Clustering computer mouse tracking data via informed priors

Michele Guindani

Department of Biostatistics
University of California, Los Angeles

2025 Biostatistics Symposium of Southern California
Newport Beach, Feb 21st, 2025

Multi-view Clustering

- 👉 Datasets often present multivariate information collected across distinct domains, with possibly different features support spaces.

Multi-view Clustering

👉 Datasets often present multivariate information collected across distinct domains, with possibly different features support spaces.

- Disease Subtyping (e.g., Cancer)
  - **Genetic Data**: Gene expression, mutations
  - **Histopathological Images**: Microscopic tissue slides
  - **Clinical Data**: Patient age, symptoms, medical history

- 👉 Datasets often present multivariate information collected across distinct domains, with possibly different features support spaces.

- Disease Subtyping (e.g., Cancer)
  - **Genetic Data**: Gene expression, mutations
  - **Histopathological Images**: Microscopic tissue slides
  - **Clinical Data**: Patient age, symptoms, medical history

  - 👉 Identifying molecular subtypes of cancer for personalized treatment.
  - 👉 Multi-view modeling may try to identify group patients with similar genetic and clinical profiles.

- Integrating Multi-Omics Data in Precision Medicine
    - **Genomics**: DNA sequencing data
    - **Transcriptomics**: RNA expression levels
    - **Proteomics**: Protein concentration data
    - **Metabolomics**: Metabolic markers in blood

👉 Predicting drug response by combining different biological levels of information.

👉 Multi-view models improve patient stratification for targeted therapies.

UCLA

- Integrating Multi-Omics Data in Precision Medicine
  - **Genomics**: DNA sequencing data
  - **Transcriptomics**: RNA expression levels
  - **Proteomics**: Protein concentration data
  - **Metabolomics**: Metabolic markers in blood
  - 👉 Predicting drug response by combining different biological levels of information.
  - 👉 Multi-view models improve patient stratification for targeted therapies.
- Many other examples: Public Health Surveillance using hospital admissions, social media data, environmental data, etc.

Our Motivating data

Ultimate Goal: During a smoking-cessation intervention, we want to learn if an individual is at risk of relapse.

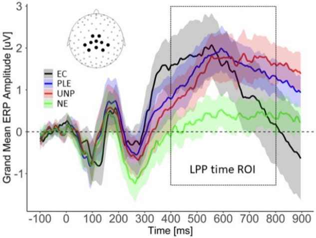
Neuroscientists may collect many data, including EEG data

- They may also collect behavioral data where smokers are asked to make decisions in front of cigarette-related cues and other cues in experiments.

👉 Thus, to understand decision-makings behind smoker's behaviors, neuroscientists end up collecting data across multiple modalities and views.

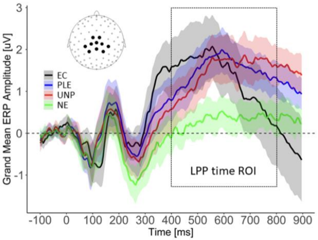
EEG data

- EEG experiment data
  event-related potentials
  (ERPs, a direct measure of brain activity)

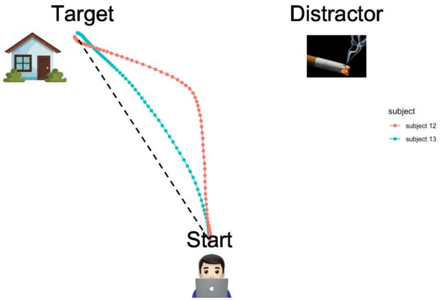


EEG data and computer-mouse tracking (CMT)

- EEG experiment data
  event-related potentials
  (ERPs, a direct measure of brain activity)

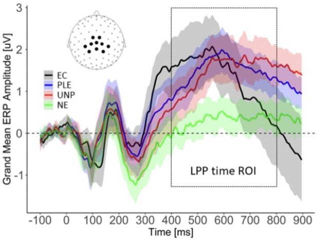


- computer-mouse trajectory tracking data
  (simpler and cheaper)

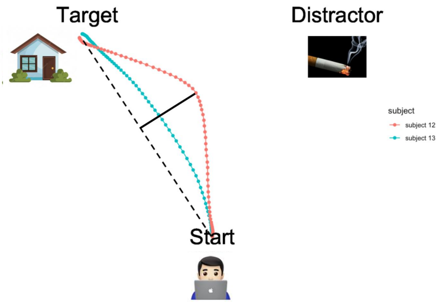


EEG data and computer-mouse tracking

- EEG experiment data
  event-related potentials
  (ERPs, a direct measure of brain activity)



