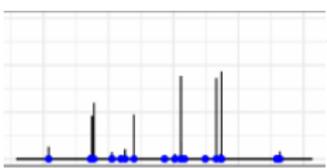
$$A_t \mid g_t = 1 \sim G_1.$$

$$A_t \mid g_t = 3 \sim G_3$$

$$G_1, G_2, G_3, \dots \sim Q$$



$$Q = \sum_{k=1}^K \pi_k \delta_{G_k^*} \quad G_k^*(\cdot) = \sum_{l=1}^{\infty} \omega_{lk} \delta_{\theta_l}(\cdot)$$



- For computational efficiency (long time series), one can employ the **generalized mixtures of finite mixtures (gMFM)** of Frühwirth-Schnatter et al. (*BA*, 2021) where the nested structure is based on the **common atom model**:

$$A_t \mid g_t = j, G_j \sim G_j.$$

$$G_1, \dots, G_J \mid Q \sim Q, \quad Q = \sum_{k=1}^K \pi_k \delta_{G_k^*}$$

where G_k^* are distributions (identifying clusters of distributions across conditions/experimental settings)

- For computational efficiency (long time series), one can employ the **generalized mixtures of finite mixtures (gMFM)** of Frühwirth-Schnatter et al. (*BA*, 2021) where the nested structure is based on the **common atom model**:

$$A_t \mid g_t = j, G_j \sim G_j.$$

$$G_1, \dots, G_J \mid Q \sim Q, \quad Q = \sum_{k=1}^K \pi_k \delta_{G_k^*}$$

where G_k^* are distributions (identifying clusters of distributions across conditions/experimental settings)

- More specifically, we assume:

$$\pi_1, \dots, \pi_K \mid K, \alpha \sim Dir_K\left(\frac{\alpha}{K}, \dots, \frac{\alpha}{K}\right)$$

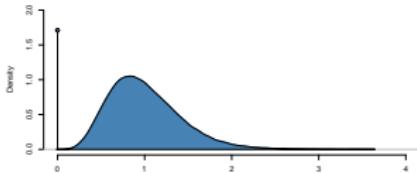
$$K - 1 \sim \text{beta-negative-binomial}$$

$$\alpha \sim F$$

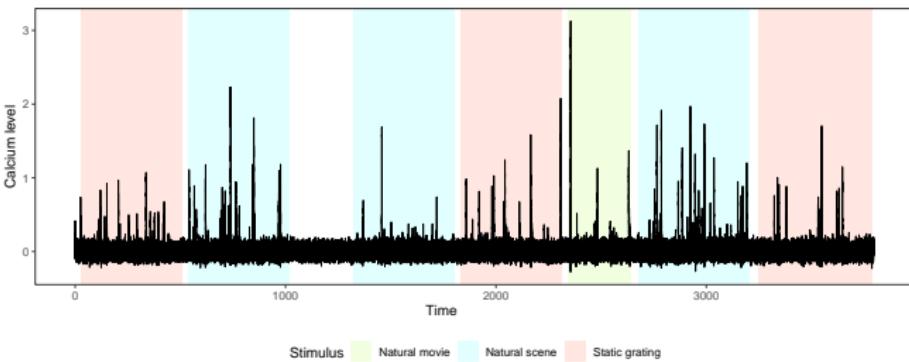
(Frühwirth-Schnatter et al., 2021)

- To enforce sparsity in the detection of spikes we model the base measure G_0 for the parameters A_i^* using a **spike-and-slab** specification:

