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

Juntang Wang†[1][0009−0008−9812−7451], Yihan Wang†[1][0009−0000−3586−9046], Hao Wu†[2][0009−0009−5792−3790], Dongmian Zou[1][0000−0002−5618−5791], and Shixin Xu⋆[1][0000−0002−8207−7313]

[1] Duke Kunshan University, Kunshan, China
[2] Sichuan University, Sichuan, China

Abstract. Infants discover categories, detect novelty, and adapt to new contexts without supervision—a challenge for current machine learning. We present a brain-inspired perspective on *configurations* [1, 2], a finite-resolution clustering framework that uses a single resolution parameter and attraction–repulsion dynamics to yield hierarchical organization, novelty sensitivity, and flexible adaptation. To evaluate these properties, we introduce `mheatmap`, which provides proportional heatmaps and reassignment algorithm to fairly assess multi-resolution and dynamic behavior. Across datasets, configurations are competitive on standard clustering metrics, achieve 87% AUC in novelty detection, and show 35% better stability during dynamic category evolution. These results position configurations as a principled computational model of early cognitive categorization and a step toward brain-inspired AI.

Keywords: Brain-inspired cognition · Clustering · Configurations

1 Introduction

Learning representations as humans do has long been a central goal in AI and cognitive science. Humans and non-human animals can discover structure without labels. Infants form categories (e.g., animals vs. vehicles; cats vs. dogs) and detect novelty long before language [3, 4, 5], and similar unsupervised sensitivities are observed in animals [6]. These findings indicate an early, hierarchical organization of experience into similarity-based groups that support recognition, generalization, and novelty detection. In sharp contrast, modern ML struggles to discover such structure without supervision: supervised models depend on massive labeled datasets [7, 8], and self-supervised or contrastive methods learn via engineered proxy tasks [9, 10, 11]. These pipelines rarely yield human-like hierarchical categories or spontaneous novelty responses without explicit supervision or carefully engineered tasks [12]. This gap matters: without human-like, label-free structure discovery, ML systems struggle with open-world generalization, out-of-distribution shifts, and evolving categories.

⋆ Corresponding author: `shixin.xu@dukekunshan.edu.cn`

Existing unsupervised representation learning and clustering methods aim to replicate this ability—k-means, spectral, agglomerative, density-based clustering [13, 14, 15, 16] and contrastive/self-supervised pipelines [9, 10, 11]. However, they typically expose multiple method-specific hyper-parameters; some require pre-specifying the number of clusters (e.g., k-means), others rely on scale parameters or linkage choices (e.g., DBSCAN, hierarchical). They are not generally designed to signal novelty or to track evolving categories, and they often lack a single global parameter that coherently sets granularity across the dataset.

What is missing is a simple framework with one finite resolution parameter that supports the three human-like abilities highlighted above: hierarchical organization, novelty sensitivity, and flexible adaptation.

In this paper, we provide a brain-inspired perspective on *configurations*, a recent finite-*resolution* clustering framework [1, 2]; we formalize it next in Section 3. We further present `mheatmap`, an evaluation tool designed to visualize generalization dynamics and quantitatively assess the emergence of these human-like clustering abilities in Section 4. We evaluate on synthetic and real-world datasets in Section 5 and find that configurations reproduce hallmarks of early cognition—hierarchical selectivity, novelty sensitivity, and graceful adaptation under evolving categories—while remaining competitive on standard clustering metrics. Our contributions are:

(1) A conceptual connection between early cognition and *configurations*.
(2) `mheatmap`: proportional visualization and a reassignment metric for dynamic clustering analysis.
(3) An empirical study showing that configurations express brain-inspired clustering behaviors not captured by standard baselines.

2 Background and Motivation

Habituation studies show that 3–4-month-old infants form categories without labels and prefer novelty, with effects observed at both superordinate (e.g., animals vs. vehicles) and basic levels (e.g., cats vs. dogs) [3, 4, 5, 17]. Computational work further supports hierarchical sensitivity in early perception [18] and rapid adaptation/meta-learning in infancy [19, 20]. We take three empirically grounded targets from this literature—unsupervised organization, hierarchical flexibility, and novelty sensitivity—as operational desiderata for our model.

