Different Brain, Same Prototype? Cognitive Variability within a Recurrent Associative Memory

Thaddé Rolon-Mérette (trolo068@uottawa.ca)
School of Psychology, 136 Jean-Jacques Lussier, Vanier Hall
Ottawa, ON, K1N 6N5, CAN

Damiem Rolon-Mérette (drolo083@uottawa.ca)
School of Psychology, 136 Jean-Jacques Lussier, Vanier Hall
Ottawa, ON, K1N 6N5, CAN

Matias Calderini (mcald052@uottawa.ca)
School of Psychology, 136 Jean-Jacques Lussier, Vanier Hall
Ottawa, ON, K1N 6N5, CAN

Sylvain Chartier (sylvain.chartier@uottawa.ca)
School of Psychology, 136 Jean-Jacques Lussier, Vanier Hall
Ottawa, ON, K1N 6N5, CAN

Abstract

When learning similar stimuli, we tend to group them together. This categorization is a behaviour which all humans share. Yet, the pathways undertaken by the brain differs between individuals. To investigate this phenomenon, a Feature Extracting Bidirectional Associative Memory (FEBAM) was used to generate representations of various grouped stimuli. It was determined that representations created by different FEBAMs were always new. However, the learning behaviour was always the same. The generated representations were always categorized into the right category. Finally, by lowering the size of these representations, prototypes of the categories could be created. Recall tests showed that reconstructed prototypes remained the same across all FEBAMs, even if the representations themselves differed. This shows that although the encoding pathways might differ between individuals, the learned cognitive concepts do not. These findings are promising steps towards better understanding how individuals exhibit common cognitive functionality despite variability in neural activity.

Keywords: Variability; Categorization; Feature-Extraction; Associative Learning; Bidirectional Recurrent Neural Networks, Cognition.

Introduction

There are billions of human beings on this planet and each one of them can understand and share complex concepts such as language, games, music and much more. This common cognitive understanding is mind boggling when individuality is considered. There is no brain that is the same as another, with each containing a unique arrangement of its neural structures and connections (Sporns, Tononi, & Kötter, 2005; Thompson, Schwartz, Lin, Khan & Toga, 1996). When presented with the same learning task, different individuals will exhibit different neural activity (Churchland et al., 2010; Mueller et al., 2013). While perception can change based on certain differences in anatomical structures, surprisingly, this variability does not seem to drastically change an individual's understanding of the world and/or its relationships with others. While the neurological pathways involved are different across individuals, the behaviour remains consistent. In other words, from different neural activities, the same cognitive functionality can be observed. That being said, the mechanisms behind such commonality from variability are yet to be fully understood.

An encouraging avenue to better understand this would be to explore the concept of associative learning and categorization. Associative learning can be seen as linking two or more stimuli together (Rescorla & Wagner., 1972). One of its interesting characteristics is the ability to recall one stimulus when only presented with a partial cue (McCleland, McNaughton & O'reilley, 1995). This process forms the basis of categorization whereas similar patterns are grouped together to form a category (Shields, Rovee & Collier, 1992). However, how these "grouped" patterns are represented in our brain remains a mystery. Are the encoded representations of stimuli different across individuals? If so, do they respect the relationship between stimuli, i.e. correctly categorized? In other words, if stimuli are similar, are their representations also similar?

In cognition, such questions can be explored using formal models (Forstmann, Wagenmakers, Eichele, Brown & Serences, 2011). Specifically, artificial neural networks (ANNs) have been an exciting approach to study various key cognitive concepts such as associative learning and categorization (Mareschal, French, & Quinn, 2000). One of the many interesting properties of ANNs dwells in the initialization of weight connections. By randomly initializing the connection weights, each individual instance of a network will be different, analogous to the variability found in human brains. However, what would be interesting is that different instances of a network would display the same behaviour when presented with the same learning task.