$$G_0 = (1-p) \delta_0 + p \text{Gamma}(h_1, h_2)$$



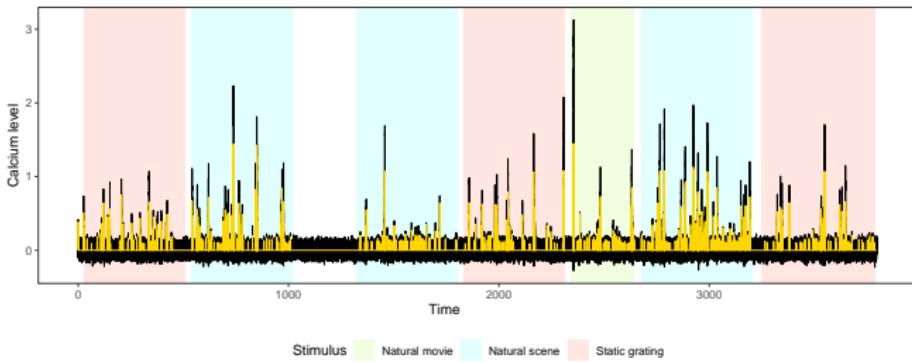
- By choosing (h_1, h_2) appropriately we can mimic the behavior of a non-local prior density (Johnson & Rossell, 2010) and induce a clear separation between zero (baseline neuronal activity) and the positive values (neuronal response)
- We do multiple simulation studies and then chose on $h_1 = h_2 = 8$ (mean=1, var=0.125).



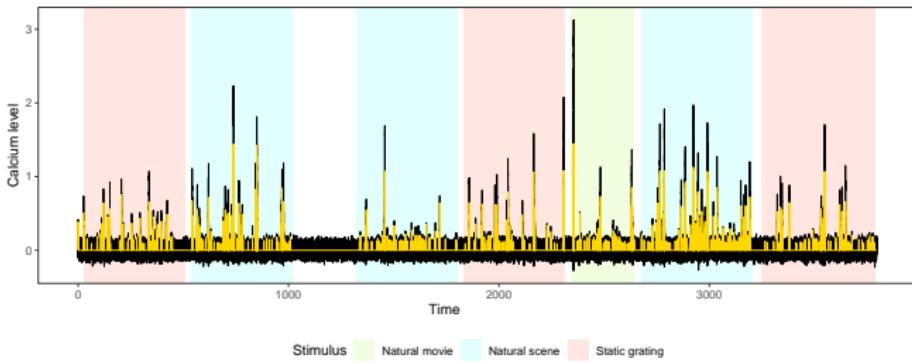
4 experimental conditions:

- 3 stimuli of increasing complexity (static grating, natural scene, natural movie)
- period of spontaneous activity (absence of stimuli)

⁶Allen Institute for Brain Science (2016). Allen brain observatory.
<http://observatory.brain-map.org/visualcoding>.



⁶Allen Institute for Brain Science (2016). Allen brain observatory.
<http://observatory.brain-map.org/visualcoding>.



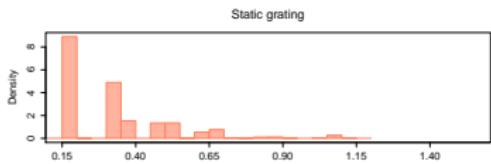
Cluster 1 = { Natural scene, Natural movie }

Cluster 2 = { Static grating }

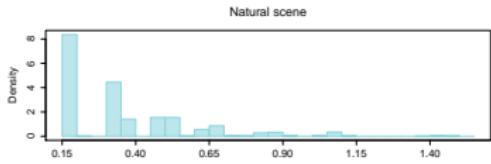
Cluster 3 = { Absence of stimuli }

⁶Allen Institute for Brain Science (2016). Allen brain observatory.
<http://observatory.brain-map.org/visualcoding>.

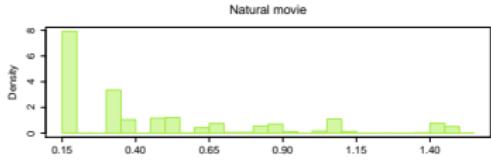
Estimated firing rate: number
of spikes per second
(posterior point estimate and credible interval)



0.287 (0.274, 0.300)



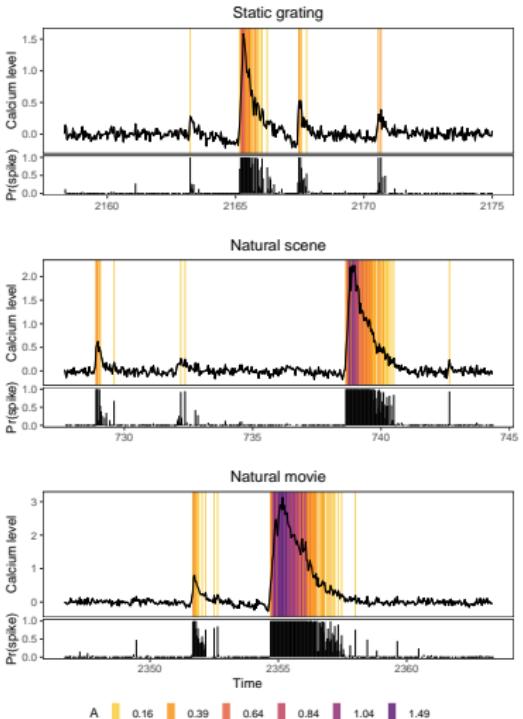
0.523 (0.504, 0.542)



