# Primacy/recency effects in infant categorisation

**Valentina Gliozzi (gliozzi@di.unito.it)**[1]
Dipartimento di Informatica, Università di Torino
Corso Svizzera 185, 10149 Torino, Italy

**Nadja Althaus (nadja.althaus@psy.ox.ac.uk )**[1]
Department of Experimental Psychology, University of Oxford
South Parks Road, Oxford OX1 3UD, UK

**Julien Mayor (julien.mayor@unige.ch)**
FPSE, University of Geneva
40, Boulevard du Pont-d'Arve 1211 Genève 4, Switzerland

**Kim Plunkett (kim.plunkett@psy.ox.ac.uk )**
Department of Experimental Psychology, University of Oxford
South Parks Road, Oxford OX1 3UD, UK

**Abstract**

We provide evidence that primacy and / or recency effects play a crucial role in infant visual categorization. First, we demonstrate that a connectionist model of infant categorization based on a self-organizing map (Gliozzi, Mayor, Hu, & Plunkett, 2009) predicts an increased influence of the first and the last stimuli during familiarization on the category boundaries. We then present data from 10-month-old infants which confirm these effects. Future research will allow to discriminate between a primacy or a recency effect.

**Keywords:**infant categorization, self-organizing maps, connectionist modelling

## Introduction

familiarization/novelty preference paradigms have been widely used in experiments on infant categorization. In these types of experiment, infants are first familiarized with a sequence of stimuli. After the familiarization phase, infants are tested by simultaneously showing them two test stimuli: a within-category test stimulus and an out-of-category test stimulus. After the test phase is completed, category formation is assessed by comparing looking time at the within-category test stimulus and looking time at the out-of-category test stimulus. Novelty preference – longer looking time at the out-of-category test stimulus than at the within-category stimulus – is taken as an indication for categorization: if looking time is indexed as a measure of surprise, this indicates that the out-of-category test stimulus is less familiar than the within-category one, and therefore that infants have formed one category over the familiarization stimuli.

The assumption underlying the novelty preference test is that infants form a category representation close to the central tendency of the stimuli. In other words, this representation is equidistant from all the stimuli and represents them equally well, in a process that is unaffected by the order of the stimuli presentation. In this paper, we question this assumption, and argue that the process of category formation is more disordered than this, and depends on many familiarization contingencies. In particular, we argue that a primacy or recency effect will affect the category formation process: the number and type of categories formed is modulated by the identity of the first, or last, stimuli presented. Future research will aim at distinguishing the relative roles of primacy and recency effects.

We will first show how the hypothesis of a primacy/recency effect was derived from the analysis of the behavior of a computational model, closely related to the model presented by Gliozzi et al. (2009). The model's predictions have been subsequently tested and validated by testing 10-month-old infants in Oxford. This manuscript results from of a strong interplay between computational simulations and experimental results.

## Literature and Previous Results

Although it is clear that infants can form categories from visual familiarization stimuli (Younger, 1985; Eimas & Quinn, 1994; Mareschal & Quinn, 2001), the way in which familiarization contingencies impact category formation remained elusive until recently (Kovack-Lesh & Oakes, 2007; P.C.Quinn & R.S.Bhatt, 2010; Bomba & Siqueland, 1983; Mather & Plunkett, 2011) and the nature of the categories formed is yet to be understood.

In a previous experiment, Mather and Plunkett (2011) showed that the *order* of presentation of the familiarization stimuli can affect categorization. In particular, Mather and Plunkett (2011) compared infant categorization under two familiarization conditions that differ in the order by which the same set of stimuli (those used by Younger (1985)) is presented to infants during familiarization. Examples of familiarization stimuli, as well as of within-category (average), and out-of-category (peripheral) test stimuli can be found in Figure 1. In the high distance condition, infants were familiarized with sequences that maximize the Euclidean distance in feature space between successive stimuli whereas in the low

[1] Both authors contributed equally to this work.