Among ANNs are Recurrent Associative Memories, which are designed to implement associative learning (Acevedo-

Mosqueda, Yáñez-Márquez, & Acevedo-Mosqueda, 2010). Particularly, there is the Feature Extracting Bidirectional Associative Memory, or FEBAM (Chartier, Giguère, Renaud, Lina & Proulx, 2007), which can create perceptual features from input patterns via feature extraction (Rolon-Merette, Rolon-Merette & Chartier, 2018). This property allows the FEBAM of category development (grouping similar patterns together based on their correlation). However, a question remains. When presented with the same stimuli, will the FEBAM always generate new representations and if so, will it exhibit the same learning behaviour? In other words, will the representations be categorized in the same manner even if they are always new? This would shed light on the mechanisms allowing common cognitive functionality found between individuals.

The next section gives a short description of the FEBAM and a cluster analysis, followed by three simulations. In simulation I, it was investigated if the representations created by different instances of the FEBAM are always new. In simulation II, the exemplar categorization was observed with a learning task consisting of grouped patterns. In simulation III, under the same learning task, the size of representations was varied to examine prototype categorization. Finally, this paper ends with a short discussion.

Model

The FEBAM is a completely unsupervised recurrent ANN, meaning it does not have any explicit teacher. The entirety of the model can be described by its architecture, transmission function and learning function.

Architecture

The FEBAM architecture is illustrated in Figure 1. The model has two layers of interconnected units in a bidirectional fashion, where the **W** and **V** layers return information to each other. Contrary to traditional bidirectional associative memories, there is only one explicit connection, **x**(0), to allow the network to perform feature extraction.

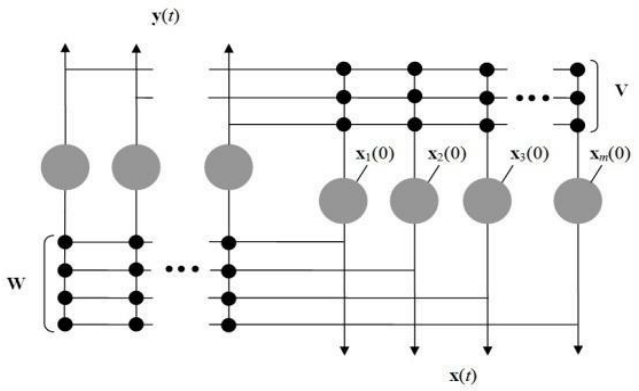

Figure 1: Architecture of the FEBAM

Output function

The transmission function is defined by the Equations 1a and 1b:

$$(1a)\ \forall i,\dots,N, y_{i(t+1)} = f\left(a_{i(t)}\right) = \begin{cases} 1, & if\ a_{i(t)} > 1 \\ -1, & if\ a_{i(t)} < -1 \\ (\delta+1)a_{i(t)} - \delta a_{i(t)}^3, & \text{Else} \end{cases}$$

$$(1b)\ \forall i,\dots,M, x_{i(t+1)} = f\left(b_{i(t)}\right) = \begin{cases} 1, & if\ b_{i(t)} > 1 \\ -1, & if\ b_{i(t)} < -1 \\ (\delta+1)b_{i(t)} - \delta b_{i(t)}^3, & \text{Else} \end{cases}$$

Where N and M are the total number of units in each layer, i is the index unit, δ is the general transmission parameter and a and b are the activations. These activations are obtained the following way: $\mathbf{a}(t)=\mathbf{W}\mathbf{x}(t)$ and $\mathbf{b}(t)=\mathbf{V}\mathbf{y}(t)$.

Learning rule

The connection weights are modified following a hebbian/anti-hebbian rule:

$$(2a)\ \mathbf{W}(k+1) = \mathbf{W}(k) + \eta\left(\mathbf{y}(0) - \mathbf{y}(t)\right)\left(\mathbf{x}(0) + \mathbf{x}(t)\right)^{\mathrm{T}}$$
$$(2b)\ \mathbf{V}(k+1) = \mathbf{V}(k) + \eta\left(\mathbf{x}(0) - \mathbf{x}(t)\right)\left(\mathbf{y}(0) + \mathbf{y}(t)\right)^{\mathrm{T}}$$