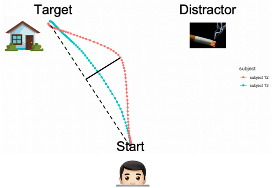
- computer-mouse trajectory tracking data
  (simpler and cheaper)



Trajectories are often summarized using
the maximum absolute deviation (MAD),
which captures the greatest point of deviation from the
idealized straight line

UCLA

Multiple-experiment conditions



	Target	v.s	Distractor	
cigarette-related (CP)		v.s		neutral (NP)
pleasant (PH)		v.s		neutral (NP)
unpleasant (UH)		v.s		neutral (NP)
neutral (NP)		v.s		cigarette-related (CP)
neutral (NP)		v.s		pleasant (PH)
neutral (NP)		v.s		unpleasant (UH)

- Analyzing a smoker's trajectories under different conditions helps us understand the individual's degree of response competition to alternatives and the individual's decision-making processes.

Goal

- Clustering Computer Mouse tracking across subject using the known clustering of subjects based on EEG data as reference

Goal

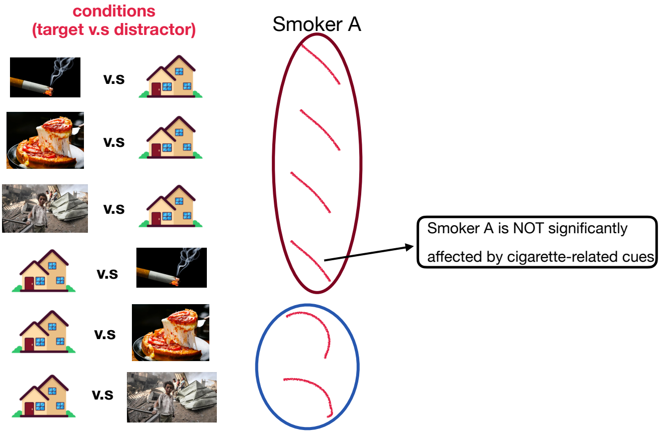
UCLA

- Clustering Computer Mouse tracking across subject using the known clustering of subjects based on EEG data as reference
- The idea is then to try to inform the probability of a partition of the data based on available information on the partition of another set of data
- We want to work on the space of partition
  - !! we may not have other information on the reference data other than a point estimate of their partition (no experiment, no data)
- In addition, we need to take into consideration the several conditions of the computer-mouse tracking experiment

Computer Mouse Tracking data

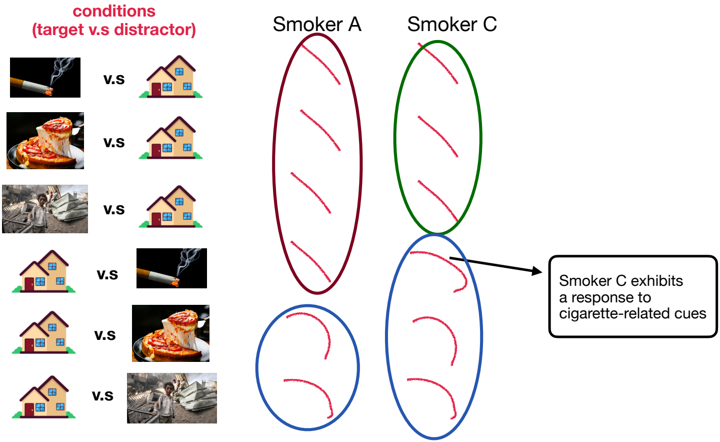
UCLA

A smoker's behavioral patterns can be revealed by clustering the smoker's mouse tracking trajectories across experimental conditions.

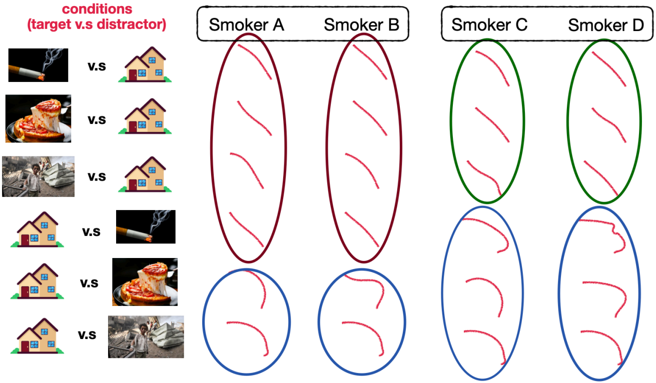


Computer Mouse Tracking data

A smoker's behavioral patterns can be revealed by clustering the smoker's mouse tracking trajectories across experimental conditions.



