

The Importance of Starting Blurry: Simulating Improved Basic-Level Category Learning in Infants Due to Weak Visual Acuity

Robert M. French, Martial Mermillod,
Quantitative Psychology and Cognitive Science
Psychology, U. of Liège, Belgium
{rfrench, mmermillod}@ulg.ac.be

Alan Chauvin,
Psychology, University of Grenoble, France
chauvin@lis-inpg.fr

Paul C. Quinn
Psychology, Washington & Jefferson College
Washington, PA, USA
pquinn@washjeff.edu

Denis Mareschal
Psychology, Birkbeck College,
London, UK
d.mareschal@bbk.ac.uk

Abstract

At the earliest ages of development, perceptual maturation is generally considered as a functional constraint to recognize or categorize the stimuli of the environment. However, using a computer simulation of retinal development using Gabor wavelets to simulate the output of the V1 complex cells (Jones & Palmer, 1987), we showed that reducing the range of the spatial frequencies from the retinal map to V1 decreases the variance distribution within a category. The consequence of this is to decrease the difference between two exemplars of the *same* category, but to increase the difference between exemplars from two *different* categories. These results show that reduced perceptual acuity produces an advantage for differentiating basic-level categories. Finally, we show that the present simulations using Gabor-filtered input instead of feature-based input coding provide a pattern of statistical data convergent with previously published results in infant categorization (e.g., Mareschal & French, 1997; Mareschal et al, 2000; French et al, 2001).

Background

This paper builds on earlier work by Quinn, Eimas, and Rosenkrantz (1993), Mareschal and French (1997), Mareschal, French, and Quinn (2000) and French, Mermillod, Quinn, and Mareschal (2001). Quinn et al. (1993) reported the following categorization asymmetry. Infants familiarized with a number of exemplars of cats show significantly increased interest when subsequently tested on an exemplar of a novel dog compared to a novel cat. However, if the infants are first familiarized with images of dogs and then tested on a novel dog and a novel cat, there is no significant difference in interest between the two test stimuli. Mareschal and French (1997) and Mareschal et al. (2000) attributed this to the greater variance of the “dog” stimuli set compared to the “cat” stimuli set, the interpretation being that the Dog category largely subsumed the Cat category. Thus, an infant familiarized on the less variable category, Cat, would, in general, view an exemplar of

a dog as a novel stimulus, whereas an infant familiarized on the more variable category, Dog, would tend to perceive a cat exemplar as simply belonging to the already-familiar Dog category. This, we claimed, explained the asymmetric levels of attention that Quinn et al. (1993) had observed. To further test this hypothesis, French et al. (2001) artificially reversed the inclusion relationship by carefully selecting breeds of dogs that were relatively similar (i.e., low variance) and highly variable breeds of cats. The connectionist computer model predicted, and the experimental results with infants subsequently confirmed (French et al., 2001), a reversal in the categorization asymmetry observed by Quinn et al. (1993).

However, one outstanding question remained. Even though, intuitively, the variability of the Cat category appears to be less than that of the Dog category, how could one be *sure* of this in any quantifiable way? Mareschal and French (1997) and Mareschal et al. (2000) handled this as follows. They originally selected ten features common to both cats and dogs (head length, head width, eye separation, ear separation, ear length, nose length, nose width, leg length, vertical extent, and horizontal extent) and measured the values associated with these features for each of the photos of the 18 Cat exemplars and 18 Dog exemplars used in their experiments. Even though this choice of features was based on experimental data where infants typically look at the head and face region of the stimulus when they observe an animal (Quinn & Eimas, 1996; Spencer, Quinn, Johnson, & Karmiloff-Smith, 1997), how could we be sure that the set of perceptual features that we had chosen corresponded to those features to which the infants were actually attending? Further, the claim was that at 3-4 months of age infants were not making use of previously acquired perceptual information (i.e., prior categorical knowledge of dogs or cats or, for that matter, ears, noses, legs, etc.); rather, they were simply relying on statistical pattern recognition. Under these circumstances, using a set of measurements of specific high-level perceptual