Where $\mathbf{x}(0)$ and $\mathbf{y}(0)$ are the initial inputs, η is the learning parameter and k is a given learning trial. Equation 2a and 2b shows that the matrix weights will converge when $\mathbf{x}(0) = \mathbf{x}(t)$ or $\mathbf{y}(0) = \mathbf{y}(t)$. To reduce the simulation time the number of iterations was set to $t = 1$. It is guaranteed that the learning will converge if the learning parameter (η) is smaller than the following value (Chartier & Boukadoum, 2006):

$$(3)\ \eta < \frac{1}{2(1-2\delta)\mathrm{Max}[M,N]}, \delta \neq \frac{1}{2}$$

FEBAM learning process

As previously mentioned, in the FEBAM, there is only one explicit connection $\mathbf{x}(0)$, meaning the $\mathbf{y}(0)$ inputs are not initially available. Instead, they are obtained after a first iteration through the network. As shown in Figure 2, $\mathbf{y}(0)$ is obtained by the iteration of $\mathbf{x}(0)$ through its corresponding weight connections $\mathbf{W}$ using the transmission function. Subsequently, $\mathbf{x}(1)$ is obtained from $\mathbf{y}(0)$ and finally, $\mathbf{y}(1)$ from $\mathbf{x}(1)$. Through the weight updates, each $\mathbf{x}(1)$ and $\mathbf{y}(1)$ will converge to a solution that will try to best reconstruct its associated initial pattern $\mathbf{x}(0)$ or its initial output $\mathbf{y}(0)$. Thus, in the case where $\mathbf{x}(1)$ does not equal $\mathbf{x}(0)$, weight convergence will be granted by $\mathbf{y}(1)$.

The number of units in the **y**–layer determined the dimensionality (size) of the generated representation.

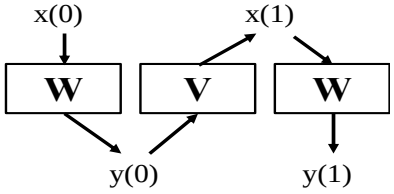

Figure 2: Iterative process for weight updates during learning in the FEBAM.

FEBAM Learning Procedure

The transmission function's parameter (δ) was set to 0.2 and the learning parameter (η) respected Equation 3 for all the simulations. Weights were randomly initialized with values between -0.1 and 0.1. Learning stopped when the network achieved a mean squared error (MSE) of less than 10^{-10} or when 5000 learning trials was reached. Learning was conducted following this procedure:

1. Creation of a list of inputs respecting preset conditions.
2. Random selection of a given exemplar from the list to obtain **x**(0).
3. Iteration through the network (as illustrated in Figure 2) using the output function to obtain **y**(0), **x**(1) and **y**(1).
4. Computation of weight updates according to the learning rule.
5. Repetition of steps 2) to 4) until the minimum mean squared error between **y**(0) and **y**(1) or maximum trials is reached.

Cluster Analysis

In order to partition the generated representations into categories, k-mean clustering was used. For a chosen number of clusters *k*, the algorithm randomly sets *k* centroids in feature space and assigns each data point to the category of its nearest centroid. The positions of each centroid are then iteratively readjusted such that the within-category distance of the resulting categories is minimized. Lloyds algorithm and K-means++ initialization were implemented with the SickitLearn library on Python (Arthur & Vassilvitskii, 2007; Kanungo, Mount, Netanyahu, Piatko, Silverman & Wu, 2002). The sum of the squared distances between data points and their centroid is presented by distortion. *A priori*, the number of clusters that would most appropriately divide the data cannot be known and its high dimensionality makes it prohibitive to determine it visually. Instead, the elbow method was applied to select the optimal number of clusters (Kodinariya & Makwana, 2013). Cluster analysis was conducted under two different scenarios. Scenario A will be used to examine variability across all FEBAMs (Simulation I and IIIb). Scenario B allows to find the average behaviour of an individual FEBAM (Simulation II and IIIa).

Scenario A

1. Creation of input patterns respecting preset conditions.
2. FEBAM learning as specified in the learning procedure.
3. Repetition of steps 1) and 2) for all FEBAMs.
4. K-Means cluster analysis on generated representations of all FEBAMs at once from step 3).