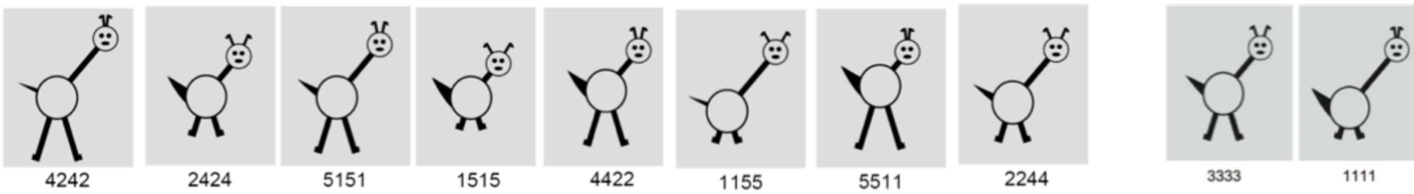


Figure 1: Example of familiarization sequences in the high distance condition with mild start/end stimuli and of the test stimuli

distance condition the Euclidean distance between successive stimuli is minimized. Mather and Plunkett (2011) found that only infants in the high distance condition successfully exhibited novelty preference at test, indicating that they had formed a category over the familiarization stimuli. Despite seeing the same items, with the only difference being the order of successive stimuli, infants in the low distance condition failed to discriminate between the test stimuli. The authors gave some potential explanations for this finding, ranging from faster habituation in the low distance condition, to the fact that infants in the high distance condition explore a bigger feature space than infants in the low distance condition, hence achieving more robust representations, until the fact that it may be more difficult to discriminate between small changes in feature space in successive stimuli in the low condition than when incremental changes in feature space are larger, as it is the case in the high distance condition. In this paper we provide a further explanation, while trying to gain further insight into the mechanisms underlying category formation with different familiarization contingencies.

Mather and Plunkett (2011)'s results are the starting point of this work. We first reproduce Mather and Plunkett (2011)'s results with a slightly-modified version of the model introduced by Gliozzi et al. (2009). As we will see, the updated model not only captures Mather and Plunkett (2011)'s results but also suggests an interpretation of its behavior which is different from the set of potential explanations provided by Mather and Plunkett (2011). Similarly to Mather and Plunkett (2011), we argue that categorization is affected by the order of presentation of the stimuli. However, in contrast from Mather and Plunkett (2011), we suggest that the largest effect impacting categorization is the identity of the first or last stimulus of the sequence, rather the average Euclidean distance in feature space between successive stimuli. In other words, we argue for a primacy/recency effect. As we will see, the experiments with infants confirm this hypothesis.

## Computational Model

### The model

The model we consider here is an adaptation of the model presented by Gliozzi et al. (2009). The model is a self-organizing map (Kohonen, 1997), which is recognized as a psychologically plausible neural network model (Kohonen, 1993), implementing a biologically plausible approach to human information processing: although our implementation is at a high level of abstraction, we can be confident that the map architecture and learning algorithms used in the paper can be implemented at a physiological level of information processing. Psychological plausibility is added to our model by the fact that the map can be trained by following the same schedule of infants: by presenting each familiarization stimulus only once (instead of hundreds of times as in standard networks). The model receives visual inputs which are vectors with four dimensions (e.g. $[1,1,5,5]$) that represent the stimuli by Younger (1985) used by Mather and Plunkett (2011) (see Figure 1). Each value in the vectors corresponds to one feature in the cartoons presented to infants: length of the neck, length of the legs, the ears' orientation and the size of the tail. The encoding of the stimuli is the same used by Gliozzi et al. (2009), following Mareschal and French (2000). The stimuli can be either "mild", containing feature values close to the overall average (items with feature values 2 and 4 in Figure 1, with mild length legs and neck, etc), or "extreme", containing features further away from the overall average (combinations of values 1 and 5 in Figure 1, with very long or very short legs, very long or very short neck, etc)).