Modern ML typically relies on labels or engineered proxy tasks [9, 10, 11, 12], and classical clustering fixes granularity or exposes many method-specific knobs without a single global control that lives in a finite space [13, 15, 16]. Critically, most methods neither signal novelty nor handle evolving categories in a principled way. We therefore seek a finite-resolution formulation with one parameter that governs granularity while supporting novelty and dynamic adaptation.

Static metrics such as ARI, NMI, and V-measure [21, 22, 23, 24] do not address three core issues in multi-resolution settings: unmatched cluster counts across resolutions, arbitrary label permutations, and semantically meaningful merge–split evolution. We introduce `mheatmap` to fill this gap with proportional visualization and a stability measure (1/ARI) that enable fair, cognitively aligned evaluation across resolutions and over time.

3 Configurations: Def. and Brain-Inspired Perspective

Building on our motivation for brain-inspired clustering, we now address the multi-granularity challenge through a finite-resolution framework that naturally bridges cognitive science and machine learning. Traditional clustering approaches either require pre-specification of granularity [3] or lack a single finite-space parameter that controls granularity [4], failing to capture the flexible, context-dependent organization observed in early cognition without "supervision" [5]. As one possible solution, we turn to *configurations*, a recent finite-γ clustering framework [1, 2].

3.1 Recap of Configuration Definition

Let there be n data items, the goal of hierarchical clustering is to find all "good" partitions, each of them can be denoted with a vector of cluster indices $\boldsymbol{\omega} \in \{0, 1, \ldots, k\}^n$. Suppose there are m such partitions, we denote them as a matrix $\boldsymbol{\Omega} = \{\boldsymbol{\omega}_i\}_{i=1}^m$, with i increasing as granularity increases.

Definition 1. *When m is finite, each $\boldsymbol{\omega}$ is called a* configuration *(Cfg.), and $\boldsymbol{\Omega}$ is called* configurations*. In this paper, when there exists a single parameter $\gamma \in [0, \infty)$ that controls granularity, we denote the configuration at γ as $\boldsymbol{\Omega}_\gamma$.*

Then an obvious proposition follows by the definition itself:

Proposition 1. *There always exist two special configurations: $\boldsymbol{\omega}_0 := \boldsymbol{\Omega}_0$ is the coarsest configuration, where all items are in the same cluster. $\boldsymbol{\omega}_\infty := \boldsymbol{\Omega}_\infty$ is the finest configuration, where each item is in a separate cluster.*

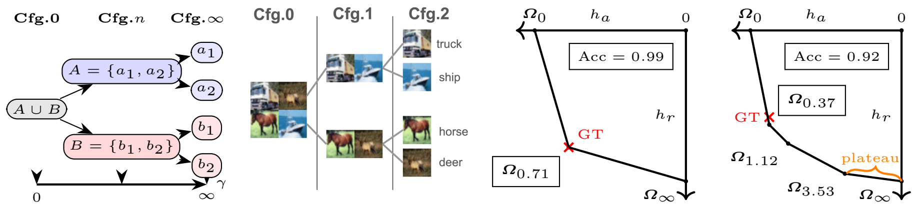
We present two illustrative examples in Figs. 1a and 1b, with each clustering a set of entities $\{a_1, a_2, b_1, b_2\}$ and entities from CIFAR-10 [26], respectively.

However, we have not yet specified (1) how to define the "good" partitions, (2) how to make m finite and (3) how to have a single parameter γ that controls granularity. Thanks to Pitsianis et al. [2], we can answer these questions with an energy function.

Definition 2. *The* Hamiltonian energy *of a partition $\boldsymbol{\omega}$ is defined as:*

$$H(\boldsymbol{\omega}) \;=\; \underbrace{-\sum_{k=1}^{|\boldsymbol{\omega}|_\infty} \sum_{i<j} w_{ij}^{+}\, \mathbf{1}_{\boldsymbol{\omega}_i = \boldsymbol{\omega}_j = k}}_{h_a} \;+\gamma \underbrace{\sum_{k=1}^{|\boldsymbol{\omega}|_\infty} \sum_{i<j} w_{ij}^{-}\, \mathbf{1}_{\boldsymbol{\omega}_i = k}}_{h_r} . \tag{1}$$

[3] e.g., k-means [13] requires the number of clusters k.
[4] e.g., Leiden and Louvain methods [25] require a resolution parameter $\gamma \in [0, \infty)$.
[5] Selection of good parameter values, such as k or γ.