Scenario B

1. Creation of input patterns respecting preset conditions.
2. FEBAM learning as specified in the learning procedure.
3. K-Means cluster analysis on generated representations of each individual FEBAM.
4. Repetition of steps 1) to 3) for all FEBAMs.
5. Calculate average distortion and number of clusters from step 4).

Simulation I: new representations

The number of different generated features was studied when the inputs were kept constant. The task consisted of three learning conditions of different input patterns and generating their associated representations. In each condition, the patterns were fed to multiple FEBAMs, mimicking the learning process of different individuals. The generated outputs, or representations, were then analyzed with k-means clustering using Scenario A.

Methodology

Three different learning conditions were studied using pixelated bipolar inputs patterns of dimension 50, where black pixels represent the value of +1 and white pixels -1. The "pattern" condition consisted of a single pattern. The "category" condition consisted of two categories of five highly correlated patterns. Each pattern within a category exhibited a correlation of 0.95 and the correlation between patterns of both categories was set to 0.15. Finally, in the "random" condition, ten inputs were generated with low correlations varying from 0.01 to 0.30. All three conditions are presented in Figure 3.

In order to have a good estimate of the behaviour, the input patterns were presented to 1000 different FEBAMs, each with a different set of randomly initialized weight connections. The size of the generated representations was kept constant at a dimension of 50. Finally, for each condition, k-means clustering analysis was conducted on the generated representations of all the FEBAMs at once, as stated in Scenario A.

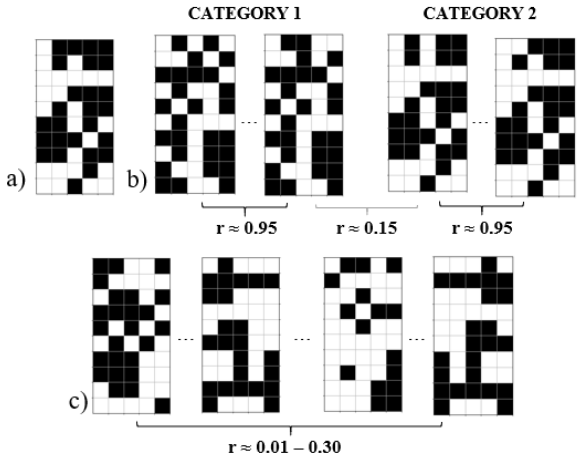

Figure 3: Input patterns for the "pattern" (a), "category" (b) and the "random" (c) conditions.

Results

Different FEBAMs generated different representations when presented with the same pattern(s). Figure 4 illustrates an example of this process. Figure 5 shows the results of k-means clustering for each condition. As the number of clusters created increased, the distortion decreased. However, for all three conditions no elbow was observed.

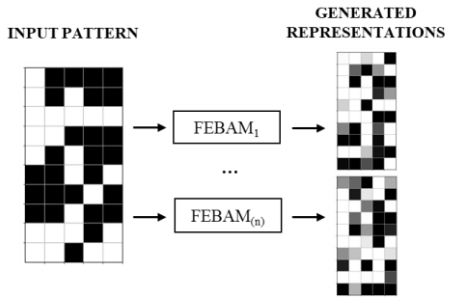

Figure 4: Generating representations for the "pattern" condition with different FEBAMs.

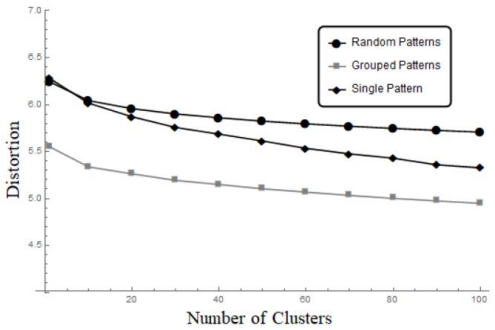

Figure 5: Cluster analysis on representations formed across 1000 different FEBAMs.

Simulation II: Exemplar categorization

In this section, we further investigate whether different FEBAMs respect the same behaviour during exemplar categorization. To do this, we extended the condition 2 of simulation I to five categories. However, in this case, clustering analysis will be conducted on individual FEBAMs and not all at once, as stated is Scenario B.