features to characterize the input seemed, if not

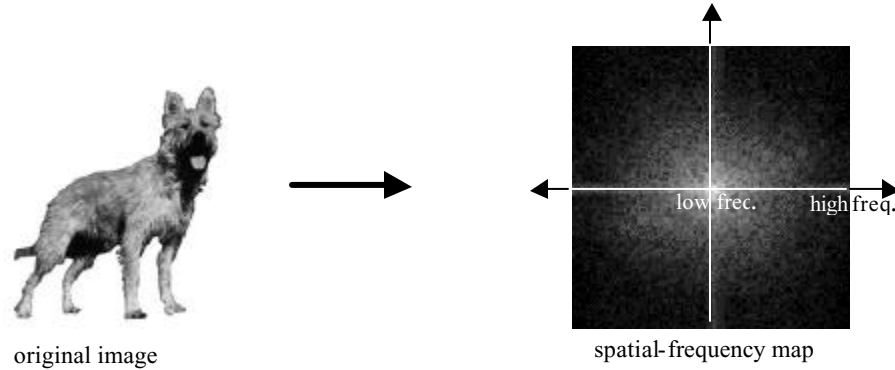


Figure 1: Transformation of the original image into a spatial-frequency map

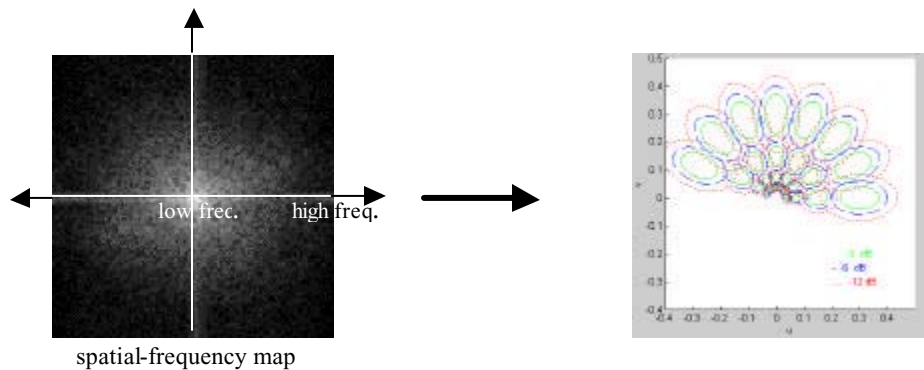


Figure 2. Once we have the map of spatial frequencies, we “cover” this map with spatial-frequency ovals along various orientations of the image. (Each of the ovals are normalized to have approximately the same energy.)

necessarily incorrect, at least somewhat inappropriate.

We therefore decided to attempt to examine this problem in a more neurobiologically plausible manner, one which sidestepped the difficulties inherent in selecting and measuring various perceptual features of the cat and dog stimuli. The dog/cat stimuli used in the simulations reported in this paper were those used in French et al. (2001), all of which had been normalized to have approximately the same size.

Organization of the present paper

We will attempt to answer two questions in the present paper.

The first is: Can we avoid the use of explicit feature coding in our autoencoder model of infant categorization and replace this coding with Gabor-filtered input known to have a neurophysiological counterpart in the infant visual system? We will show that this can, indeed, be done successfully.

The second issue that we will address starts from the well-known fact that the 34 month old infant visual system is not sensitive to high spatial frequency information (Banks & Salapatek, 1981; Dobson & Teller, 1978). However, instead of this being a disadvantage for the infant, we will show that, somewhat counter-intuitively, this low visual acuity is

actually an *advantage* in learning basic-level categories. The claim is that high spatial frequency information in the input signal produces an “information overload” in the infant cognitive system, adding information that is not necessary for correct categorization but that must, nonetheless, still be processed. In other words, when the infant is attempting to learn basic-level categories, high spatial frequency information in the input is very much like noise (Turkewitz & Kenny, 1982; Turkewitz & Kenny, 1985) and, as such, the less there is, the better.

Spatial frequency maps

It is well known that different columns in V1 are sensitive to different ranges of spatial frequencies (De Valois & De Valois, 1988; Tootell, Silverman, & De Valois, 1981). A scene reconstructed from only low spatial frequency information (i.e., with fine details blurred out) appears to us to be blurry. On the other hand, an image composed of high spatial frequencies would show *only* the fine details and would have no global perspective (rather like seeing many individual trees, but having no sense of the global entity, a forest). In any case, in order to have an optimal perception of a scene, we need the entire range of spatial frequencies. Therefore, by means of a 2D

Fourier transform, we first decomposed each of the images in the stimulus set into its component spatial frequencies and plotted them on a spatial frequency map (see Figure 1).

We then covered the frequency diagram with a “flower-petal” arrangement of 26 oval spatial frequency areas (“filters”) corresponding to various orientations emanating from the center of the spatial-frequency diagram (Figure 2). Gabor functions were used to simulate the 2D spatial and spectral structure of simple cells in visual primary cortex. (Jones & Palmer, 1987; Jones, Stepnowski, & Palmer, 1987). The smaller petals near the center of the map encompass the low frequencies, while the larger ovals further from the center group together high spatial frequencies. For each of these 26 filters, we calculate an “energy” value based on the local energy spectra, thereby simulating the activity of V1 complex cells (Sakai & Tanaka, 1999). This value determines the importance of that particular filter. If there are many spatial-frequency points that fall in a particular oval, it is given a high energy value; few points in a particular oval mean a low energy value.

Recall that in prior experiments and simulations (Mareschal et al., 2000; French et al., 2001), the dog/cat stimuli were characterized by a vector of ten values, with each value corresponding to a particular “high-level” feature. Now, instead of using ten features, we characterize each of the images by a vector of 26 values, each of which corresponds to the weighting of a group of spatial frequencies along various orientations of the image.