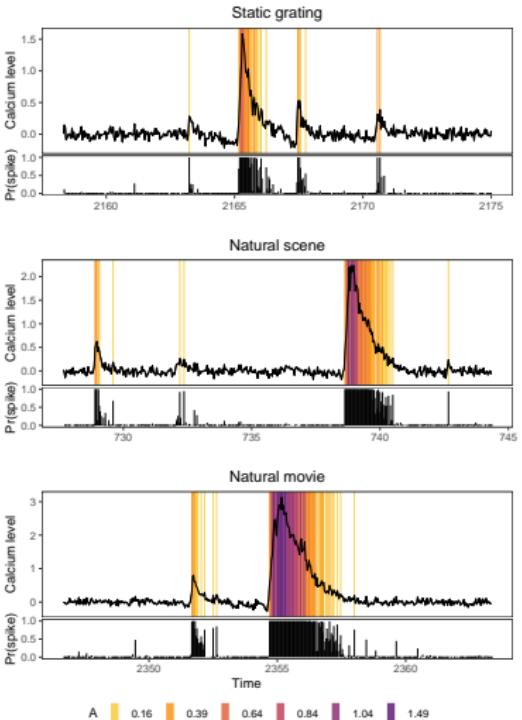
0.612 (0.575, 0.644)

Figure: Distribution of the spikes' amplitudes

Spikes amplitudes



Spikes amplitudes

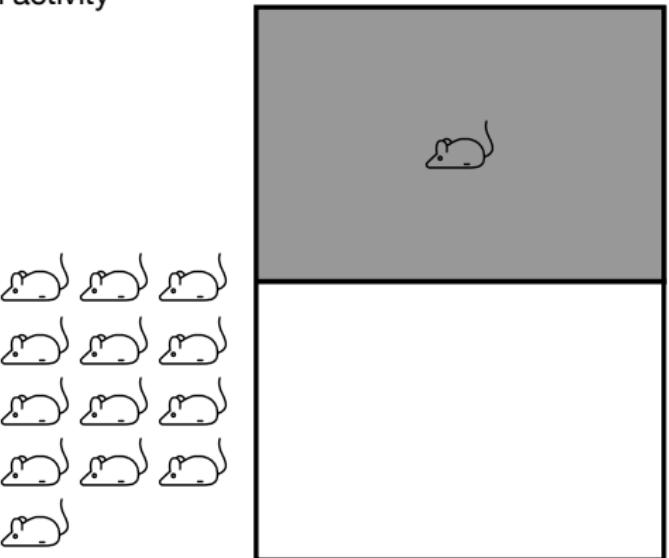


D'Angelo, Canale, Yu, Guindani (2022+), *Bayesian nonparametric analysis for the detection of spikes in noisy calcium imaging data*, *Biometrics*, In press. <https://arxiv.org/abs/2102.09403>

Bayesian mixtures for screening in large-scale testing

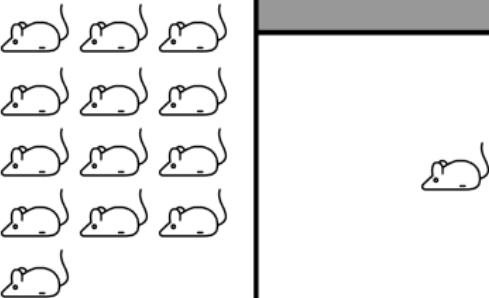
The light-sheet fluorescence microscopy dataset

- Fourteen mice were **individually housed in the dark** for 24 hours to establish baseline visual activity