Methodology

The same method described in simulation I was used to generate input patterns. Here, two to five categories were generated. Each category contained five patterns. The correlation of patterns within the categories was approximately 0.95 and the correlation of patterns between the categories was set to approximately 0.15. The dimensionality of representations (outputs **y** units) was set again to 50. Each set of patterns were presented to 1000 different FEBAMs with the same learning procedure and parameters as previously described. Subsequently, following Scenario B, k-means cluster analysis was conducted on the generated patterns of individual FEBAMs only. Within-category and between-category correlation of generated representations were also examined. Finally, a recall test was performed to verify that patterns were correctly categorized.

Results

In Figure 6, an example of exemplar categorization is presented. In Figure 7, the mean number of clusters and distortion for the 1000 FEBAMs are presented. Results show that the generated representations respected the number of categories found in the input patterns (e.g. two categories, two clusters of generated representations). Furthermore, the average within-category correlation of generated representations was 0.75 and the average between-category correlation was <0.05. Lastly, the recall test yielded a performance of 100% correct pattern categorization.

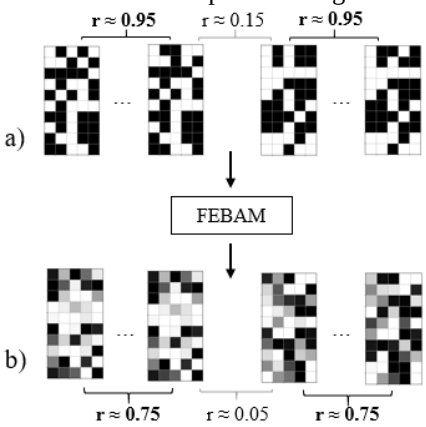

Figure 6: Example of exemplar categorization. Within category (black) and between category (gray) correlation for input (a) and output (b) patterns are presented.

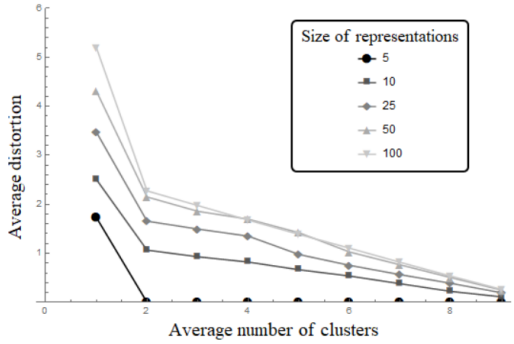

Figure 7: Average distortion and number of clusters for generated representations in function to the number of categories.

Simulation III: Prototype categorization

In this last simulation, the goal was to examine the behaviour of the FEBAM during prototype categorization. A previous study showed that if the dimensionality of the representations is small enough when compared to the number of patterns, prototypes are formed (Giguère, Chartier, Proulx & Lina, 2007). However, the variability of recalled prototypes formed across different FEBAMs was not investigated.

Methodology

Two categories of input patterns, each containing five patterns, were generated in the same fashion as in Simulation Ib and II. The dimensionality of generated representations (number of **y** units) was varied from 5, 10, 25, 50 to 100 dimensions. The patterns were presented to 1000 different FEBAMs using the same learning procedure and parameters as in simulation I and II. Two clustering analyses were conducted.

Simulation IIIa. First, to determine the relationship between distortion and size of representations, k-means cluster analysis was conducted on generated representations from individual FEBAMs. This was done following the procedure described in Scenario B.

Simulation IIIb. Second, to determine the variability of recalled patterns, k-means clustering analysis was conducted on generated representation of dimension 5 and their recalled patterns for all FEBAMs at once. This was done following the procedure described in Scenario A.

Results

Figure 8 shows the first cluster analysis. The average number of clusters and distortion is presented. In all cases, two clusters were formed. Additionally, by lowering the dimensionality of the representations, clusters with lower distortion began to appear. With representations of dimension 5, two clusters accounted for all the distortion, suggesting that prototypes were created.