The model, like any self-organizing map, consists of a set of units, spatially organized in regular grids. Each map unit $u$ is associated with a weight vector $W_u$ of the same dimension as the input vectors. All weight vectors taken together can be seen as the map's representation of the world. The weight vectors are initialized to small random values. During training, the input vectors are presented to the network. After each presentation of a vector, its best matching unit is identified. This is the unit whose weight vector is closest to the input vector itself (in Euclidean distance). Next, the best matching unit's weights are adjusted to decrease the difference between the associated weight vector and the current input vector, according to the equation

$$W_u(t+1) = W_u(t) + a(t)(I(t) - W_u(t))$$

where $W_u(t+1)$ and $W_u(t)$ are the weight vectors associated to unit $u$ at time $t+1$ and $t$ respectively. $I(t)$ is the input vector presented to the network at time $t$. For the best matching unit $u$ and for input $I(t)$, the difference $I(t) - W_u(t)$ is called the *quantization error* (*qerr*) of the network with respect to $I(t)$. Adjusting the weights can be seen as corresponding to an adaptation of the map's internal representation to accommodate for the new incoming familiarization stimulus. With respect to standard self-organizing maps our model is simplified and does not have any neighborhood function, due to its limited size. Results extend to a larger version of the model in which there is a (non-shrinking) neighborhood function.

Finally, $a(t)$ is the learning rate at time $t$, defined as $max(0, min(1, \beta * exp^{\alpha * \sqrt{qerr}}))$ (with $\alpha = 4.5$, $\beta = 0.05$). Results are robust (hold in more than 50% of the cases) when $\alpha$ ranges from 1 to 10, and $\beta$ ranges from 0.04 to 0.4. We have also studied a decreasing variant of the learning rate $a(t)' = max(0, min(1, \beta * exp^{\alpha * \sqrt{qerr}}))/t$ that allows to replicate results under some parameters' combinations. In the following we restrict our attention to the non-decreasing learning rate. The model's adaptive learning rate has two important properties. The first of these is that it is usually higher than in standard self-organizing maps. This allows the network to be trained in analogy to an infant familiarization procedure: rather than training the network over hundreds of epochs, effectively presenting each stimulus many times, here each stimulus is only presented to the map once. The second property of the learning rate is that it depends on the quantization error: roughly speaking, the more novel the incoming stimulus is, the higher the learning rate will be. The consequence of this is that the learning rate can be considered as a computational counterpart of attention in infants: the adaptive learning rate corresponds to the general finding that infants pay more attention to novel stimuli rather than to familiar ones.

## The model's predictions

In order to replicate Mather and Plunkett (2011)'s results, we have trained our model in the same way in which infants have been familiarized in Mather & Plunkett's (2011) study: we produced 24 maps per condition (low distance condition versus high distance condition), and each of these was trained with the encoding corresponding to the sequence presented to an infant, with the same schedule used in infant familiarization, i.e. presenting each stimulus exactly once.

After each network was trained, we assessed whether a category had been formed by measuring the *network looking time*, defined as the quantization error (as in Mareschal and French (2000) and Westermann and Mareschal (2004)). In analogy to the infant experiments, network categorization was assessed during a test phase in which network looking time at the overall average test stimulus was compared to looking time at the peripheral test stimulus: a proportion lower than chance indicates that the stimuli presented during familiarization have been organized in a cluster whose centroid is closer to the overall average test stimulus than to the peripheral one.

For each condition, the average of the ratios for all networks was calculated, and compared to the corresponding ratio calculated by Mather and Plunkett (2011). The model reproduces Mather and Plunkett (2011)'s results with infants: networks familiarized in the high distance condition exhibit a stronger novelty preference for the peripheral test stimulus than those familiarized in the low distance condition.

Although the model successfully reproduces Mather and Plunkett (2011)'s results, the organization of its internal representation during training suggests an explanation of the results which is different from that provided by Mather and Plunkett (2011). Indeed, the model predicts that the nature of the start and end stimuli impacts categorization more than the Euclidean distance, as suggested by Mather and Plunkett (2011). In particular, novelty preference on test is stronger for maps familiarized with sequences starting and ending with mild values than for those familiarized with sequences starting and ending with extreme values.

