Visual Cognition, 2013
Vol. 21, No. 6, 726–738, http://dx.doi.org/10.1080/13506285.2013.819825

Routledge
Taylor & Francis Group

The importance of feature distribution and correlation for simulating 3 to 4-month-old infants' visual categorization processes

Martial Mermillod[1,2], Nicolas Vermeulen[3,4], Gwenael Kaminski[5], Edouard Gentaz[6], and Patrick Bonin[7,2]

[1]Université Grenoble Alpes, LPNC, F-38040, Grenoble & CNRS, LPNC UMR 5105, F-38040, Grenoble, France
[2]Institut Universitaire de France
[3]Psychological Sciences Research Institute, Université Catholique de Louvain, Belgium
[4]Fund for Scientific Research, Belgium
[5]Laboratoire CLLE-LTC, Toulouse University, Maison de la Recherche, France
[6]Université de Genève, Switzerland
[7]Université de Bourgogne and CNRS, Dijon, France

(Received 2 April 2013; accepted 21 June 2013)

Mareschal, French, and Quinn (2000) and Mareschal, Quinn, and French (2002) have proposed a connectionist model of visual categorization in 3- to 4-month-old infants that simulates and predicts previously unexplained behavioural effects such as the asymmetric categorization effect (French, Mareschal, Mermillod, & Quinn, 2004). In the current paper, we show that the model's ability to simulate the asymmetry depends on the correlational structure of the stimuli. These results are important given that adults (Anderson & Fincham, 1996) as well as infants (Younger & Cohen, 1986) are able to rely on correlation information to perform visual categorization. At a behavioural level, the current paper suggests that pure bottom-up processes, based on the correlational structure of the categories, could explain the disappearance of the asymmetry in older 10-month-old infants (Furrer & Younger, 2005). Moreover, our results also raise new challenges for visual categorization models that attempt to simulate the shift from asymmetric

Please address all correspondence to Martial Mermillod, Laboratoire de Psychologie et NeuroCognition, Université Pierre Mendès France, BP 47, 38040 Grenoble Cedex 9, France. E-mail: Martial.Mermillod@upmf-grenoble.fr

This work was supported by the French CNRS (UMR 5105) and the Institut Universitaire de France.

categorization in 3- to 4-month-old to symmetric categorization in 10-month-old infants (Shultz & Cohen, 2004; Westermann & Mareschal, 2004, 2012).

Keywords: Visual categorization; Infancy; Neural network modelling.

Science evolves by revealing the limitations and constraints of theoretical models that attempt to explain empirical results. In psychology, parallel and distributed processes have been shown to provide an extremely reliable and parsimonious account of many different cognitive processes including infants' visual categorization skills, i.e., the capacity to mentally group together similar things that exist in the world. The aim of this brief report is to reveal the limitations and constraints of parallel and distributed processes for the simulation of categorization capacities in 3- to 4-month-old infants (French, Mareschal, Mermillod, & Quinn, 2004; French, Mermillod, Quinn, & Mareschal, 2001; Mareschal & French, 2000; Mareschal, French, & Quinn, 2000; Mareschal, Quinn, & French, 2002). As stated in a recent paper (Westermann & Mareschal, 2012, p. 6), "[w]hile these models were successful in explaining mechanisms underlying object categorization in 3 to 4-month-olds, they did not account for developmental change". In the current paper, we will show that the original model (Mareschal & French, 2000; Mareschal et al., 2000), like other models designed to simulate this asymmetry (Shultz & Cohen, 2004; Westermann & Mareschal, 2004, 2012), is subject to significant constraints when attempting to simulate 3- to 4-month-olds' visual categorization processes and is actually a better model of the processes at work in older infants.