(a) Lineage (schem.) (b) Lineage (CIFAR) (c) Energy (MNIST) (d) Energy (WHU)

Fig. 1: Configuration lineage and energy landscapes. **(a)&(b)** Configuration lineages: γ controls hierarchical granularity from coarse to fine. **(c)&(d)** Energy landscapes: Axes show attraction h_a and repulsion h_r.

where $w_{ij}^{+} \geq 0$ and $w_{ij}^{-} \geq 0$ are pairwise attraction and repulsion weights, larger for similar and dissimilar pairs, respectively. These weights can be derived from graph-based similarities (e.g., kNN with stochastic reweighting [2]), learned embeddings, or application-specific affinity measures. $\gamma \in [0, \infty)$ is the resolution parameter controlling granularity: small γ favors coarse groupings (attraction dominates), large γ favors fine partitions (repulsion dominates). A partition is considered better than another if its energy is lower.

Proposition 2. *When w_{ij}^{+} and w_{ij}^{-} are derived from graph-based similarities (i.e., edge weights of a graph), minimizing $H(\boldsymbol{\omega})$ is equivalent to maximizing the modularity as defined in Blondel et al. [27].*

Actually, the implementation of Pitsianis et al. [2] (also the implementation of this paper), when γ is fixed, is nothing big but the Leiden method [25] on k nearest neighbor graph, with some minor modifications. As the graph is sparse with a small k, the runtime is essentially linear in n, i.e., $O(n)$. However, it can be exhaustive to search $\gamma \in [0, \infty)$ for "best" configurations. Luckily, Pitsianis et al. [2] provides a solution called *Parallel-DT*. Parallel-DT is guaranteed to find $m+1$ segments of $(0, \infty)$ with dominant $\boldsymbol{\Omega}_i$ at the division points, and such segments are called *plateaus*. Combining all $\boldsymbol{\Omega}_i$, we get the desired $\boldsymbol{\Omega}$.

We provide two illustrative examples of such $\boldsymbol{\Omega}$ on the *energy landscapes* (2D plots of h_a and h_r) in Figs. 1c and 1d on the MNIST [28] and WHU-Hi-Hanchuan [29] datasets, respectively. These two examples also demonstrate (1) validity and high accuracy of the configuration framework by ground truth (GT) lying in the front-end and high accuracy (Acc) values, and (2) an example of the plateaus.

Design and proofs of the whole $\boldsymbol{\Omega}$ framework, including Parallel-DT, are not our focus, and we refer the reader to Liu et al. [1] and Pitsianis et al. [2] for more details.

3.2 Brain-Inspired Properties of Configurations

The mathematical structure of configurations naturally embodies the three cognitive desiderata from infant studies. We demonstrate how unsupervised organization, hierarchical flexibility, and novelty sensitivity emerge directly from the attraction–repulsion dynamics and energy minimization framework.

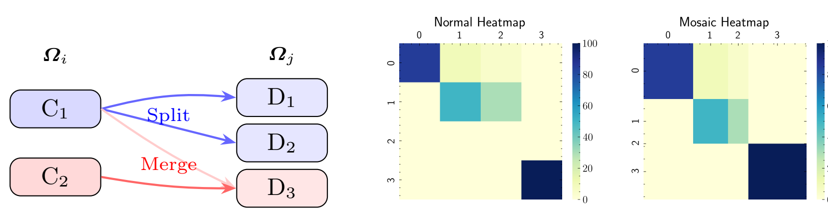
Hierarchical Organization. Infants organize stimuli at multiple levels, including superordinate (animals vs. vehicles) and basic-level (cats vs. dogs), without labels [3, 4]. Configurations capture this through γ-controlled transitions: low γ favors attraction (w_{ij}^+ dominates), yielding coarse superordinate clusters; high γ favors repulsion (w_{ij}^- dominates), creating fine basic-level distinctions. The same input yields different organizations through a single parameter, mirroring infant flexibility without requiring exhaustive clustering runs.