In order to understand how Euclidean distance, on the one side, and the nature of start-end stimuli, on the other side, influence the model's behavior, we have conducted simulations in a 2*2 design considering four different conditions. The conditions are obtained by varying the average Euclidean distance between successive stimuli as well as the nature of the start and end stimuli (whether mild or extreme). We thus consider the four possible combinations: low distance & mild start/end stimuli (Low/Mild); low distance & extreme start/end stimuli (Low/Extreme); high distance & mild start/end stimuli (High/Mild); high distance & extreme start/end stimuli (High/Extreme). In all conditions start and end stimuli are either both mild or both extreme

The model predicts a main effect of start/end stimuli on categorization. For some choices of the learning rate's parameters ($\alpha$ and $\beta$) one obtains an interaction between start/end stimuli and Euclidean distance.

In the following we give an intuitive idea of the model's mechanisms that lead to the prediction. Roughly speaking, the prediction derives from the way in which successive stimuli are organized throughout the training phase: an internal representation (or several internal representations) corresponding to the stimuli experienced is formed and updated run-time, after each stimulus presentation (in line with several other models, as Gliozzi et al. (2009); Gureckis and Love (2004); Westermann and Mareschal (2004)). Depending on the strength of the update of this internal representation after each stimulus presentation (i.e. depending on the value of the learning rate), at the end of the familiarization phase the internal representation is close to the first or last stimulus experienced during familiarization. For our sequences, where start and end type were bound, sequences starting and ending with mild stimuli lead to internal representations of the familiarization stimuli containing mild attributes' values, whereas sequences starting and ending with extreme stimuli lead to internal representations containing extreme attributes' values. For this reason, maps familiarized in the mild condition will find the average test stimulus (that also contains mild values) much more familiar than the peripheral test stimulus, whereas for maps familiarized in the extreme condition the difference will be much less dramatic.

Do infants tested with the same 2*2 design exhibit the same behavior? Can we say that they process the familiarization stimuli in a way similar to the model?

We will see in the next section that infant data reflect the model's predictions. The question naturally arises on how precisely the model's behavior and infant behavior parallel each other. We address this question by considering looking time throughout familiarization/training. As we will see in the next section, infant looking time decreases throughout the familiarization phase in the low-distance condition while remaining stable in the high distance condition. However, the original model does not exhibit this kind of behavior. In order to achieve this behavior in the model we have to add two elements to the learning mechanism : (i) a form of weight decay: the weights associated to the maps' units that are not involved in training (because they are not selected as the best matching unit) slowly decay towards the initial values, and (ii) a form of habituation: the learning rate decreases if the same unit is the best match over multiple trials. With these two new elements, the network looking time mimics infant looking time also in the familiarization phase.

## Experiments

### Methods

**Participants** In total, 104 infants (mean age: 310 days; 52 females) took part in this study. An additional 31 infants were excluded due to technical reasons (N=12) or a failure to reach the looking time criterion (N=19; criterion: a minimum of 6 trials with looking time data including trials 1 and 8, total looking time greater than two standard deviations below the mean). Infants were recruited at the maternity ward of the local hospital.

**Procedure** Infants were seated on the caregiver's lap in front of a large television screen (110 cm x 95 cm) at a distance of approximately 90 cm. They were presented with eight familiarization trials, followed by four test trials (see Figure 1); all trials were 10 seconds in duration. During the eight familiarization trials, a single familiarization image (subtending ca. 14 degrees visual angle) was displayed either on the left or right hand side of the screen. During the test trials, two images were shown side by side. The first two test trials paired one of the peripheral stimuli with the overall average, with a location switch between the trials, and counterbalancing the position of the average stimulus on Test trial 1 across subjects. Test trials 3 and 4 involved one pairing of the novel stimulus with the average stimulus, and one pairing of the novel stimulus with the peripheral stimulus shown during tests 1 and 2 (order of trials and location of stimuli were counterbalanced). The infant's face was filmed by two cameras mounted above the screen to the left and right. Throughout the procedure, the experimenter monitored infants' gaze from a control room next to the testing booth. Trials were initiated manually by the experimenter after confirming that the infants gaze was directed at the screen, or re-directing the infant's gaze at the screen through verbal communication via microphone (e.g. "Look (baby's name)! What s next?").