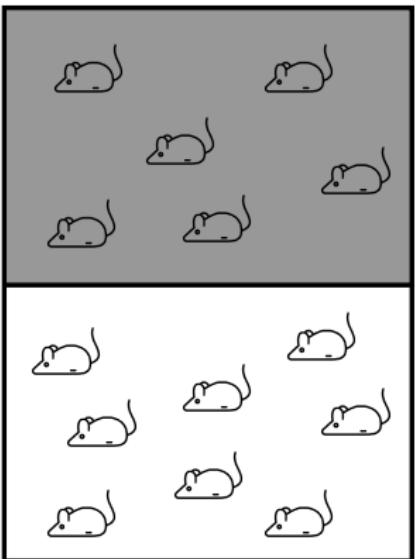
The light-sheet fluorescence microscopy dataset

- Fourteen mice were **individually housed in the dark** for 24 hours to establish baseline visual activity
- Mice were then transferred into a new cage **exposed to ambient light**

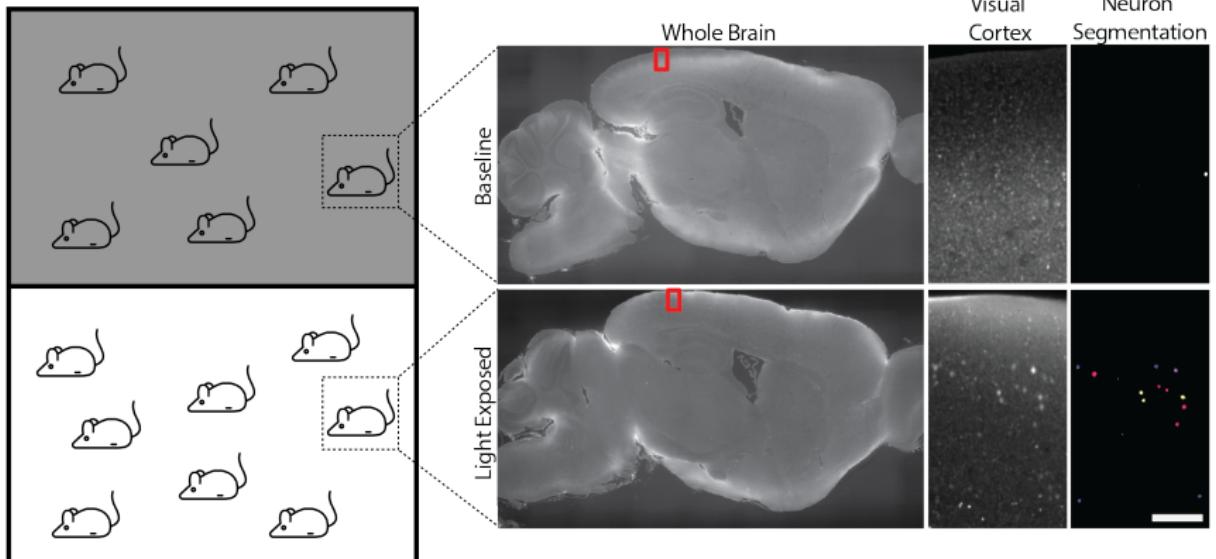


The light-sheet fluorescence microscopy dataset

- Fourteen mice were **individually housed in the dark** for 24 hours to establish baseline visual activity
- Mice were then transferred into a new cage **exposed to ambient light**
- The brains of six mice were examined **0-15 minutes** (i.e., no light) after light exposure to serve as the **baseline** group
- The brains of another eight mice were examined 30-120 minutes after light exposure, within the window of **Npas4** protein up-regulation (Ramamoorthi et al., 2011)



The light-sheet fluorescence microscopy dataset

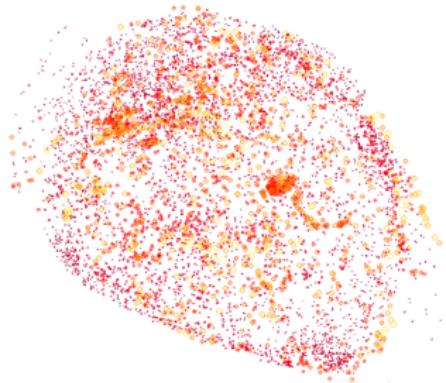


- The light-sheet fluorescence microscopy imaging techniques allows the detection of activated cells at high resolution *in vivo* in the whole-brain of the mouse.