Simulating the asymmetric categorization effect observed in 3- to 4-month-old infants

The empirical effect, as revealed by basic-level visual categories in Quinn, Eimas, and Rosenkrantz (1993), was as follows: When 3- to 4-month-old infants are familiarized on the Cat category, they seem to form a perceptual representation of cats that excludes dogs. On the other hand, when they are familiarized on the Dog category, they are not able to categorize a new cat as coming from a new category. That is, the Cat and Dog categories have asymmetric exclusivity: Cat excludes dogs, but Dog does not exclude cats (Quinn et al., 1993). In an attempt to propose a theoretical underpinning for this effect, Mareschal et al. (2000) suggested that the exclusivity difference might reflect an asymmetric relation in the distribution of feature values used to characterize the two sets of images. Indeed, these authors found that, in the image set used by Quinn et al. (1993), the large majority of cat feature values were *subsumed within* the distribution of the broader dog feature values. Furthermore, the majority of dog feature values did not fall within the distribution of cat feature values. Thus, at the level of individual features, most cats were plausible dogs, but most dogs were not plausible cats.

On the basis of connectionist modelling, Mareschal and French (2000) and Mareschal et al. (2000, 2002) have proposed a connectionist autoencoder which makes it possible to replicate the effect observed in 3- to 4-month-olds. The analogy between the amount of error produced by the neural network and the infant's fixation time is as follows. During fixation, the infant is thought to construct an interactive representation (Sokolov, 1963). The infant ceases to examine a stimulus once he or she has developed a reliable internal representation of that stimulus. Similarly, learning in an autoencoder consists of an iterative process of representation construction. The network tries to adjust its internal representation of the input (i.e., by adjusting the network weights) until it encodes enough information about the input stimulus to be able to reconstruct that stimulus in the output. The central assumption of this account is that stimuli that produce a high level of initial output errors (that are poorly autoencoded) will take longer to encode (and thus reduce the output errors) than stimuli that produce low initial output error rates. This is because the network will require more iterations in order to adjust the internal representations appropriately. Hence, the output error rate produced when a novel stimulus is presented to the network is interpreted as being equivalent to the time the infant spends looking at the stimulus. The model has been found to successfully capture the category-based looking time behaviours of 3- to 4-month-olds, including the subtle asymmetric exclusivity in the extensions of the categories tested, such as Cat and Dog (French et al., 2004; Mareschal et al., 2000, 2002; Quinn et al., 1993).

Interestingly, Furrer and Younger (2005) reported that the asymmetric categorization effect completely disappears in 10-month-old infants. Research in the field of developmental psychology has therefore attempted to investigate the shift in the asymmetry effect from that observed in the behaviour of 3- to 4-month-old infants to that observed in 10-month-old infants. The latest thinking in the field of developmental psychology therefore assumes that what still needs to be understood is the shift in behaviour between the ages of 3–4 months and 10 months (Furrer & Younger, 2005; Westermann & Mareschal, 2012). For instance, Furrer and Younger suggested that the disappearance of the asymmetric categorization effect in 10-month-old infants could be related to the top-down conceptual knowledge of these older infants (related to the ability to assign "cat" and "dog" label categories). Here, we propose the alternative hypothesis that the disappearance of this asymmetry could be the result of simple bottom-up processes that relate to the ability of 8- to 10-month-old infants to take account of the correlational structure of the categories.

The importance of correlational structures for visual categorization

A variety of papers have shown that adults (Anderson & Fincham, 1996; Crawford, Huttenlocher, & Hedges, 2006; R. D. Thomas, 1998), as well as infants as young as 10 months of age (Gureckis & Love, 2004; Shultz & Cohen, 2004; Younger & Cohen, 1986), are able to use information about correlations among the different features of the stimuli to produce correct categorizations of visual and natural objects. Conversely, 4-month-old infants do not seem to be able to use correlation information to perform categorization tasks (Younger & Cohen, 1986). Therefore, investigating how the theoretical model proposed in the literature responds to correlated attributes is both *important* in the light of the fact that this information is broadly used by humans to perform categorization (Anderson & Fincham, 1996; Younger & Cohen, 1986) and *crucial* when we consider that most natural categories have features that naturally covary. This idea is supported by the fact that the differences in the size of different exemplars of any given species remain constant between scales (Mandelbrot, 1977), thus leading to strong correlations among the different features of each exemplar. This characteristic of living categories is an old and well-known property in the life sciences and is often used for the characterization of different species (Cavallini, 1995) or in studies of the maturation of living organisms (Hughes & Tanner, 1970).