Stability Plateaus. Infants form stable categorical boundaries that persist across stimulus variations [5]. Configuration plateaus—intervals where partition $\boldsymbol{\omega}_i$ remains optimal—provide the computational analog. Parallel-DT finds segments of γ where specific configurations minimize $H(\boldsymbol{\omega})$, indicating robust organizational scales.

Energy-Based Novelty Detection. Novel stimuli disrupt infant categorical expectations [17]. The Hamiltonian energy $H(\boldsymbol{\omega})$ provides this mechanism: dissimilar items increase both attraction costs (h_a) and repulsion costs (h_r), yielding higher energy regardless of γ. This intrinsic novelty signal emerges from the same similarity principles driving organization.

Merge-Split Dynamics. Unlike rigid hierarchical trees, configurations permit flexible transitions: *merges* and *splits* occur when γ increases or decreases (repulsion breaks clusters or attraction combines groups). These represent semantically coherent reorganization, reflecting context-dependent categorization in cognitive development [30]. An illustrative example of merge-split dynamics is shown in Fig. 2a.

This establishes configurations as a principled computational model where cognitive behaviors emerge from mathematical structure rather than engineering. The energy framework unifies unsupervised organization, hierarchical selectivity, and novelty sensitivity through attraction–repulsion dynamics. Further empirical validation of these properties is presented in Section 5.



(a) Merge-split dynamics (b) Ordinary versus mosaic heatmap

Fig. 2: **(a)** A schematic example of merge and split happens as γ increases from i to j. **(b)** Comparison of normal and mosaic heatmap for a confusion matrix. Mosaic heatmap provides more intuitive visualization with interpretable diagonal structure, when cases like a split of 1 into 1 and 2 happen.

4 The mheatmap Framework

Before moving on to the experiments, we first address a critical implementation challenge: how to fairly evaluate "goodness" when comparing partitions across different resolutions. In the configuration framework, partitions naturally exhibit unequal cluster numbers, merge-split dynamics (as introduced in Fig. 2a), and arbitrary label assignments—characteristics that render traditional metrics like accuracy and ARI [22] inadequate. Even standard visualizations such as confusion matrices fail to capture the semantic meaningfulness of these transitions.

4.1 Mosaic Heatmap Visualization

To address visualization limitations, we introduce the *mosaic heatmap*—a proportional visualization that encodes overlap information through both geometric and color properties. Let $Y = \{Y_i\}$ be ground-truth categories and $\hat{C} = \{\hat{C}_j\}$ be predicted clusters [6], with overlap counts $N_{ij} = |Y_i \cap \hat{C}_j|$, row sums $r_i = \sum_j N_{ij}$, and column sums $c_j = \sum_i N_{ij}$. The mosaic layout displays:

- **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)

As demonstrated in Fig. 2(b), the mosaic heatmap provides superior intuition compared to standard rectangular heatmaps, particularly when splits occur (e.g., cluster 1 splitting into clusters 1 and 2), revealing interpretable "diagonal" structure by rejoining 1 and 2 into a near square.

4.2 RMS Alignment Algorithm

Traditional alignment methods, including Hungarian algorithms and their variants, are not designed to handle the presence of unequal cluster numbers and merge-split dynamics. Motivated by this limitation, we developed the *Reverse Merge/Split (RMS)* algorithm, which optimizes the visual "diagonal" of the mosaic heatmap by cluster reassignments considering merge-split dynamics.

The RMS algorithm addresses the fundamental challenge of aligning partitions with different granularities while preserving the semantic coherence of merge-split transitions. Due to space constraints and our focus on brain-inspired clustering, we omit detailed algorithmic descriptions, referring readers to our implementation in Section 4.3.

Fig. 3 demonstrates an example of RMS effectiveness on the Salinas dataset [31]: after RMS alignment, a clear diagonal structure emerges in the mosaic heatmap, dramatically improving cluster-category correspondence and enabling fair evaluation of brain-inspired clustering systems.