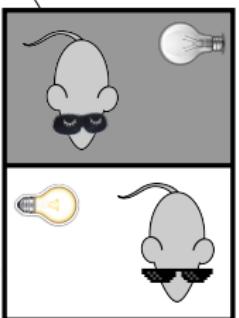
Light-sheet fluorescence microscopy (LSFM)

UCLA

CONTROL GROUP: DARKNESS



CASE GROUP: LIGHT-EXPOSED

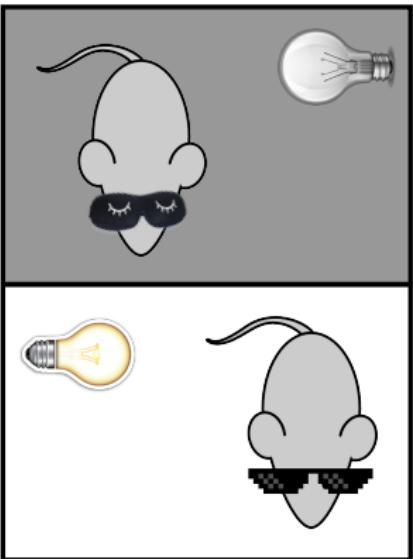


The light-sheet fluorescence microscopy dataset

- **GOAL of the study:**

- Assess **differentially activated regions** by comparing the baseline and light-exposed groups
- The activation is measured in terms of Npas4 expression (we will refer to this as **fluorescence**)
- Data are pre-processed eventually organized into 281 brain regions of interest and z-scores

$$Z_\nu = \beta_\nu + \varepsilon_\nu, \quad \varepsilon_\nu \sim N_T(0, \sigma)$$



The light-sheet fluorescence microscopy dataset

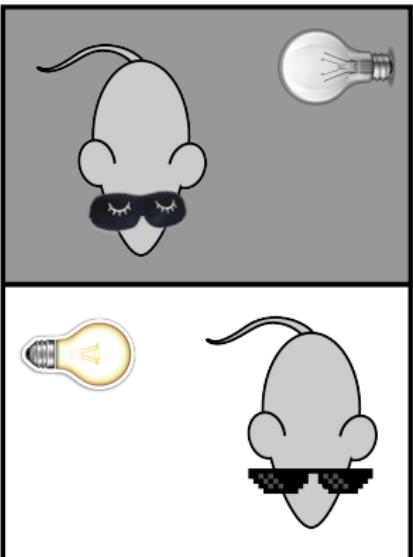
- **GOAL of the study:**

- Assess **differentially activated regions** by comparing the baseline and light-exposed groups
- The activation is measured in terms of Npas4 expression (we will refer to this as **fluorescence**)
- Data are pre-processed eventually organized into 281 brain regions of interest and z-scores

$$Z_\nu = \beta_\nu + \varepsilon_\nu, \quad \varepsilon_\nu \sim N_T(0, \sigma)$$

- BH discovers 142 regions (50%) too liberal!

The local FDR method (Efron, 2004) flags only 38 brain regions as relevant, however missing many regions known to be associated with the visual task.

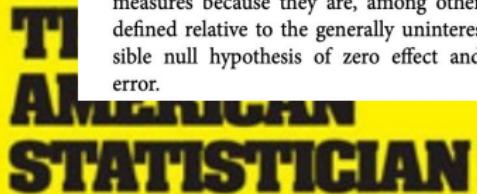


7.4. Adopting More Holistic Approaches

**McShane, B., Gal, D., Gelman, A., Robert, C., and Tackett, J.,
Abandon Statistical Significance**

1. Treat p -values (and other purely statistical measures like confidence intervals and Bayes factors) continuously rather than in a dichotomous or thresholded manner. In doing so, bear in mind that it seldom makes sense to calibrate evidence as a function of p -values or other purely statistical measures because they are, among other things, typically defined relative to the generally uninteresting and implausible null hypothesis of zero effect and zero systematic error.

5. Accept uncertainty and embrace variation in effects: we can learn much (indeed, more) about the world by forsaking the false promise of certainty offered by dichotomous declarations of truth or falsity—binary statements about there being “an effect” or “no effect”—based on some p -value or other statistical threshold being attained.