At a computational level, this question raises several related issues since parallel and distributed connectionist networks are able to use not only the distribution of feature values but also correlational information in order to categorize visual stimuli, for example among different Gabor filters simulating V1 neuron receptive fields (Mermillod, Bonin, Mondillon, Alleysson, & Vermeulen, 2010; Mermillod, Vermeulen, Lundqvist, & Niedenthal, 2009). Therefore, based on the assumption that parallel and distributed processes are able to simulate visual categorization behaviour in 3- to 4-month-old infants, the developmental sciences have embarked on a wide-ranging debate aimed at understanding how artificial cognitive systems might explain the shift from feature-based to correlation-based visual categorization stimuli (Gureckis & Love, 2004; Shultz & Cohen, 2004; Westermann & Mareschal, 2004; but see also M. S. C. Thomas, 2004).

However, one recent paper (Mermillod, Vermeulen, Kaminski, Gentaz, & Bonin, in press) has shown that the way in which the stimuli were encoded in French et al. (2004) removed almost all the information relating to the correlations among features of the cats and dogs categories. Therefore, in order to test the impact of feature distribution in addition to that of correlation information, we artificially created a pattern of correlated features, which, however, had the same distribution properties as the original simulations (French et al., 2004; Mareschal et al., 2000). The aim of the current paper is to show that a connectionist network is able to isolate and differentiate between the two categories on the basis of correlational information, even without the same distributions of feature values and inclusion relationships present in the original simulations.

CONNECTIONIST SIMULATION: USING A CORRELATIONAL STRUCTURE TO ELIMINATE THE ASYMMETRIC CATEGORIZATION EFFECT

Stimuli

As in the original simulations, the networks were trained on the same 10 measurements of the stimuli used to familiarize the infants (i.e., horizontal extent, vertical extent, leg length, head length, head width, eye separation, ear separation, nose length, nose width). However, in order to test the effect of correlated attributes, we artificially modified the input pattern (i.e., the feature values of the original stimuli used in Simulation 1 as input/output for neural network simulations) in order to increase correlations among features within each category. The set of correlated features was generated by using one dog vector and one cat vector from the original simulation (French et al., 2004). This vector was then multiplied 18 times by a scalar number in order to create a pattern of perfectly correlated features for 18 exemplars of one and the same category. Finally, we added some noise in this matrix of perfectly correlated features by adding a small random number in the range 0–4 for each feature. The aim was to obtain realistic values (i.e., not perfectly correlated) of correlated features (Tables 1 and 2). Finally, we ensured that all the features had the same distribution and inclusion relationship (Figure 1) as in the original simulations (French et al., 2004; Mareschal et al., 2000). We predicted that the introduction of correlation information might allow the neural network to differentiate the original categories despite their feature value distributions and inclusion relationships.

Neural network procedure

We used exactly the same standard 10-8-10 feedforward backpropagation autoencoder network as used by French et al. (2004) to model infant categorization. This means that we had 10 input and output units (the measurements of the 10 features) and 8 hidden units. Training parameters were identical to those used in the original simulation.[1] The vectors used for this simulation were normalized between 0 and 1, feature by feature, across

[1] Learning rate: 0.1, momentum: 0.9, Fahlman offset: 0.1.