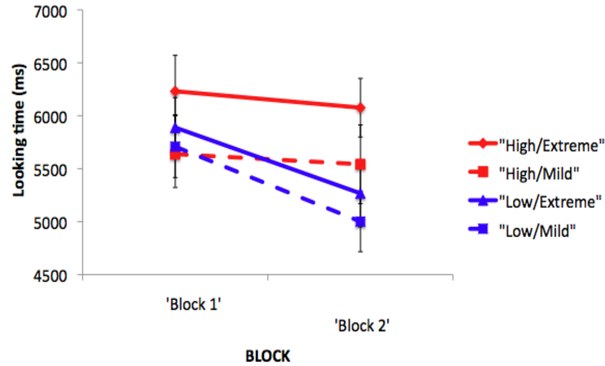


Figure 2: Looking time during familiarization.

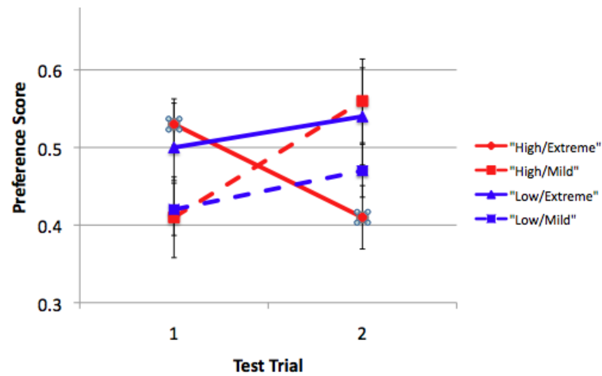


Figure 3: Looking time during categorization test trials

### Results

The video streams from left and right cameras were manually scored for infants' gaze direction (left vs. right).

**Looking time during familiarization** A mixed ANOVA on the looking times for familiarization trials (see Figure 2) with within-subjects factor Block (Block 1: trials 1-4, Block 2: trials 5-8) and between-subjects factors Distance (low, high) and Start/End Stimulus (mild, extreme) revealed a main effect of Block (F(1,98)=8.253, p=.005) and a Block x Distance interaction (F(1,98)=4.072, p=.046). T-Tests confirmed that looking time decreased between Block 1 and 2 in the low-distance conditions, but remained the same in the high-distance conditions.

**Categorization: Test trials 1 and 2** In order to assess categorization performance, looking preference scores were obtained for each test trial from each participant by dividing the time spent looking at the average stimulus by the time spent looking at either test stimulus, average or peripheral (see Figure 3 for results). The resulting preference scores from the first test trial were subjected to an ANOVA with factors distance (low vs. high) and start and end stimulus (extreme vs. mild). This revealed a main effect of start/end stimulus ($F(1,92)=6.242$, $p=.014$). All other effects remained non-significant (all $F$s $< .31$, $p$s $> .57$). Follow-up t-tests showed that infants in the mild start/end stimulus conditions exhibited a preference for the peripheral stimulus on Test trial 1 (Looking proportion for average stimulus 3333: M=41.7%, SE=2.9%; t(50)=2.882, p=.006), whereas infants in the extreme start/end stimulus condition exhibited no preference (Looking proportion for stimulus 3333: M=51.4%, SE=2.4%; *t*(44)=.564, *p*=.576). On Test trial 2, the observed pattern of behavior was different. An ANOVA with factors distance and start/end stimulus revealed a significant interaction between distance and start/end stimulus (F(1,93)=5.534, p=.021). No other effects were significant (all $F$s $< .75$, $p$s $> .39$). Further analysis of the interaction showed that only infants in the high/extreme condition had a significant preference, again for the peripheral stimulus (*t*(23)=2.198, *p*=.038). Preferential looking in all other conditions did not differ from chance (0.5; all $t$s $< 1.2$, $p$s $> .24$).