RONALD L. WASSERSTEIN, ALLEN L. SCHIRM & NICOLE A. LAZAR
THE AMERICAN STATISTICIAN
2019, VOL. 73, NO. S1, 1–19: Editorial
<https://doi.org/10.1080/00031305.2019.1583913>

Moving to a World Beyond " $p < 0.05$ "

👉 Continuous scale mixtures of Gaussians (Carvalho et al, 2010, Polson et al 2012) do not lead to an immediate “selection” of relevant parameters

$$\beta_\nu \mid \tau, \lambda_\nu \sim \mathcal{N} (0, \tau^2 \cdot \lambda_\nu^2)$$

with

$\tau \sim g$ a global shrinkage parameter

and

$\lambda_\nu \sim h_\nu$ a local shrinkage parameter

- However, the decisions on the “significance” of the β coefficients are typically dichotomized (e.g., based on 90% credible intervals or shrinkage factor)
or other decision theoretic-based procedures (Chandra, Mueller, Sarkar, ArXiv, 2022+; Lee, Hussain, Warnick, et al, ArXiv, 2022+)

- We can consider a mixture:

$$\beta_\nu \mid \tau, \lambda_K, \pi, \sigma^2 \sim \sum_{k=1}^K \pi_k \phi(\beta_\nu; 0, \sigma^2 \cdot \tau^2 \cdot \lambda_k^2)$$

where λ_k^2 is a mixture shrinkage component.

The smallest variance component is typically such that $\tau \lambda_{(1)} \approx 0$ and represents the null distribution

The other components can be sorted according to the magnitudes of λ_k 's.

The alternative distribution gets segmented into different levels

- 👉 One can rank the β_ν 's into **tiers of relevance**

- We can consider a mixture:

$$\beta_\nu \mid \tau, \lambda_K, \pi, \sigma^2 \sim \sum_{k=1}^K \pi_k \phi(\beta_\nu; 0, \sigma^2 \cdot \tau^2 \cdot \lambda_k^2)$$

where λ_k^2 is a mixture shrinkage component.

The smallest variance component is typically such that $\tau \lambda_{(1)} \approx 0$ and represents the null distribution

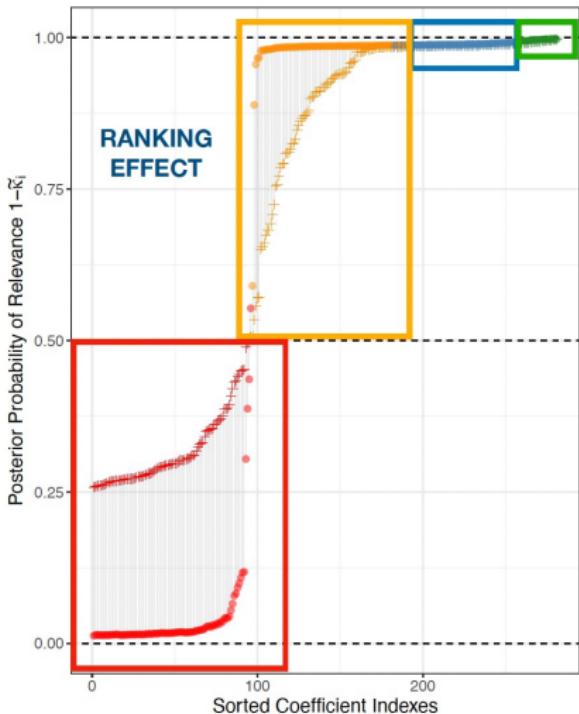
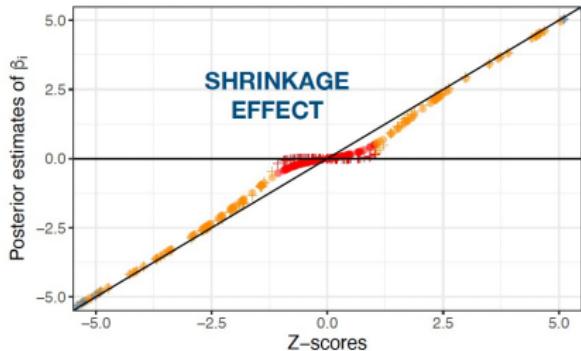
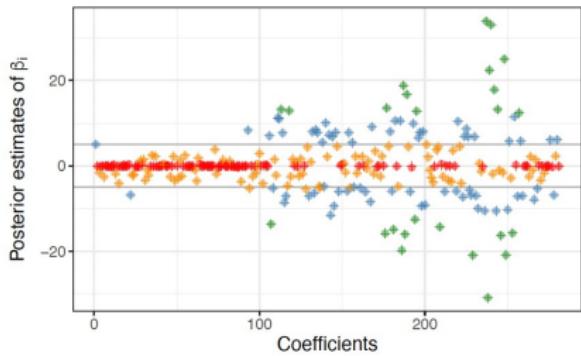
The other components can be sorted according to the magnitudes of λ_k 's.