We are also interested in clustering the smokers to identify population subgroups with similar behavior patterns.



- ‼ The clusters across experimental conditions is typically more interesting than the actual values of the trajectory measurements
- 👉 Random partition models

‼️ The clusters across experimental conditions is typically more interesting than the actual values of the trajectory measurements

👉 Random partition models

🤔 It is reasonable to expect that a group of smokers is identified by partitions of conditions that are just more similar (rather than identical) within a group than between groups.

Goals:

1. Reveal patterns of the smoker's decision-making behaviors via clustering individual responses by experimental conditions;
2. Identify smokers' subgroups within the subgroups that exhibit similar behavioral patterns.
3. incorporate prior domain knowledge on the partitioning of smokers and experiment conditions

Computer Mouse Tracking data

UCLA

Prior Behavioral Neuroscience knowledge and expectations about the partitioning of either smokers or experiment conditions

ERP — Brain Activity

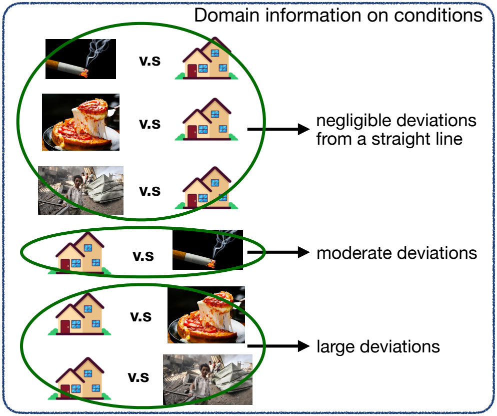
Domain information on 43 smokers

43 smokers

Group 1: less sensitive to cigarettes

Group 2: more sensitive to cigarettes

Computer Mouse Tracking



Possible Strategies: PPM

- A possible way to do this is to use product partition models (Hartigan 1990; Barry and Hartigan, 1992)

Possible Strategies: PPM

UCLA

- A possible way to do this is to use product partition models (Hartigan 1990; Barry and Hartigan, 1992)
- In product partition models (PPMs), we consider a partition $\rho = \{S_1, S_2, \ldots, S_k\}$, of a set of n units, where each S_j is a subset (cluster) of indices. The prior probability of the partition is given by

$$p(\rho) \propto \prod_{j=1}^{k} h(S_j),$$

where $h(S_j) \geq 0$ is the *cohesion function*.

The cohesion function quantifies how similar the elements within the subset S_j are considered a priori. In other words, it reflects our belief that items within the same cluster should be similar.

Possible Strategies: PPM

- A possible way to do this is to use product partition models (Hartigan 1990; Barry and Hartigan, 1992)
- In product partition models (PPMs), we consider a partition $\rho = \{S_1, S_2, \ldots, S_k\}$, of a set of n units, where each S_j is a subset (cluster) of indices. The prior probability of the partition is given by

$$p(\rho) \propto \prod_{j=1}^{k} h(S_j),$$

  where $h(S_j) \geq 0$ is the *cohesion function*.

- Park and Dunson (2010) and Müller, Quintana, and Rosner (2011) modified the cohesion functions in product partition models to allow for effects of covariates

- items closer in the covariate space have a higher probability of being grouped together a priori (nonexchangeable PPMs)

- Smith and Allenby (2020) introduce a location-scale partition (LSP) prior modifying the Polya-Urn scheme with similarity functions, incorporating the distance from a reference partition, say ρ_n

Possible Strategies: Shrinkage priors

Smith and Allenby (2020) introduce a location-scale partition (LSP) prior modifying the Polya-Urn scheme with similarity functions, incorporating the distance from a reference partition, say ρ_n

$$\theta_i \mid \theta_{<i}, \rho_n, \tau \sim w_0(\rho_n, \tau) G_0(\theta_i) + \sum_{k=1}^{K^{(i)}} w_k(\rho_n, \tau) \delta_{\phi_k^*}(\theta_i)$$