**Novelty preference: Test trials 3 and 4** In order to establish that looking on the first test trials was driven by novelty preference rather than familiarity preference, preference scores were obtained for test trials 3 and 4 by dividing the amount of looking at the novel stimulus by the total looking time for each trial. The preference scores were subjected to an ANOVA with factors Test type (novel vs. average, novel vs. peripheral), Trial order, (novel vs. average first, novel vs. average second), Distance, and Start- and End-stimulus. This revealed a main effect of Trial order (F(1,84)=4.895, p=.03). All other effects were non-significant. Follow-up t-tests showed that there was always a significant novelty preference on the first of the two trials (Novel vs. Peripheral: *M*=.66, *SE*=.04, *t*(50)=4.6, $p < .001$; Novel vs. Average: *M*=.57, *SE*=.03; *t*(46)=2.03, *p*=.048), but on the second test trial infants only exhibited a (marginally) significant novelty preference if they had previously seen the pairing of the novel stimulus and a peripheral stimulus, and were now looking at the average and the novel stimulus (*M*=.57, *SE*=.04; $t(46) = 2.0$,p=.051). Infants who saw the novel stimulus paired with the overall average first did not exhibit a preference on the second novelty preference test trial (*M*=.55, *SE*=.03;*t*(46)=1.6, *p*=.107). These results are consistent with Mather & Plunkett's (2011) findings.

### Discussion of Experimental Findings

The main effect of start and end stimulus found for Test trial 1 suggests that a recency or primacy effect determines looking on Test trial 1. This is consistent with the model predictions described above. As expected, infants who saw a mild stimulus on familiarization trials 1 and 8 exhibited a preference for the peripheral stimulus on Test trial 1. For these groups, the average stimulus appears particularly familiar when they get to Test trial 1. Infants in the groups with extreme start and end stimuli on the other hand do not exhibit any preference on Test trial 1. This is, empirically, the more surprising result: Younger (1985) reported merely an overall preference for the peripheral stimulus on the equivalent test trial. A conservative interpretation of our data would assume that no category was formed in the extreme conditions. However, the model's performance indicates that instead of no category being formed the category's central tendency is merely closer to the peripheral stimulus than in the mild condition. This suggests that the null preference we observe is merely due to the fact that with this shifted category representation the average and peripheral test stimuli are equally interesting to the infants.

Test trial 2 is harder to interpret, as the pattern of preferences is very different from Test trial 1. Such order effects are common in familiarization / novelty preference paradigms (for a discussion see Schöner and Thelen (2006)). A likely cause for this is that learning does not stop at the end of familiarization: infants may incorporate both test stimuli presented on Test trial 1 in their category, and this will influence looking preferences on Test trial 2. Further work is required in order to explain the exact patterns observed, but the fact that all four conditions differ on this test trial indicate that Euclidean distance has a secondary impact, i.e. Mather and Plunkett (2011) assumption still holds. Looking times during familiarization imply that Euclidean distance is an important factor for maintaining infants interest during learning. Infants in the high distance conditions maintained looking, whereas infants looking times in the low distance conditions decreased, indicating that they began to habituate. This behavior is consistent with Mather and Plunkett (2011) interpretation of the impact of Euclidean distance on infants' attention.

## General Discussion

Decades of research on early categorization have assumed that categorization patterns were not impacted by the order of presentation of the familiarization stimuli. familiarization sequences were randomised and results averaged over different realisations. Recently, Mather and Plunkett (2011) challenged this view and showed that the order of presentation of the familiarization stimuli had an impact on infant category formation. Reasons for this behavior are yet unclear, which is why we decided to implement a model so as to evaluate the role of the order of presentation of the stimuli on the pattern of categorization.

