Brain-Inspired Perspective on Configurations: Unsupervised Similarity and Early Cognition

Juntang Wang†, Yihan Wang†, Hao Wu†, Dongmian Zou, Shixin Xu

Duke Kunshan University, Kunshan, China
Sichuan University, Sichuan, China

BICS, November 30, 2025

A Central Goal in AI

Learning representations as humans do

has long been a central goal in AI and cognitive science.

The Question

Can machines discover structure without labels, like infants do?

The Challenge: Learning Like Infants

Infants have three key abilities:

1. **Unsupervised Organization**
   - Form categories without labels
2. **Hierarchical Flexibility**
   - Both superordinate and basic-level
   - Rapid adaptation/meta-learning
3. **Novelty Sensitivity**

(3-4 months old infants; According to Quinn and Eimas [1], Behl-Chadha [2], and Mareschal and Quinn [3])

Modern ML often struggles with:

- Labels or engineered proxy tasks
- Lack of hierarchical categories or spontaneous novelty detection

Key Gap

Without human-like, label-free structure discovery, ML systems struggle with open-world generalization, out-of-distribution shifts, and evolving categories.

What's Missing?

Some existing methods aim to replicate the three abilities, but often struggle with:

(clustering methods, contrastive/self-supervised pipelines)

- Multiple method-specific hyper-parameters
  (pre-specification of granularity in k-means, scale parameters in DBSCAN, or linkage choices in hierarchical clustering)
- Not designed to signal novelty or to track evolving categories
- Lack of a single global parameter that coherently sets granularity

What We Need

A simple framework with **one finite-resolution parameter** supporting all three abilities.

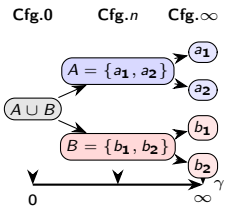
Recap of Configurations Framework

(Liu et al. [4] and Pitsianis et al. [5])

Having n data items, use a vector of cluster indices $\boldsymbol{\omega} \in \{0, 1, \ldots, k\}^n$ to denote a "good" partition. Use a matrix $\boldsymbol{\Omega} = \{\boldsymbol{\omega}_i\}_{i=1}^m$ to denote m such partitions, with $i \uparrow$ as granularity $\uparrow$.

Definition

When m is finite, each $\boldsymbol{\omega}$ is called a *configuration* (Cfg.), and $\boldsymbol{\Omega}$ is called *configurations*. When there exists a single parameter $\gamma \in [0, \infty)$ that controls granularity, we denote the configuration at γ as $\boldsymbol{\Omega}_\gamma$.



Like a camera zoom: One dial controls the level of detail

Recap of Configuration Framework (cont.)

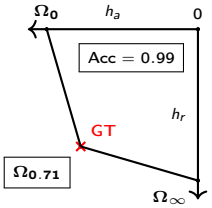
Definition

The *Hamiltonian energy* $\downarrow$ of a partition $\boldsymbol{\omega}$ is defined as:

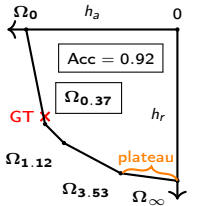
$$H(\boldsymbol{\omega}) = \underbrace{-\sum_{k=1}^{|\boldsymbol{\omega}|_\infty} \sum_{i<j} w_{ij}^{+}\, 1_{\omega_i=\omega_j=k}}_{h_a} + \underbrace{\gamma \sum_{k=1}^{|\boldsymbol{\omega}|_\infty} \sum_{i<j} w_{ij}^{-}\, 1_{\omega_i=k}}_{h_r}. \quad (1)$$

w_{ij}^{+}/w_{ij}^{-}: pairwise similarities/dissimilarities.
$\gamma \in [0, \infty)$: the resolution parameter controlling granularity.

- Uses Leiden (column-stochastic) as inner step. Efficient ($O(n)$) on sparse graphs.
- Scans γ, finds *stability plateaus*: ω remains under small perturbations to γ
- Parallel-DT (Pitsianis et al. [5]): discovers chain of stable configurations (plateaus)
- Multiple local searches run in parallel for efficiency
- Produces a set of representative stable partitions covering all granularity levels