Here $w_0(\cdot)$ and $w_k(\cdot)$ are positive similarity functions that satisfy $w_0(\cdot) + \sum_k w_k(\cdot) = 1$. More specifically, they build a model for auxiliary variables

$$w_0(s_i, \tau) \propto \int \mathrm{Cat}\left(\xi_1, \ldots, \xi_{C^{(i)}+1}\right) \mathrm{Dir}\left(\tau_1, \ldots, \tau_{C^{(i)}+1}\right) d\boldsymbol{\xi} = \frac{\tau + 1\left(s_i = C^{(i)} + 1\right)}{\tau C^{(i)} + \tau + 1}$$

$$w_k(\{s_i, S_k\}, \tau) \propto \int \mathrm{Cat}\left(\xi_1, \ldots, \xi_{C^{(i)}+1}\right) \mathrm{Dir}\left(\tau_1^*, \ldots, \tau_{C^{(i)}+1}^*\right) d\boldsymbol{\xi} = \frac{\tau + n_{S_k}^{s_i}}{\tau C^{(i)} + \tau + n_k}$$

with $n_{S_k}^c$ counting the number of elements in S_k equal to c.

Possible Strategies: Shrinkage priors

- Paganin et al (2021) propose a centered partition process which defines the prior on set partitions as proportional to a baseline EPPF multiplied by a penalization term of the type

$$p\left(\boldsymbol{c} \mid \boldsymbol{c}_0, \psi\right) \propto p_0(\boldsymbol{c})\, e^{-\psi d(\boldsymbol{c}, \boldsymbol{c}_0)}$$

Possible Strategies: Shrinkage priors

- Paganin et al (2021) propose a centered partition process which defines the prior on set partitions as proportional to a baseline EPPF multiplied by a penalization term of the type

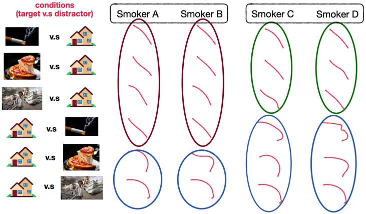
$$p\left(\boldsymbol{c} \mid \boldsymbol{c}_0, \psi\right) \propto p_0(\boldsymbol{c})\, e^{-\psi d(\boldsymbol{c}, \boldsymbol{c}_0)}$$

- Dhal, Warr and Jensen (2022, video) and Dhal, Warr and Jensen (arXiv:2312) propose dependent random partitions by shrinking toward an anchor partition

Shrinkage Partition Prior (Dhal, Warr and Jensen, 2022, 2023)

- $(I \times J)$ data matrix of entries $Y_{i,j}$'s, from J smokers and I experiment conditions.

- Each $y_{i,j}$ is a MAD computed for the trajectory drawn by smoker j under condition i, for $i = 1, \ldots, I$ and $j = 1, \ldots, J$.



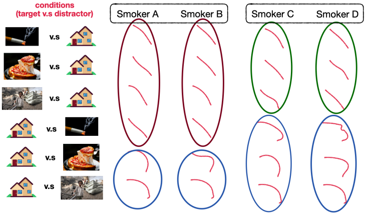
Shrinkage Partition Prior (Dhal, Warr and Jensen, 2022, 2023)

- $(I \times J)$ data matrix of entries $Y_{i,j}$'s, from J smokers and I experiment conditions.

- Each $y_{i,j}$ is a MAD computed for the trajectory drawn by smoker j under condition i, for $i = 1, \ldots, I$ and $j = 1, \ldots, J$.

- condition partition $\pi_j = \left\{ \pi_{1,j}, \ldots, \pi_{I,j} \right\}$: a partition of the I conditions MAD's for each smoker j

- subject partition $c = \left\{ c_1, \ldots, c_J \right\}$: a partition of the J smokers



- For any partition π, the probability function of a SP prior is

$$\pi \sim \mathrm{SP}\left(\pi \mid \nu, \lambda, \delta, p_{\mathrm{b}}\right)$$

Shrinkage Partition Prior (Dhal, Warr and Jensen, 2022, 2023)

UCLA

- For any partition π, the probability function of a SP prior is

$$\pi \sim \mathrm{SP}\left(\pi \mid \nu, \lambda, \delta, p_{\mathrm{b}}\right)$$

A base (anchor) partition suggested
by available expert information

Shrinkage Partition Prior (Dhal, Warr and Jensen, 2022, 2023)