First, we created a variant of the neural network model introduced by Gliozzi et al. (2009). The model is built with a simple self-organizing map and successfully reproduces Mather and Plunkett (2011)'s results. However, the model proposes an explanation of these results which is different from that provided by Mather and Plunkett (2011). In particular, the model predicts a primacy/recency effect: category formation depends on the nature of the first or last stimuli used in the training sequence.

The model's predictions have been confirmed by data from infants. 104 10-month-old infants were familiarized with sequences in the same four conditions presented to the network. Novelty preference scores on test indicate that responses are mainly driven by primacy/recency effects, whereas the average Euclidean distance influenced looking time during familiarization. This implies that, at odds to common assumptions about familiarization, 10-month-old novelty preference responses can be heavily influenced by familiarization stimuli at the start or end of the familiarization sequence, a factor which is often ignored in infant familiarization studies.

Our results are consistent with both primacy and recency effects, and future research will determine whether category formation is more heavily influenced by either primacy or recency.

In conclusion, this paper questions the traditional view underlying the novelty preference procedure suggesting that familiarization stimuli are categorized in an abstract representation of all the stimuli. In this traditional view, the representation formed is independent from familiarization contingencies. On the contrary, our results show that infants are sensitive to the order of presentation of the stimuli and support models that advocate infant category learning as an incremental process by which, on a moment-by-moment basis, infant refine the boundaries of new categories (Gliozzi et al., 2009; Gureckis & Love, 2004; Westermann & Mareschal, 2004). In contrast, our results cannot be explained by models in which the infants only establish the category boundaries once they have sampled all familiarization items.

## Acknowledgments

This research was supported by Wellcome Grant 084386/Z/07/Z to Kim Plunkett and the Winkler Career Development Fellowship at St Hugh's College, University of Oxford, to Nadja Althaus. We wish to thank all parents and infants who participated in this study.

## References

Bomba, P. C., & Siqueland, E. R. (1983). The nature and structure of infant form categories. *Journal of Experimental Child Psychology*, *35*, 294–328.

Eimas, P., & Quinn, P. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development*, *65*, 903–917.

Gliozzi, V., Mayor, J., Hu, J.-F., & Plunkett, K. (2009). Labels as Features (Not Names) for Infant Categorization: A Neurocomputational Approach. *Cognitive Science*, *33*, 709–738.

Gureckis, T., & Love, B. (2004). Common mechanisms in infant and adult category learning. *Infancy*, *5(2)*, 173–198.

Kohonen, T. (1993). Physiological interpretation of the self-organizing map algorithm. *Neural Networks*, *6*(7), 895–905.

Kohonen, T. (1997). Self-Organizing Maps, volume 30 of Springer Series in Information Sciences. *Springer, Berlin*.

Kovack-Lesh, K. A., & Oakes, L. M. (2007). Hold your horses: How exposure to different items influences infant categorization. *Journal of Experimental Child Psychology*, *98*, 69–93.

Mareschal, D., & French, R. (2000). Mechanisms of categorization in infancy. *Infancy*, *1*, 59–76.

Mareschal, D., & Quinn, P. C. (2001, Sep). Categorization in infancy. *Trends in Cognitive Sciences*, *5*, 443–450.

Mather, E., & Plunkett, K. (2011). Same items, different order: Effects of temporal variability on infant categorization. *Cognition*, *119*, 438–447.

P.C.Quinn, & R.S.Bhatt. (2010). Learning perceptual organization in infancy:The effect of simultaneous versus sequential variability experience. *Perception*, *39*, 795–806.

Schöner, G., & Thelen, E. (2006). Using dynamic field theory to rethink infant habituation. *Psychological Review*, *113*, 273–298.

Westermann, G., & Mareschal, D. (2004). From Parts to Wholes: Mechanisms of Development in Infant Visual Object Processing. *Infancy*, *5*(2), 131–151.

Younger, B. (1985). The segregation of items into categories by ten-month-old infants. *Child Development*, *56*, 1574–1583.