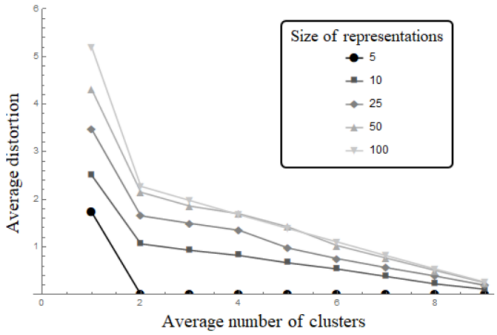

Figure 8: Simulation IIIa. Relationship between the number of clusters and number of y units.

In Figure 9, the second k-means clustering is shown. When looking at the generated representations across the 1000 FEBAMs, it is quickly noted that no clusters were observed. This is consistent with the results from Simulation I, different FEBAMs will always generate different representations. Furthermore, when looking at the recalled patterns, two clusters are shown. However, these accounted for almost all the distortion. This suggests that although coming from 1000 different FEBAMs, the same two patterns were recalled. Figure 10 illustrates this process.

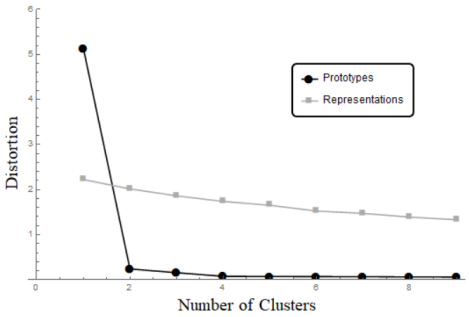

Figure 9: Simulation IIIb. Recalled prototypes and generated representations clustering across 1000 FEBAMs.

Discussion

The goal of this paper was to determine if the FEBAM could shed light on the categorization process found within and between individuals. Results from simulation I showed that when learning the same stimuli, different FEBAMs will generate diverse representations of these input patterns. As seen by the absence of clusters during a k-means clustering analysis. This result was expected since connection weights were initialized randomly.

However, in simulation II, it was found that although different FEBAMs generate different representations, their learning behaviour remained the same. This was first shown by looking at the correlation of generated representations from each FEBAM. The within-category correlation (≈ 0.75) was far greater than the between-category correlation (≈ 0.05). This was further shown with a k-means clustering analysis on the representations. The analysis put forward the fact that the number of clusters corresponded to the number of categories. In addition, recalled patterns were correctly categorized into individual exemplars. These findings are keys since it proposed that the FEBAM will have the same encoding behaviour even if the initial connection weights are different. This also contributes to previous work by showing that both representations and reconstructed patterns are categorized in the same manner (Giguère, Chartier, Proulx & Lina, 2007).

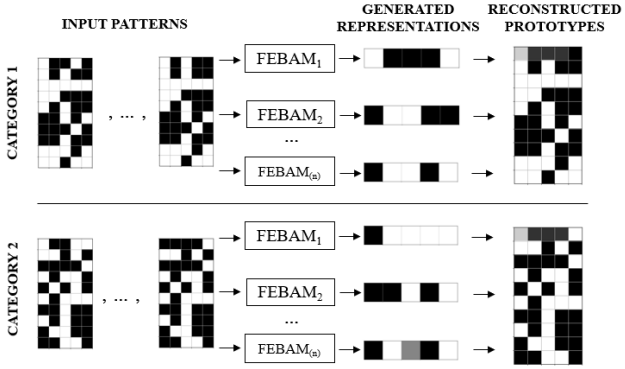

Figure 10: Example of Prototype categorization. Two categories of input patterns are presented to (n) different FEBAMs. These generate representations of dimension 5. Although representations are always different, the same two patterns are recalled. These recalled patterns act as a prototype for each input category.

This characteristic was further explored with prototype categorization. Simulation III showed that different FEBAMs constructed the same prototypes even if the stored representations were different. If the size of the representations is equal or lower than the number of patterns of a given category, then the same pattern was always recalled. This recalled pattern was a prototype of all input patterns within a given category. Thus, even if the initial learning conditions and subsequent generated representations are different, the network will still create the same prototypes.

To sum up, this study showed that the FEBAM is a good model for categorization, capable of both exemplars and prototypes encoding while also accounting for individual differences. The findings are a promising step towards better understanding how individuals exhibit common cognitive functionality despite variability in neural activity and may help in defining the optimal conditions to perform a classification task.