UCLA

- For any partition π, the probability function of a SP prior is

$$\pi \sim \mathrm{SP}\left(\pi \mid \nu, \lambda, \delta, p_{\mathrm{b}}\right)$$

A base (anchor) partition suggested by available expert information

A nonnegative shrinkage parameter indicates how much we trust the base partition ν for inference

Shrinkage Partition Prior (Dhal, Warr and Jensen, 2022, 2023)

UCLA

- For any partition π, the probability function of a SP prior is

Permutation parameter

$$\pi \sim \mathrm{SP}\left(\pi \mid \nu, \lambda, \delta, p_{\mathrm{b}}\right)$$

A base (anchor) partition suggested by available expert information

A nonnegative shrinkage parameter indicates how much we trust the base partition ν for inference

Shrinkage Partition Prior (Dhal, Warr and Jensen, 2022, 2023)

UCLA

- For any partition π, the probability function of a SP prior is

Permutation parameter

$$\pi \sim \mathrm{SP}\left(\pi \mid \nu, \lambda, \delta, p_{\mathrm{b}}\right)$$

A random partition prior distribution, e.g., CRP or Pitman-Yor process

A base (anchor) partition suggested by available expert information

A nonnegative shrinkage parameter indicates how much we trust the base partition ν for inference

Shrinkage Partition Prior (Dhal, Warr and Jensen, 2022, 2023)

- For any partition π, the probability function of a SP prior is

Permutation parameter

$$\pi \sim \mathrm{SP}\left(\pi \mid \nu, \lambda, \delta, p_{\mathrm{b}}\right)$$

A random partition prior distribution, e.g., CRP or Pitman-Yor process

A base (anchor) partition suggested by available expert information

A nonnegative shrinkage parameter indicates how much we trust the base partition ν for inference

- If shrinkage parameter $\lambda \to \infty$, the simulated partition π is **shrinks towards and aligned with the base partition ν**, i.e., $\pi = \nu$;

Shrinkage Partition Prior (Dhal, Warr and Jensen, 2022, 2023)

UCLA

- For any partition π, the probability function of a SP prior is

$$\pi \sim \mathrm{SP}\left(\pi \mid \nu, \lambda, \delta, p_{\mathrm{b}}\right)$$

Permutation parameter

A random partition prior distribution, e.g., CRP or Pitman-Yor process

A base (anchor) partition suggested by available expert information

A nonnegative shrinkage parameter indicates how much we trust the base partition ν for inference

- If shrinkage parameter $\lambda \to \infty$, the simulated partition π is **shrinks towards and aligned with the base partition ν**, i.e., $\pi = \nu$;

- If shrinkage parameter $\lambda = 0$, the partition $\pi \sim p_{\mathrm{b}}$ and π is **independent of the base partition ν.**

The SP prior assigns the joint probabilities $P_{SP}(\pi)$ through a sequential allocation scheme:

$$\mathrm{Pr}_{\mathrm{SP}}(\pi_{\delta_t} = s \mid \pi_{\delta_1}, \ldots, \pi_{\delta_{t-1}}, \nu_{\delta_1}, \ldots, \nu_{\delta_t}, \lambda, p_{\mathrm{b}})$$
$$\propto \begin{cases} \mathrm{Pr}_{\mathrm{b}}\left(\pi_{\delta_t} = s \mid \pi_{\delta_1}, \ldots, \pi_{\delta_{t-1}}\right) \times & \exp\left(\lambda \, \frac{\sum_{k=1}^{t-1} \mathrm{I}\left\{\pi_{\delta_k} = s\right\} \mathrm{I}\left\{\nu_{\delta_k} = \nu_{\delta_t}\right\}}{\sum_{k=1}^{t-1} \mathrm{I}\left\{\nu_{\delta_k} = \nu_{\delta_t}\right\}}\right) \\ & \text{if } s \in \left\{\pi_{\delta_1}, \ldots, \pi_{\delta_{t-1}}\right\}, \\ \mathrm{Pr}_{\mathrm{b}}\left(\pi_{\delta_t} = s \mid \pi_{\delta_1}, \ldots, \pi_{\delta_{t-1}}\right) \times & \exp\left(\lambda \, \mathrm{I}\left\{\sum_{k=1}^{t-1} \mathrm{I}\left\{\nu_{\delta_k} = \nu_{\delta_t}\right\} = 0\right\}\right) \\ & \text{if } s \notin \left\{\pi_{\delta_1}, \ldots, \pi_{\delta_{t-1}}\right\}, \end{cases}$$

$t = 1, \ldots, n.$

Typically, $\lambda_t = \lambda$, $t = 1, \ldots, n$.