TABLE 1
Correlations among features of the Dog category

	Head length	*Head width*	*Eye separation*	*Ear separation*	*Ear length*	*Nose length*	*Nose width*	*Leg length*	*Vertical extent*	*Horizontal extent*
Head length	1	0.88	0.47	0.79	0.66	0.66	0.29	0.87	0.88	0.93
Head width	0.88	1	0.46	0.83	0.66	0.73	0.40	0.91	0.87	0.90
Eye separation	0.47	0.46	1	0.65	0.60	0.42	0.50	0.52	0.64	0.60
Ear separation	0.79	0.83	.065	1	0.61	0.72	0.53	0.86	0.92	0.92
Ear length	0.66	0.66	0.60	0.61	1	0.49	0.43	0.74	0.71	0.74
Nose length	0.66	0.73	0.42	0.72	0.49	1	0.38	0.72	0.68	0.73
Nose width	0.29	0.40	0.50	0.53	0.43	0.38	1	0.41	0.39	0.43
Leg length	0.87	0.91	0.52	0.86	0.74	0.72	0.41	1	0.93	0.95
Vertical extent	0.88	0.87	0.64	0.92	0.71	0.68	0.39	0.93	1	0.95
Horizontal extent	0.93	0.90	0.60	0.92	0.74	0.73	0.43	0.95	0.95	1

TABLE 2
Correlations among features of the Cat category

	Head length	*Head width*	*Eye separation*	*Ear separation*	*Ear length*	*Nose length*	*Nose width*	*Leg length*	*Vertical extent*	*Horizontal extent*
Head length	1	0.97	0.80	0.94	0.94	0.84	0.58	0.98	0.97	0.97
Head width	0.97	1	0.82	0.95	0.94	0.86	0.62	0.96	0.98	0.98
Eye separation	0.80	0.82	1	0.79	0.87	0.80	0.52	0.83	0.86	0.83
Ear separation	0.94	0.95	0.79	1	0.92	0.89	0.59	0.97	0.97	0.97
Ear length	0.94	0.94	0.87	0.92	1	0.88	0.62	0.95	0.96	0.95
Nose length	0.84	0.86	0.80	0.89	0.88	1	0.58	0.85	0.88	0.87
Nose width	0.58	0.62	0.52	0.59	0.62	0.58	1	0.59	0.62	0.63
Leg length	0.98	0.96	0.83	0.97	0.95	0.85	0.59	1	0.99	0.99
Vertical extent	0.97	0.98	0.86	0.97	0.96	0.88	0.62	0.99	1	0.99
Horizontal extent	0.97	0.98	0.83	0.97	0.95	0.87	0.63	0.99	0.99	1

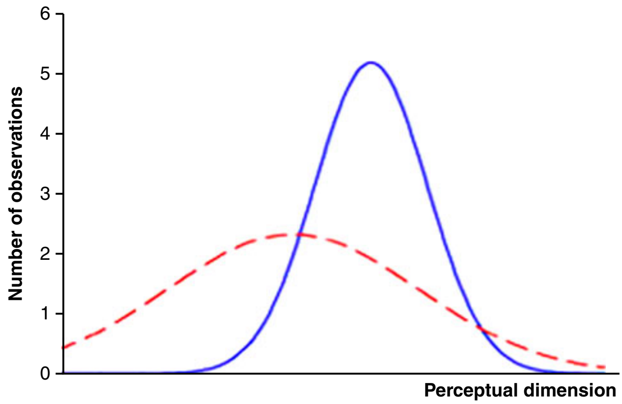

Figure 1. Example of the feature distributions for the cat and dog exemplars. The dog distribution (in red) subsumes the cat distribution (in blue). To view this figure in colour, please see the online issue of the Journal.

all of the 36 stimuli (i.e., 18 dogs and 18 cats). The networks were trained in six batches of two patterns each until all outputs were within 0.2 of their target values or for a maximum of 250 weight updates. Each batch of two patterns was trained fully before the next batch of two patterns was presented. This procedure was intended to mimic the paired presentation technique used with infants. The error, as in the original simulations (French et al., 2004; Mareschal & French, 1997; Mareschal et al., 2000), was the maximum MSE produced by the activation of the output nodes compared to the desired output. The results were averaged over 50 neural networks with different initial random weights. Each network was first trained on 12 randomly selected exemplars selected from either the set of cat images or the set of dog images. Once a network had learned to autoencode these images to criteria, it was tested on six novel images from each of the two categories. The output error level produced in response to the novel images was a measure of the novelty of the images and was assumed to reflect the infant's looking time.