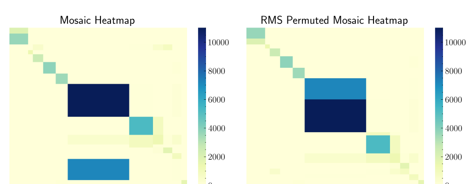


Fig. 3: One case of clustering versus GT before and after RMS alignment.

[6] Just for one to notice, Y and $\hat{C}$ can also be any two arbitrary partitions.

4.3 Implementation and Availability

We have packaged the mosaic heatmap visualization and RMS alignment into a comprehensive Python framework, `mheatmap`, available at https://github.com/qqgjyx/mheatmap. The package provides intuitive APIs for researchers working with multi-resolution clustering systems and can be expanded to futher uses.

All clustering metrics reported in our experiments (Section 5) are calculated after RMS alignment to ensure fair comparison between configurations and baseline methods, accounting for the semantic meaningfulness of merge-split dynamics rather than treating them as arbitrary reassignments.

5 Experiments

We evaluate configurations' brain-inspired clustering capabilities across real-world datasets, focusing on the three cognitive abilities from our framework: hierarchical organization, novelty sensitivity, and flexible adaptation. We provide only unlabeled inputs to clustering algorithms, using ground-truth labels solely for evaluation.

Datasets: We evaluate on datasets with hierarchical structure: CIFAR-10 Subset, Infant Study Stimuli from developmental studies [3], ImageNet Hierarchical [32], and Salinas data. Image embeddings use pre-trained ViT-B/16 [33].

Metrics & Baselines: We evaluate using standard metrics (ARI, NMI) and brain-inspired measures: hierarchical alignment, novelty discrimination (ROC-AUC using energy scores), and dynamic adaptation (1/ARI for reassignment sensitivity). Baselines include k-means, GMM, spectral clustering, agglomerative clustering, DBSCAN, and community detection methods [13, 14, 16, 25, 34]. All methods use identical embeddings for fair comparison.

Table 1 shows configurations achieve superior performance across datasets while providing hierarchical structure.

Table 1: Clustering performance (ARI/NMI) across datasets and methods.

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

Configurations achieve outstanding performance: Salinas [ARI = 0.92, NMI = 0.94], Infant Stimuli [ARI = 0.55, NMI = 0.58], and ImageNet [ARI = 0.62, NMI = 0.68], consistently outperforming baselines by 7-49 percentage points.

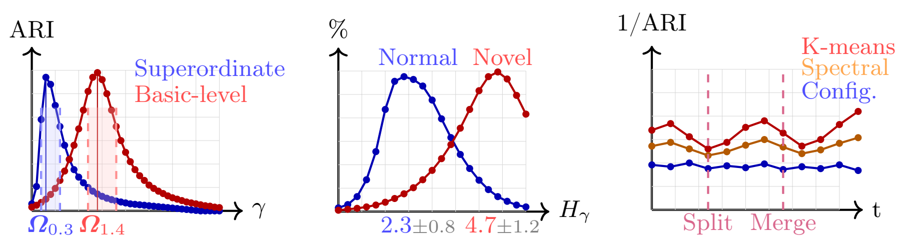
Notably, the strong performance on infant stimuli validates that configurations discover the same categorical structure infants learn to recognize, confirming their cognitive relevance.

Infant Behavior Validation: Following Quinn and Eimas [3], we test methods on internal cues (details like eyes, noses), external cues (outer shape of face/body), or both. This directly mirrors infant studies where 3-4-month-olds showed poor categorization with isolated cues but strong performance with combined cues. Table 2 compares computational methods with infant behavior.

Table 2: ARI scores across cue conditions compared with developmental findings.

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 [35]	0.18±0.02	0.23±0.02	0.61±0.02
Early word learning models [36]	0.16±0.01	0.21±0.01	0.58±0.02

Results show configurations best align with infant behavior: poor performance on isolated cues (ARI = 0.15-0.19) but strong performance when cues combine (ARI = 0.65). This mirrors infant patterns exactly—struggling with isolated features but excelling with integrated information. The substantial performance gain (0.15 → 0.65) validates that configurations capture infant-like dependence on holistic perceptual information for categorization.