For large λ, the SP distribution favors partitions drawn from the CRP that are very similar to the base partition. This behavior is induced by the exponential term, which encourages alignment with the base partition

Hierarchical Shrinkage prior

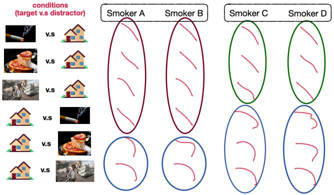
Hierarchical Shrinkage prior

base partition of smokers
given by ERPs results

shrinkage
parameter

smokers
clustering

$$c \sim \mathrm{SP}\left(c \mid c_0, \tau, \zeta, \mathrm{CRP}\left(\alpha_0\right)\right)$$



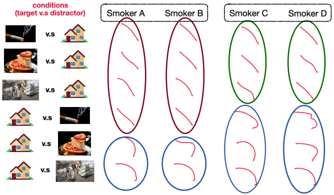
Hierarchical Shrinkage prior

base partition of smokers given by ERPs results

shrinkage parameter

smokers clustering

$$c \sim \text{SP}\left(c \mid c_0, \tau, \zeta, \text{CRP}\left(\alpha_0\right)\right)$$



experiment conditions clustering for each smoker

$$\pi_j \sim \text{SP}\left\{\pi_j \mid \nu_j, \lambda, \delta_j, \text{CRP}(\beta)\right\}$$

base condition partition for smoker $j = 1, \ldots, J$

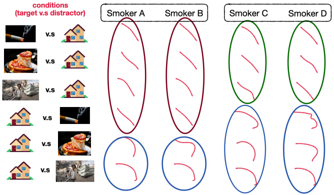
Hierarchical Shrinkage prior

base partition of smokers
given by ERPs results

shrinkage
parameter

smokers
clustering

$$c \sim \mathrm{SP}\left(c \mid c_0, \tau, \zeta, \mathrm{CRP}\left(\alpha_0\right)\right)$$



Smokers within a group share a common base partition of conditions ν_k^*

$$\nu_j = \nu_{j'} = \nu_k^* \text{ if } c_j = c_{j'}$$

If smoker j and j' are in the same group according to c, we anticipate they exhibit similar behaviors π_j and $\pi_{j'}$, so we let them share an identical base partition of conditions, i.e., $\nu_j = \nu_{j'} = \nu_k^*$

experiment conditions
clustering for each
smoker

$$\pi_j \sim \mathrm{SP}\left\{\pi_j \mid \nu_j, \lambda, \delta_j, \mathrm{CRP}(\beta)\right\}$$

base condition partition for smoker $j = 1,\ldots,J$

Hierarchical Shrinkage prior

base partition of smokers
given by ERPs results

shrinkage
parameter

smokers
clustering

$$c \sim \mathrm{SP}\left(c \mid c_0, \tau, \zeta, \mathrm{CRP}\left(\alpha_0\right)\right)$$

Smokers within a group share a common base partition of conditions ν_k^*

$$\nu_j = \nu_{j'} = \nu_k^* \text{ if } c_j = c_{j'}$$

$$\nu_k^* \sim \mathrm{SP}\left(\nu_k^* \mid \nu_0, \rho, \epsilon_k^*, \mathrm{CRP}(\beta_0)\right)$$

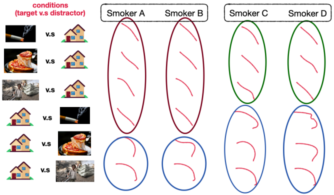
base condition partition
given by collaborator
on population level

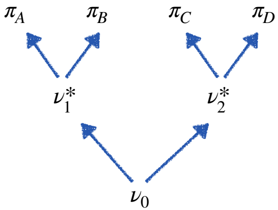
how much we trust the
base condition partition ν_0

experiment conditions
clustering for each
smoker

$$\pi_j \sim \mathrm{SP}\left\{\pi_j \mid \nu_j, \lambda, \delta_j, \mathrm{CRP}(\beta)\right\}$$