RESULTS

We observed a significant interaction between the type of training category and the test category effect, $F(1, 98) = 164.19$, $p < .001$, $\eta^2 = .59$. The model brought about a significant increase in errors for cats after training on the broad Dog category, $F(1, 98) = 74.34$, $p < .001$. However, after being trained on the narrow Cat category, the model also produced more errors on the novel dog than on the novel cat exemplars, $F(1, 98) = 90.23$, $p < .001$. As expected, the correlation information provided by this new pattern of data

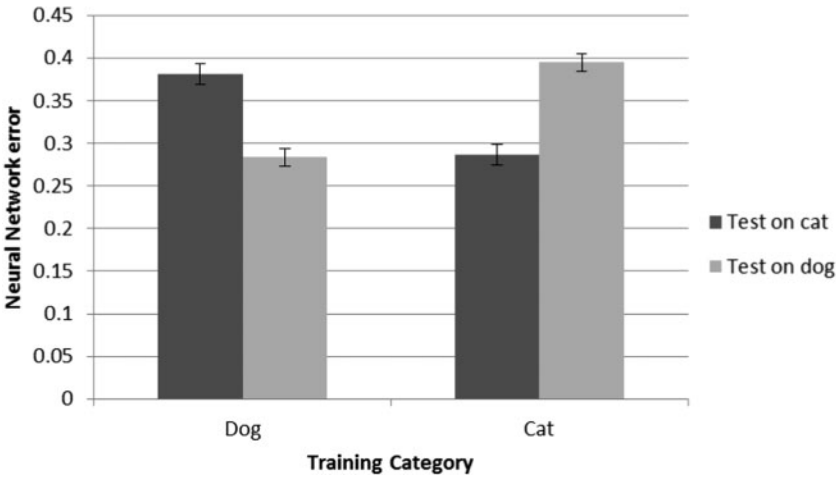

Figure 2. Average neural network error produced in response to novel dogs or cats following training with either the Cat or Dog category.

was sufficient to eliminate the asymmetry despite the feature value distribution and inclusion relationship assumed to produce the asymmetry in the original simulations (Figure 2).

DISCUSSION

No asymmetry was observed despite the variance distribution and inclusion relationship present in the inputs. The connectionist model is able to use the correlational structure of the two categories to identify the exemplars as coming from different categories. It therefore seems that the original model revealed an asymmetric categorization effect because of poor correlations in the input data (Mermillod et al., in press). We assume that this could be due to the way the data was encoded (measurements of responses to pictures having different scales or orientations, for instance) and suggest that a more realistic model of vision, for example using Gabor patches as a simulation of V1 neural processes, would probably retain correlations across features (Kaminski, Méary, Mermillod, & Gentaz, 2010, 2011; Mermillod, Vuilleumier, Peyrin, Alleysson, & Marendaz, 2009).

CONCLUSION

These data clearly show that the manipulation of the distribution and inclusion relationship is *necessary* but not in itself *sufficient* to explain the asymmetric categorization effect reported by Quinn et al. (1993). Feature correlation was not considered as a factor in the original model. However, new simulations with correlated feature sets suggest that it has a massive impact on the performance of the model. Regarding the visual categorization properties at a behavioural level, 3- to 4-month-old infants do not as yet seem to be sensitive to correlation information (Younger & Cohen, 1983, 1986). This behavioural characteristic therefore makes them sensitive only to the feature value distribution and inclusion relationship, irrespective of the correlational structure of the categories, and consequently allows the asymmetry to appear. This hypothesis supports and extends empirical evidence showing that the asymmetric categorization effect was reproduced in 4-month-old but disappeared in 10-month-old infants (Furrer & Younger, 2005). However, Furrer and Younger (2005) assumed the use of top-down conceptual knowledge (although this top-down hypothesis was not clearly set out in their paper). Contrary to this hypothesis that top-down conceptual knowledge is involved in differentiating the two categories, and in accordance with various papers in the literature showing the importance of correlations for visual categorization (Gureckis & Love, 2004; Shultz & Cohen, 2004; Younger & Cohen, 1986), we propose here the parsimonious hypothesis that simple bottom-up processes, related to the use of correlational information provided by the stimuli, could be sufficient to remove the asymmetry in 10-month-old infants. It might be possible to test these hypotheses at the empirical level using stimuli similar to those used by Younger and Cohen (1986), namely artificial animals comprising a set of perceptually controlled features (legs, body, neck, head, etc.).