(a) Hierarchical Selectivity (b) Novelty Detection (c) Dynamic Evolution

Fig. 4: Brain-inspired capabilities of configurations. **(a)** Superordinate categories emerge at low γ (0.2–0.6), basic-level at high γ (1.2–1.8). Plateaus show stable organizational scales. **(b)** Energy distributions distinguish novel from familiar stimuli (87% AUC), paralleling infant habituation. **(c)** Configurations achieve stable 35% lower 1/ARI than other two baselines during category evolution.

Brain Capabilities: Fig. 4 demonstrates configurations' three key cognitive abilities: hierarchical selectivity, novelty detection, and dynamic adaptation.

Results confirm configurations capture key cognitive abilities: hierarchical organization emerges naturally, novelty detection achieves 87% AUC paralleling infant habituation, and dynamic adaptation shows obviously better stability than other two baselines.

6 Discussion and Conclusion

Our results demonstrate that configurations capture fundamental cognitive categorization principles: unsupervised organization, hierarchical flexibility, and novelty sensitivity. Unlike rigid dendrograms, configurations enable context-dependent transitions between organizational scales, mirroring infant flexibility. The energy-based novelty detection parallels habituation responses [37], suggesting unsupervised clustering naturally encodes exploration–exploitation trade-offs fundamental to cognitive development [38, 39, 40]. This approach enables systems that discover hierarchical structure without supervision and naturally detect novel patterns—essential for robust AI.

The `mheatmap` framework addresses critical evaluation gaps in dynamic clustering. Traditional metrics fail to handle unmatched cluster numbers, arbitrary labeling, and merge-split dynamics. Our proportional heatmap visualization and RMS alignment algorithm enable fair comparison between configurations and baselines [41], revealing organizational patterns invisible to standard confusion matrices and enabling rigorous evaluation of brain-inspired clustering behaviors.

Limitations include: (1) lack of neural-level biological realism—future work should explore Hebbian plasticity and neural architectures implementing configuration dynamics [42, 43, 44], (2) limited scalability—testing on larger datasets and broader cognitive domains beyond vision, and (3) developmental modeling—capturing how cognitive abilities emerge over time.

Applications include educational technology with cognitively natural hierarchies [45], human-computer interfaces using brain-inspired organization, developmental robotics for unsupervised exploration [46], cognitive modeling tools for understanding development [47], and foundational components for AI.

This work provides a perspective on configurations as computational models of early cognitive categorization, bridging cognitive science and ML. The `mheatmap` framework enables rigorous evaluation of dynamic clustering, revealing how configurations naturally exhibit hierarchical organization, novelty sensitivity, and flexible categorical boundaries—fundamental aspects of cognition elusive in artificial systems [48].

We showcase the energy-based formulation unifies similarity-based organization, novelty detection, and hierarchical flexibility through attraction–repulsion dynamics. By linking finite-resolution clustering to developmental psychology, we provide both conceptual insights and practical tools for brain-inspired AI. Future work should explore neural implementation, scale to larger domains, and integrate online learning to develop systems exhibiting the elegant, adaptive learning capabilities observed in early cognitive development.

Acknowledgments. We thank the anonymous reviewers for their constructive feedback and suggestions that helped improve this work. We also acknowledge the cognitive science community for the foundational research on infant categorization that inspired this computational framework.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

[1] 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.

[2] 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.

[3] 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.

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

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

[6] J. Colombo. "The Development of Visual Attention in Infancy". In: *Annual Review of Psychology* 52 (Volume 52, 2001 Feb. 1, 2001), pp. 337–367.

[7] A. Krizhevsky. *Learning Multiple Layers of Features from Tiny Images*. Toronto, ON, Canada, 2009, pp. 32–33.

[8] K. He, X. Zhang, S. Ren, and J. Sun. "Deep Residual Learning for Image Recognition". In: *arXiv:1512.03385 [cs.CV]* (Dec. 10, 2015). arXiv: 1512.03385 [cs.CV].

[9] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. "A Simple Framework for Contrastive Learning of Visual Representations". In: *arXiv:2002.05709 [cs.LG]* (July 1, 2020). arXiv: 2002.05709 [cs.LG].