The alternative distribution gets segmented into different levels

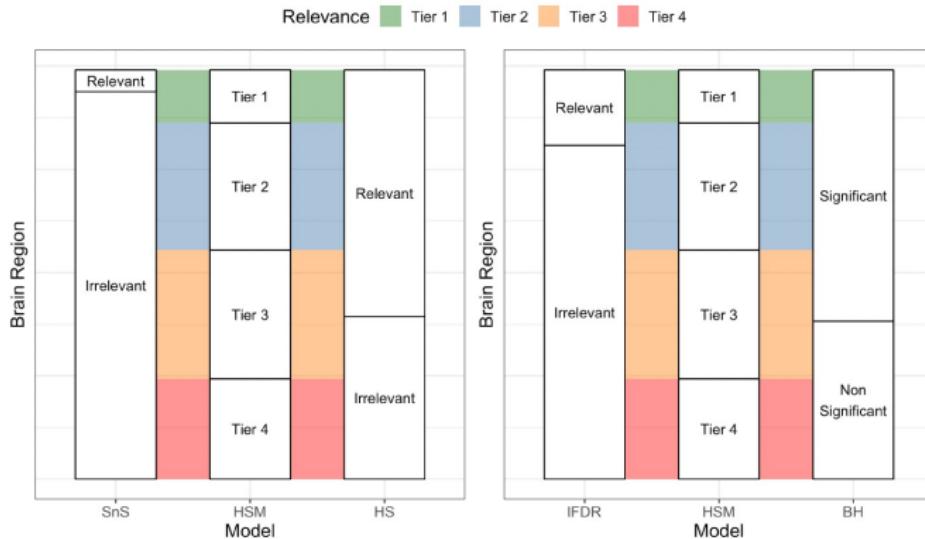
- 👉 One can rank the β_ν 's into **tiers of relevance**
- 👉 One can choose a **Half-Cauchy prior** for the mixture shrinkage component,

$$\lambda_l \sim \text{Cauchy}^+(0, 1), \forall l \quad (\text{Horseshoe pit})$$

We can segment the results into 4 tiers of activation, from high-activity (**Tier 1**) to no activity (**Tier 4**)



We compare the findings with other well-known methods: Local-FDR (IFDR), Horseshoe prior (HS), Spike-and-Slab (SnS), and Benjamini-Hochberg (BH)



- The HSM model mediates between the more conservative IFDR and SnS methods and the numerous discoveries of the BH and HS models.
- Denti et al (2022+), *A Horseshoe mixture model for Bayesian screening with an application to light sheet fluorescence microscopy in brain imaging*, Submitted. <https://arxiv.org/abs/2106.08281>

- ❑ The development of neuroimaging biomarkers for targeted interventions requires to take into account the complexity and heterogeneity of brain functioning
 - ⇒ Still a lot of work to do; many opportunities
- ❑ Statistical approaches play a crucial role in this quest
- ❑ Close collaboration with neuroscientists is essential
- ❑ Hierarchical Bayesian methods allow to elegantly borrow information across and within subjects
- ❑ Challenges: Computational scalability → dimension reduction techniques & approximate inference

Collaborators - Thanks



Laura D'Angelo
Universita' Bicocca, U. Padova, Italy



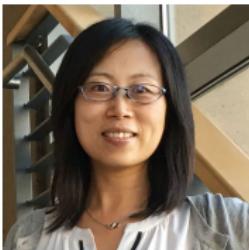
Francesco Denti
U. Cattolica, Milan, Italy



Babak, Shababa
UCI



Antonio Canale
U. Padova, Italy



Zhaoxia Yu
UCI



Norbert Fortin
Anatomy & Neurobiology
UCI