Future work could focus on how manipulating weight initialization may influence learning. A change in initial weight connections between different FEBAMs could result in a corresponding change in their generated representations. Furthermore, depending on the size of the network, the FEBAM exhibits different behaviours during reconstruction of the input patterns (prototype or exemplar recall). An interesting property would be to grow (increase y-units) or prune (decrease y-units) the network based on a task. This would help to surpass the current task specific problem and allow the model to be more generalized.

Achnowledgements

This research was partly supported by the Natural Sciences and Engineering Research Council of Canada.

References

Acevedo-Mosqueda, M. E., Yáñez-Márquez, C., & Acevedo-Mosqueda, M. A. (2013). Bidirectional associative memories: Different approaches. *ACM Computing Surveys (CSUR)*, *45*(2), 18.

Arthur, D., & Vassilvitskii, S. (2007, January). k-means++: The advantages of careful seeding. In Proceedings of the eighteenth annual *ACM-SIAM symposium on Discrete algorithms* (pp. 1027-1035). Society for Industrial and Applied Mathematics.

Chartier, S., & Boukadoum, M. (2006). A bidirectional heteroassociative memory for binary and grey-level patterns. *IEEE Transactions on Neural Networks*, *17*(2), 385-396.

Chartier, S., Giguère, G., Renaud, P., Lina, J. M., & Proulx, R. (2007, August). FEBAM: A feature-extracting bidirectional associative memory. *In Neural Networks, 2007. IJCNN 2007. International Joint Conference on* (pp. 1679-1684). IEEE.

Chervyakov, rolA. V., Sinitsyn, D. O., & Piradov, M. A. (2016). Variability of Neuronal Responses: Types and Functional Significance in Neuroplasticity and Neural Darwinism. *Frontiers in human neuroscience*, *10*, 603.

Churchland, M. M., Byron, M. Y., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., ... & Bradley, D. C. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature neuroscience*, *13*(3), 369.

Forstmann, B. U., Wagenmakers, E. J., Eichele, T., Brown, S., & Serences, J. T. (2011). Reciprocal relations between cognitive neuroscience and formal cognitive models: opposites attract*? Trends in cognitive sciences*, *15*(6), 272-279.

Giguère, G., Chartier, S., Proulx, R., & Lina, J. M. (2007). Category development and reorganization using a bidirectional associative memory-inspired architecture. In *Proceedings of the 8th international conference on cognitive modeling* (pp. 97-102). Ann Arbor, MI: University of Michigan.

Kanai, R., Bahrami, B., & Rees, G. (2010). Human parietal cortex structure predicts individual differences in perceptual rivalry. *Current biology*, *20*(18), 1626-1630.

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. IEEE *Transactions on Pattern Analysis & Machine Intelligence*, (7), 881-892.

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal*, *1*(6), 90-95.

Mareschal, D., French, R. M., & Quinn, P. C. (2000). A connectionist account of asymmetric category learning in early infancy. Developmental psychology, 36(5), 635.

McClelland, J. L., McNaughton, B. L., & O'reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, *102*(3), 419.

Mueller, S., Wang, D., Fox, M. D., Yeo, B. T., Sepulcre, J., Sabuncu, M. R., ... & Liu, H. (2013). Individual variability in functional connectivity architecture of the human brain. *Neuron*, *77*(3), 586-595.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, *2*, 64-99.

Rolon-Merette, T., Rolon-Merette, D., & Chartier, S. (2018). Generating Cognitive Representations with Feature-Extracting Bidirectional Associative Memory. *Procedia computer science*, *145*, 428-436.

Shields, P. J., & Rovee-Collier, C. (1992). Long-term memory for context-specific category information at six months. *Child Development*, *63*(2), 245-259.

Sporns, O., Tononi, G., & Kötter, R. (2005). The human connectome: a structural description of the human brain. *PLoS computational biology*, *1*(4), e42.

Thompson, P. M., Schwartz, C., Lin, R. T., Khan, A. A., & Toga, A. W. (1996). Three-dimensional statistical analysis of sulcal variability in the human brain. *Journal of Neuroscience*, *16*(13), 4261-4274.