At a computational level, the mainstream of research in visual categorization based on prototype models (Rosch, 1978), exemplar models (Kruschke, 1992; Nosofsky, 1986), or even connectionist models (French et al., 2004) has largely focused on feature value distributions and, to a lesser extent, correlated feature (Anderson & Fincham, 1996; Shultz & Cohen, 2004; Younger & Cohen, 1986). However, our current results raise new challenges for visual categorization models that attempt to simulate the shift from asymmetric categorization in 3- to 4-month-old to symmetric categorization in 10-month-old infants (Shultz & Cohen, 2004; Westermann & Mareschal, 2004, 2012). Although feature value distribution is an important factor that should be taken into account when simulating visual categorization processes during infancy, correlations among features are at least as important, and a reliable model of visual categorization has to take both constraints into account at the level of the input data in order to simulate this shift from asymmetric categorization to symmetric categorization (M. S. C. Thomas, 2004). Note that the current connectionist simulations clearly show that the ability (or the inability) of a simple connectionist classifier to simulate (or not) the asymmetric categorization effect is largely depending on the method used to encode the stimuli at a perceptual level. We can assume that a more realistic model of visual perception, allowing the precise encoding of the correlations among different features of natural objects, will be crucial to test the respective effects of variance distribution and correlation information. Indeed, empirical studies investigating the effects of these factors also have to ensure a correct coding of the statistical properties of the stimuli by the participants. We can assume that it should be the case in the haptic modality when the infants can manipulate the objects (i.e., encoding the correlation information among the different features) but not necessarily for static images presented in the visual modality. Therefore, we suggest that a 3-D video of the stimuli could be probably necessary to ensure a correct coding of the correlations among features for subsequent empirical experiments.

To our knowledge, no computational model in the literature has addressed this question by simultaneously taking account of all three factors—(1) feature value distribution, (2) inclusion relationship, and (3) correlated (vs. uncorrelated) features—when simulating this surprising asymmetry. In our view, this represents a major shortcoming in the literature since different computational models, as well as humans, are sensitive to these different factors when performing visual categorization (Gureckis & Love, 2004; Shultz & Cohen, 2004; Westermann & Mareschal, 2004; Younger & Cohen, 1986). A reliable computational model of visual categorization should be able to simulate human behaviour even if we add new constraints to the model. By taking into account these three factors in combination with the behavioural data available in the literature (Furrer & Younger, 2005; Younger & Cohen, 1986), we have clearly shown that a simple connectionist neural network probably provides a better model of 10-month-old infants' categorization capacities, but not of the asymmetry observed among 3- to 4-month-olds that still remains to be understood.

REFERENCES

Anderson, J. R., & Fincham, J. M. (1996). Categorization and sensitivity to correlation. *Journal of experimental psychology: Learning, Memory, and Cognition*, *22*, 259–277.

Cavallini, P. (1995). Variation in the body size of the red fox. *Annales Zoologici Fennici*, *32*(4), 421–427.

Crawford, L. E., Huttenlocher, J., & Hedges, L. V. (2006). Within category feature correlations and Bayesian adjustment strategies. *Psychonomic Bulletin and Review*, *13*, 245–225.

French, R. M., Mareschal, D., Mermillod, M., & Quinn, P. C. (2004). The role of bottom-up processing in perceptual categorization by 3- to 4-month-old infants: Simulations and data. *Journal of Experimental Psychology: General*, *133*(3), 382–397.

French, R. M., Mermillod, M., Quinn, P. C., & Mareschal, D. (2001). Reversing category exclusivities in infant perceptual categorization: Simulations and data. In K. Stenning & J. Moore (Eds.), *Proceedings of the 23rd annual conference of the Cognitive Science Society* (pp. 307–312). Mahwah, NJ: Lawrence Erlbaum Associates.