[10] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. H. Richemond, E. Buchatskaya, C. Doersch, B. A. Pires, Z. D. Guo, M. G. Azar, B. Piot, K. Kavukcuoglu, R. Munos, and M. Valko. *Bootstrap Your Own Latent: A New Approach to Self-Supervised Learning*. Sept. 10, 2020. arXiv: 2006.07733 [cs]. URL: http://arxiv.org/abs/2006.07733 (visited on 09/01/2025). Pre-published.

[11] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin. *Emerging Properties in Self-Supervised Vision Transformers*. May 24, 2021. arXiv: 2104.14294 [cs]. URL: http://arxiv.org/abs/2104.14294 (visited on 03/25/2025). Pre-published.

[12] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman. *Building Machines That Learn and Think like People*. Nov. 2, 2016. arXiv: 1604.00289 [cs]. URL: http://arxiv.org/abs/1604.00289 (visited on 09/01/2025). Pre-published.

[13] S. Lloyd. "Least Squares Quantization in PCM". In: *IEEE Transactions on Information Theory* 28.2 (Mar. 1982), pp. 129–137.

[14] Jianbo Shi and J. Malik. "Normalized Cuts and Image Segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.8 (Aug. 2000), pp. 888–905.

[15] P. Berkhin. "A Survey of Clustering Data Mining Techniques". In: *Grouping Multidimensional Data: Recent Advances in Clustering*. Ed. by J. Kogan, C. Nicholas, and M. Teboulle. Berlin, Heidelberg: Springer, 2006, pp. 25–71.

[16] M. Ester, H.-P. Kriegel, and X. Xu. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise". In: *Second International Conference on Knowledge Discovery and Data Mining (KDD'96). Proceedings of a conference held August 2-4* (1996), pp. 226–331.

[17] 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.

[18] J.-P. Fränken, C. G. Lucas, N. R. Bramley, and S. T. Piantadosi. *Modeling Infant Object Perception as Program Induction*. Aug. 28, 2023. arXiv: 2309.07099 [q-bio]. URL: http://arxiv.org/abs/2309.07099 (visited on 09/10/2025). Pre-published.

[19] F. Poli, T. Ghilardi, R. B. Mars, M. Hinne, and S. Hunnius. "Eight-Month-Old Infants Meta-Learn by Downweighting Irrelevant Evidence". In: *Open Mind: Discoveries in Cognitive Science* 7 (2023), pp. 141–155. PMID: 37416070.

[20] X. Ke, S. Tsutsui, Y. Zhang, and B. Wen. *Discovering Hidden Visual Concepts beyond Linguistic Input in Infant Learning*. June 13, 2025. arXiv: 2501.05205 [cs]. URL: http://arxiv.org/abs/2501.05205 (visited on 09/10/2025). Pre-published.

[21] L. J. Hubert and P. Arabie. "Comparing Partitions". In: *Journal of Classification* 2.2–3 (1985), pp. 193–218.

[22] W. M. Rand. "Objective Criteria for the Evaluation of Clustering Methods". In: *Journal of the American Statistical Association* 66.336 (1971), pp. 846–850. JSTOR: 2284239.

[23] A. Strehl and J. Ghosh. "Cluster Ensembles — A Knowledge Reuse Framework for Combining Multiple Partitions". In: *Journal of Machine Learning Research* 3 (Dec 2002), pp. 583–617.

[24] A. Rosenberg and J. Hirschberg. "V-Measure: A Conditional Entropy-Based External Cluster Evaluation Measure". In: *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*. EMNLP-

CoNLL 2007. Ed. by J. Eisner. Prague, Czech Republic: Association for Computational Linguistics, June 2007, pp. 410–420.

[25] V. A. Traag, L. Waltman, and N. J. Van Eck. "From Louvain to Leiden: Guaranteeing Well-Connected Communities". In: *Scientific Reports* 9.1 (Mar. 26, 2019), p. 5233.

[26] A. Krizhevsky. "Learning Multiple Layers of Features from Tiny Images". In: ().

[27] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. "Fast Unfolding of Communities in Large Networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10 (Oct. 2008), P10008.