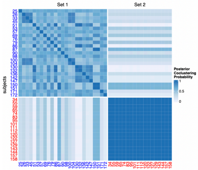
base condition partition for smoker $j = 1, \ldots, J$

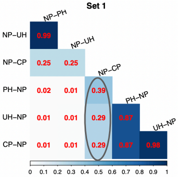


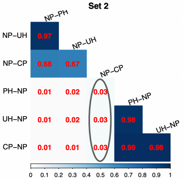


- **Strong constrain**: shrink the estimated partitions of subjects towards the ERP groups

- **Strong constrain**: shrink the estimated partitions of subjects towards the ERP groups

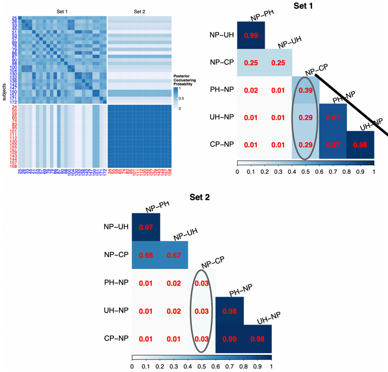


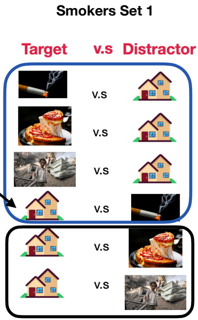




clustering of subjects and mouse-tracking experimental conditions

- **Strong constrain**: shrink the estimated partitions of subjects towards the ERP groups

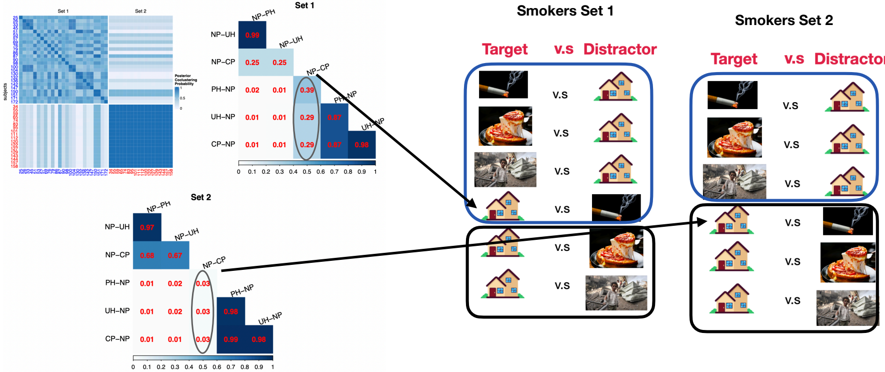




clustering of subjects and mouse-tracking experimental conditions

Exploratory Data Analysis

- **Strong constrain**: shrink the estimated partitions of subjects towards the ERP groups

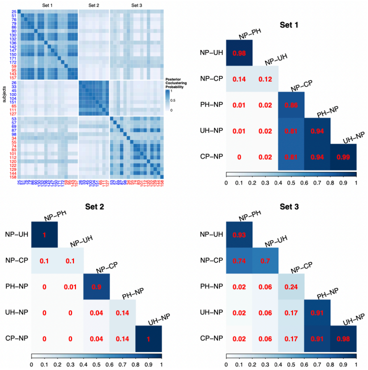


clustering of subjects and mouse-tracking experimental conditions

- **No ERP information** taken into account

UCLA

- **No ERP information** taken into account

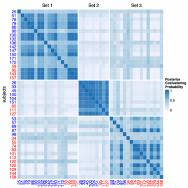


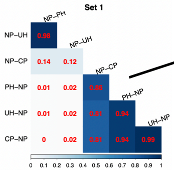
clustering of subjects and mouse-tracking experimental conditions