Furrer, S. D., & Younger, B. A. (2005). Beyond the distributional input: A developmental investigation of asymmetry in infants' categorization of cats and dogs. *Developmental Science*, *8*(6), 544–550.

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

Hughes, P. C., & Tanner, J. M. (1970). A longitudinal study of the growth of the black-hooded rat: Methods of measurement and rates of growth for skull, limbs, pelvis, nose-rump and tail lengths. *Journal of Anatomy*, *106*(2), 349–370.

Kaminski, G., Méary, D., Mermillod, M., & Gentaz, E. (2010). Factors affecting the ability to assess facial resemblance between parents and newborns in humans. *Perception*, *39*(6), 807–818.

Kaminski, G., Méary, D., Mermillod, M., & Gentaz, E. (2011). Is it a girl or a boy? Behavioural and computational approaches of gender categorization. *Attention, Perception and Psychophysics*, *73*(5), 1344–1349.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.

Mandelbrot, B. B. (1977). *The fractal geometry of nature*. New York, NY: Freeman.

Mareschal, D., & French, R. M. (1997). A connectionist account of interference effects in early infant memory and categorization. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the 19th annual conference of the Cognitive Science Society* (pp. 484–489). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Mareschal, D., French, R. M., & Quinn, P. C. (2000). A connectionist account of asymmetric category learning in early infancy. *Developmental Psychology*, *36*, 635–645.

Mareschal, D., Quinn, P. C., & French, R. M. (2002). Asymmetric interference in 3- to 4-month-olds' sequential category learning. *Cognitive Science*, *26*, 377–389.

Mermillod, M., Bonin, P., Mondillon, L., Alleysson, D., & Vermeulen, N. (2010). Coarse scales are sufficient for efficient categorization of emotional facial expressions: Evidence from neural computation. *Neurocomputing*, *73*, 2522–2531.

Mermillod, M., Vermeulen, N., Kaminsky, G., Gentaz, E., & Bonin, P. (in press). Theoretical and computational limitations in simulating 3- to 4-month-old infants' categorization processes. In J. Mayor (Ed.), *Proceedings of the 13th Neural Computation and Psychology workshop* (Vol. 21). Singapore: World Scientific.

Mermillod, M., Vermeulen, N., Lundqvist, D., & Niedenthal, P. M. (2009). Neural computation as a tool to differentiate perceptual from emotional processes: The case of anger superiority effect. *Cognition*, *110*(3), 346–357.

Mermillod, M., Vuilleumier, P., Peyrin, C., Alleysson, D., & Marendaz, C. (2009). The importance of low spatial frequency information for recognizing fearful facial expressions. *Connection Science*, *21*(1), 75–83.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3- and 4-month-old infants. *Perception*, *22*, 463–475.

Rosch, E. (1978). *Principles of categorization*. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shultz, T. R., & Cohen, L. B. (2004). Modeling age differences in infant category learning. *Infancy*, *5*(2), 153–171.

Sokolov, E. N. (1963). *Perception and the conditioned reflex*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Thomas, M. S. C. (2004). How do simple connectionist networks achieve a shift from "featural" to "correlational" processing in categorization? *Infancy*, *5*(2), 199–207.

Thomas, R. D. (1998). Learning correlations in categorization tasks using large, ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 119–143.

Westermann, G., & Mareschal, D. (2004). From parts to wholes: Mechanisms of development in infant visual object processing. *Infancy*, *5*(2), 131–151.

Westermann, G., & Mareschal, D. (2012). Mechanisms of developmental change in infant categorization. *Cognitive Development*, *27*, 367–382.

Younger, B., & Cohen, L. B. (1983). Infant perception of correlations among attributes. *Child Development*, *54*, 858–867.

Younger, B., & Cohen, L. B. (1986). Developmental change in infants' perception of correlations among attributes. *Child Development*, *57*, 803–815.