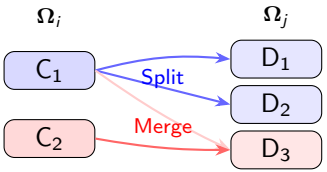
MNIST



WHUHi

Brain-Inspired Properties of Configurations

1. **Hierarchical Organization** ✓
   - Low γ: superordinate (animals vs. vehicles)
   - High γ: basic-level (cats vs. dogs)
   - Same input, different organizations
2. **Energy-Based Novelty Detection** ✓
   - Dissimilar items increase both h_a and $h_r \rightarrow$ higher energy
   - Intrinsic signal from same similarity principles driving organization
   - No separate detector needed
3. **Flexible Adaptation** ✓
   - Merge-split (not rigid trees)
   - Merges: attraction combines groups
   - Splits: repulsion breaks clusters
   - Context-dependent reorganization



Before We Test: An Evaluation Challenge

Problem: Traditional metrics struggle with multi-resolution clustering

- Unmatched cluster counts and arbitrary label permutations
- Semantically meaningful merge-split evolution not captured

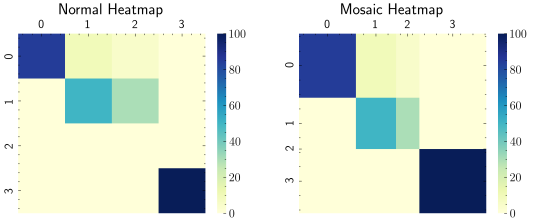
mheatmap framework

- **Mosaic heatmap**: proportional, interpretable heatmaps
- **RMS**: label alignment that respects merge-split dynamics

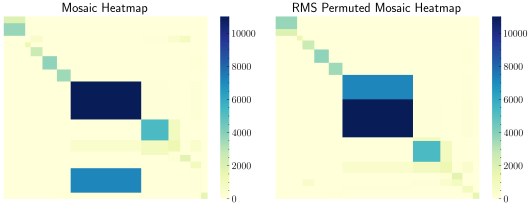
- **Cell width:** proportional to N_{ij}/r_i (fraction of ground-truth category i)
- **Cell height:** proportional to N_{ij}/c_j (fraction of predicted cluster j)
- **Cell color:** proportional to N_{ij} (magnitude, as in ordinary heatmaps)

mheatmap in Action

Mosaic vs. ordinary heatmap



RMS alignment (Salinas)



RMS improves diagonal structure and metrics without changing the underlying clustering.

Let's See If This Works in Practice

Our Evaluation

We test configurations on real-world datasets to see if they:

1. Achieve **competitive performance on standard clustering metrics**
2. Exhibit the three key brain-inspired cognitive abilities
3. Align with infant behavior patterns

Key Question: Can configurations be both competitive AND brain-inspired?

Clustering Performance

Method	Salinas		InfantS		ImageNet	
	ARI	NMI	ARI	NMI	ARI	NMI
K-means	0.45±0.04	0.55±0.03	0.22±0.03	0.32±0.02	0.18±0.03	0.28±0.02
GMM	0.85±0.02	0.88±0.01	0.25±0.02	0.35±0.02	0.21±0.02	0.31±0.02
Spectral	0.50±0.03	0.60±0.02	0.26±0.02	0.36±0.02	0.22±0.02	0.32±0.02
Agglomerative	0.43±0.04	0.53±0.03	0.21±0.03	0.31±0.02	0.17±0.03	0.27±0.02
DBSCAN	0.38±0.06	0.48±0.04	0.19±0.04	0.29±0.03	0.15±0.04	0.25±0.03
Mean Shift	0.41±0.03	0.51±0.03	0.20±0.02	0.30±0.02	0.16±0.03	0.26±0.02
Birch	0.39±0.04	0.49±0.03	0.18±0.03	0.28±0.02	0.14±0.03	0.24±0.02
Configurations	0.92±0.01	0.94±0.01	0.55±0.02	0.58±0.02	0.62±0.02	0.68±0.01

Competitive on standard metrics: Salinas (0.92), Infant (0.55), ImageNet (0.62)
Outperforming baselines by 7–49 points, **while also** exhibiting brain-inspired behaviors.