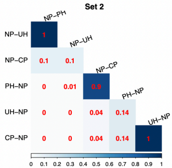
Exploratory Data Analysis

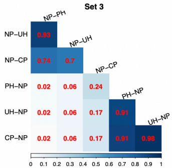
UCLA

- **No ERP information** taken into account



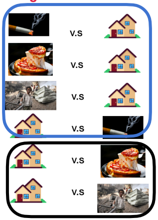






Smokers Set 1

Target v.s Distractor

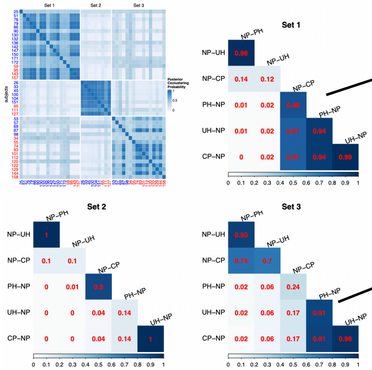


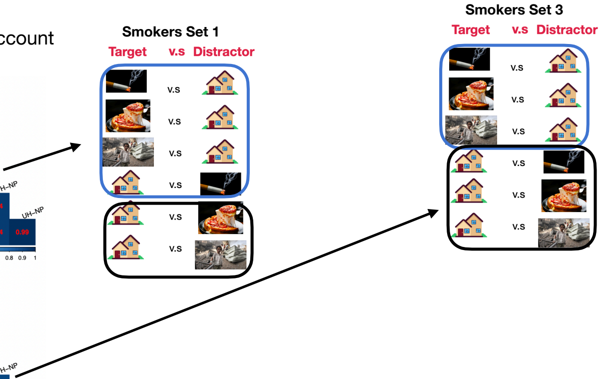
clustering of subjects and mouse-tracking experimental conditions

Exploratory Data Analysis

UCLA

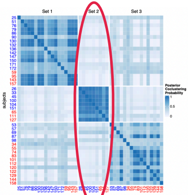
- **No ERP information** taken into account

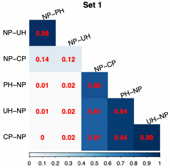


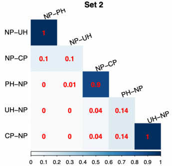


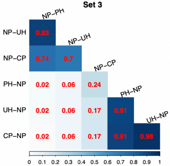
clustering of subjects and mouse-tracking experimental conditions

- **No ERP information** taken into account





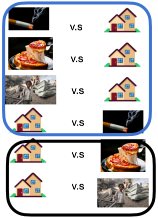




clustering of subjects and mouse-tracking experimental conditions

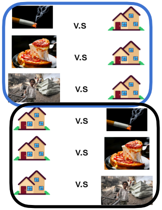
Smokers Set 1

Target v.s Distractor



Smokers Set 3

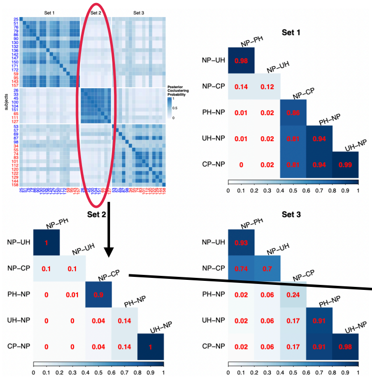
Target v.s Distractor



Exploratory Data Analysis

UCLA

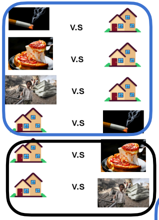
- **No ERP information** taken into account



clustering of subjects and mouse-tracking experimental conditions

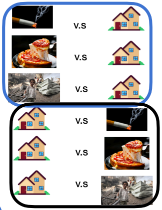
Smokers Set 1

Target v.s Distractor



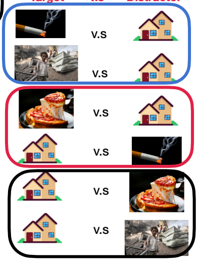
Smokers Set 3

Target v.s Distractor



Smokers Set 2

Target v.s Distractor



Conclusions

- Incorporate valuable expert information through the base partitions for rows and columns.
- It is important to understand how much either base partition should be trusted.
  - E.g., in considering a base partition of rows, it is important to account for the degree of heterogeneity across all subjects in the original data.

Conclusions

UCLA

- Incorporate valuable expert information through the base partitions for rows and columns.
- It is important to understand how much either base partition should be trusted.
  - E.g., in considering a base partition of rows, it is important to account for the degree of heterogeneity across all subjects in the original data.
- Extensions:
  - Move beyond CRP for the baseline partition distributions;
  - Priors over the shrinkage parameters;
  - Directly model the curve trajectories to capture more complex patterns, such as velocity, acceleration, or curvature.
- Ziyi Song, Weining Shen, Marina Vannucci, Alexandria Baldizon, Paul M Cinciripini, Francesco Versace, Michele Guindani, (2024) Clustering computer mouse tracking data with informed hierarchical shrinkage partition priors, Biometrics