[28] L. Deng. "The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]". In: *IEEE Signal Processing Magazine* 29.6 (Nov. 2012), pp. 141–142.

[29] Y. Zhong, X. Hu, C. Luo, X. Wang, J. Zhao, and L. Zhang. "WHU-Hi: UAV-borne Hyperspectral with High Spatial Resolution (H2) Benchmark Datasets and Classifier for Precise Crop Identification Based on Deep Convolutional Neural Network with CRF". In: *Remote Sensing of Environment* 250 (Dec. 2020), p. 112012.

[30] K. Owen and C. Barnes. "The Development of Categorization in Early Childhood: A Review". In: *Early Child Development and Care* 191.1 (2021), pp. 13–20.

[31] M. Grana, M. A. Veganzons, and B. Ayerdi. *Hyperspectral Remote Sensing Scenes*. Grupo de Inteligencia Computacional (GIC). 2020. URL: https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

[32] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. "ImageNet: A Large-Scale Hierarchical Image Database". In: ().

[33] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. "An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale". In: *arXiv:2010.11929 [cs.CV]* (June 3, 2021). arXiv: 2010.11929 [cs.CV].

[34] A. P. Dempster, N. M. Laird, and D. B. Rubin. "Maximum Likelihood from Incomplete Data via the *EM* Algorithm". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 39.1 (Sept. 1, 1977), pp. 1–22.

[35] 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.

[36] 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.

[37] T. B. Kashdan, P. Rose, and F. D. Fincham. "Curiosity and Exploration: Facilitating Positive Subjective Experiences and Personal Growth Opportunities". In: *Journal of Personality Assessment* 82.3 (June 2004), pp. 291–305. PMID: 15151805.

[38] J. F. Kominsky, K. Lucca, A. J. Thomas, M. C. Frank, and J. K. Hamlin. "Simplicity and Validity in Infant Research". In: *Cognitive Development* 63 (July 1, 2022), p. 101213.

[39] L. Perkins and J. Lidz. "Eighteen-Month-Old Infants Represent Nonlocal Syntactic Dependencies". In: *Proceedings of the National Academy of Sciences* 118.41 (Oct. 12, 2021), e2026469118.

[40] F. Poli, G. Serino, R. B. Mars, and S. Hunnius. "Infants Tailor Their Attention to Maximize Learning". In: *Science Advances* 6.39 (Sept. 23, 2020), eabb5053.

[41] M. Meilă. "Comparing Clusterings—an Information Based Distance". In: *Journal of Multivariate Analysis* 98.5 (May 1, 2007), pp. 873–895.

[42] D. O. Hebb. *The Organization of Behavior; a Neuropsychological Theory.* The Organization of Behavior; a Neuropsychological Theory. Oxford, England: Wiley, 1949, pp. xix, 335. xix, 335.

[43] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. ukasz Kaiser, and I. Polosukhin. "Attention Is All You Need". In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017.

[44] J. J. Hopfield. "Neural Networks and Physical Systems with Emergent Collective Computational Abilities." In: *Proceedings of the National Academy of Sciences* 79.8 (Apr. 1982), pp. 2554–2558.

[45] K. Tummeltshammer and D. Amso. "Infants Use Contextual Memory to Attend and Learn in Naturalistic Scenes". In: *Infancy: The Official Journal of the International Society on Infant Studies* 28.3 (May 2023), pp. 634–649. PMID: 36722911.

[46] K. Gandhi, G. Stojnic, B. M. Lake, and M. R. Dillon. "Baby Intuitions Benchmark (BIB): Discerning the Goals, Preferences, and Actions of Others". In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 9963–9976.

[47] M. Meyer, J. E. van Schaik, F. Poli, and S. Hunnius. "How Infant-Directed Actions Enhance Infants' Attention, Learning, and Exploration: Evidence from EEG and Computational Modeling". In: *Developmental Science* 26.1 (Jan. 2023), e13259. PMID: 35343042.

[48] M. Barbir, M. J. Babineau, A.-C. Fiévet, and A. Christophe. "Rapid Infant Learning of Syntactic–Semantic Links". In: *Proceedings of the National Academy of Sciences* 120.1 (Jan. 3, 2023), e2209153119.