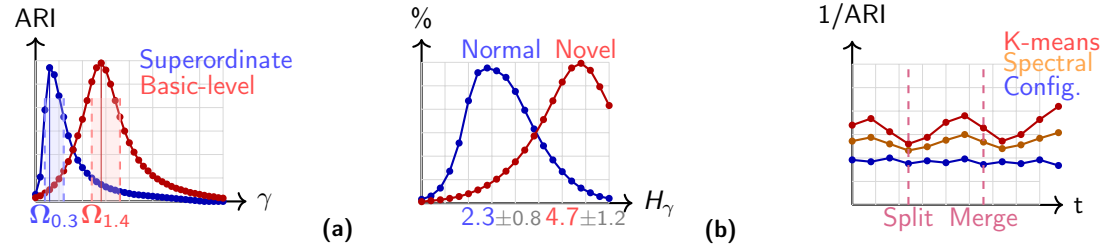
Infant Behavior Validation

Method	Internal	External	Both
K-means	0.18±0.02	0.22±0.02	0.55±0.03
GMM	0.21±0.01	0.25±0.02	0.58±0.02
Spectral Clustering	0.19±0.02	0.24±0.02	0.56±0.02
Agglomerative	0.20±0.02	0.23±0.02	0.57±0.03
Configurations	0.15±0.01	0.19±0.01	**0.65**±0.02
Infant behavior (Quinn et al., 1996):			
Infant 3-4 months	Poor	Poor	Good
Recent computational validation:			
ML prediction models [6]	0.18±0.02	0.23±0.02	0.61±0.02
Early word learning models [7]	0.16±0.01	0.21±0.01	0.58±0.02

Cognitive Alignment

Following Quinn and Eimas [1], configurations best align with infant behavior: Poor on isolated cues (ARI=0.15-0.19), strong when combined (ARI=0.65)

The Three Abilities in Action



Configurations reproduce hallmarks of early cognition—hierarchical selectivity, novelty sensitivity, and graceful adaptation—**while remaining competitive on standard metrics**.

What We Learned

Can machines learn like infants?

Yes—with configurations!

The Three Abilities Achieved

- ✓ **Hierarchical**: Multi-level via single parameter
- ✓ **Novelty**: 87% AUC, paralleling infant habituation
- ✓ **Adaptation**: 35% better stability

Key Insight

Mathematical structure captures cognitive principles—no engineering needed.

Future Directions

- **Neural implementation**: Hebbian plasticity, neural architectures
- **Scalability**: larger datasets, broader cognitive domains
- **Developmental modeling**: emergence of cognitive abilities over time
- **Applications**: educational technology, human-computer interfaces, developmental robotics

Code and Resources

mheatmap framework available at:
`https://github.com/qqgjyx/mheatmap`

References I

[1] P. C. Quinn and P. D. Eimas. "Perceptual Cues That Permit Categorical Differentiation of Animal Species by Infants". In: *Journal of Experimental Child Psychology* 63.1 (Oct. 1996), pp. 189–211. PMID: 8812045.

[2] G. Behl-Chadha. "Basic-Level and Superordinate-like Categorical Representations in Early Infancy". In: *Cognition* 60.2 (Aug. 1, 1996), pp. 105–141.

[3] D. Mareschal and P. C. Quinn. "Categorization in Infancy". In: *Trends in Cognitive Sciences* 5.10 (Oct. 1, 2001), pp. 443–450.

[4] T. Liu, D. Floros, N. Pitsianis, and X. Sun. "Digraph Clustering by the BlueRed Method". In: *2021 IEEE High Performance Extreme Computing Conference (HPEC)*. 2021 IEEE High Performance Extreme Computing Conference (HPEC). Sept. 2021, pp. 1–7.

[5] N. Pitsianis, D. Floros, T. Liu, and X. Sun. "Parallel Clustering with Resolution Variation". In: *2023 IEEE High Performance Extreme Computing Conference (HPEC)*. 2023 IEEE High Performance Extreme Computing Conference (HPEC). Boston, MA, USA: IEEE, Sept. 25, 2023, pp. 1–8.

References II

[6] S. Gibbon, A. Attaheri, Á. Ní Choisdealbha, S. Rocha, P. Brusini, N. Mead, P. Boutris, H. Olawole-Scott, H. Ahmed, S. Flanagan, K. Mandke, M. Keshavarzi, and U. Goswami. "Machine Learning Accurately Classifies Neural Responses to Rhythmic Speech vs. Non-Speech from 8-Week-Old Infant EEG". In: *Brain and Language* 220 (Sept. 1, 2021), p. 104968.

[7] S. Tsutsui, A. Chandrasekaran, M. A. Reza, D. Crandall, and C. Yu. *A Computational Model of Early Word Learning from the Infant's Point of View*. June 4, 2020. arXiv: 2006.02802 [cs]. URL: http://arxiv.org/abs/2006.02802 (visited on 09/10/2025). Pre-published.

[8] R. L. Fantz. "Visual Experience in Infants: Decreased Attention to Familiar Patterns Relative to Novel Ones". In: *Science* 146.3644 (1964), pp. 668–670. JSTOR: 1714550.

Thank You!

Questions?

Presenter email: jw853@duke.edu