Visual acuity in infants

We know that the visual acuity in infants is not the same as that of adults (Banks & Salapatek, 1981; Dobson & Teller, 1978). In particular, infants do not perceive high-spatial frequencies (i.e., fine details), or perceive them only poorly. Certain authors (Turkewitz & Kenny, 1982; Turkewitz & Kenny, 1985) have claimed that, rather than being a problem, this reduced visual acuity may actually improve perceptual efficiency by eliminating the “information overload” caused by too many extraneous fine details likely to overwhelm their cognitive system. An implication is that basic-level category learning may be facilitated by reduced visual acuity.

In both of the simulations below we removed most of the high spatial frequencies from the input given to the autoencoder network that was used in Mareschal and French (1997), Mareschal et al. (2000), and French et al. (2001). This was done by weighting the contribution of each of the spatial frequencies according to a normal distribution (with the low spatial frequencies at the center) and cutting off all spatial frequencies above 7.1 cycles/degree. The spatial frequencies are Gaussian-filtered in such a way that spatial frequencies above 3-4 cycles/degree contribute very little to the input vector associated

with each image; the cut-off of 7.1 cycles/degree completely removes the highest spatial frequencies.

Overview of the simulations

The 26-16-26 autoencoder network used in the two simulations presented in this paper is based on a simple encode-compare-adjust principle (Sokolov, 1963; Charlesworth, 1969; Cohen, 1973): When an infant sees a perceptual stimulus, this stimulus is encoded as an internal representation, which is continually compared to the external stimulus and adjusted to match it. As long as there is a significantly large discrepancy between the internal representation and the external stimulus, the infant continues to look at the external stimulus. As this discrepancy falls, the infant becomes less interested in the external stimulus. In the autoencoder model, this is equivalent to the network’s correctly generalizing on output to match the network input (i.e., if the error on each of the 26 outputs is less than 0.5). In particular, we will use this criterion of generalization to measure the network performance on the category-learning task in Simulation 2.

In the simulations reported here we hope to establish two claims – namely:

- i) Simulation 1: that the use of a vector of 26 weighted spatial-frequency values, as described above, does, indeed, produce autencoder network results that are similar to those produced by infants tested on the same images and
- ii) Simulation 2: that the reduced visual acuity produced by largely eliminating high-spatial frequency information from the input (i.e., “blurry” vision) actually significantly *improves* the network’s ability to categorize the images presented to it.

Simulation 1: The adequacy of Gabor-filtered spatial-frequency input

In the first simulation we used the dog/cat stimulus set used in French et al. (2001). These authors used an encoding technique developed in Mareschal and French (1997) and Mareschal et al. (2000) in which 10 features of the animal images were measured and used as input to a 10-8-10 autoassociative network. Using feature-based input to this autencoder, we obtained categorization results that qualitatively matched experimental data with infants. In contrast, in the present simulation, we decomposed each image into a vector of values consisting of the energy values from each Gabor filter for a given orientation and spatial frequency. These values correspond, at least approximately, to what V1 neurons are known to “perceive.”

Each value of the 26-element vector represents an “energy” level associated with that particular spatial frequency. For this simulation, frequencies above 3-4 cycles per degree of visual arc are given a very low

energy value (very high frequencies, i.e., those above 7.1 cycles/degree are simply removed), which means that they contribute very little to the overall input vector (i.e., they contribute very little to the overall characterization of the image). The removal of this high spatial-frequency information was done to simulate the reduced visual acuity of 3-4 month old infants (Courage & Adams, 1995).

The simulation reported here was done on a standard 26-16-26 feedforward backpropagation autoencoder network (learning rate: 0.1, momentum: 0.9, Fahlman offset: 0.1). The stimulus set and the training regime was identical to that used in French et al. (2001). (It is important to recall that in French et al., 2001, the Dog stimuli were selected to be the less varied category, while the exemplars making up the Cat category were chosen to be considerably more varied than the dogs.)

Networks were trained in batches of 2 patterns for a maximum of 250 epochs. This simulated familiarization with pairs of pictures for a fixed period before being presented with a new familiarization pair. All results were averaged over 100 runs.

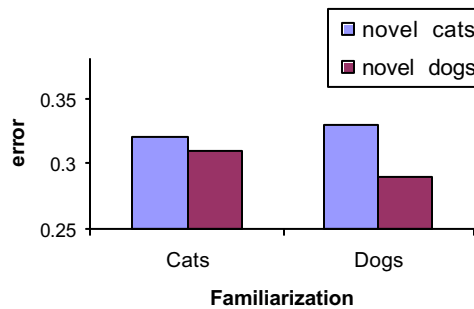


Figure 3a: Network generalization errors on novel cats/dog exemplars as a function of familiarization category.

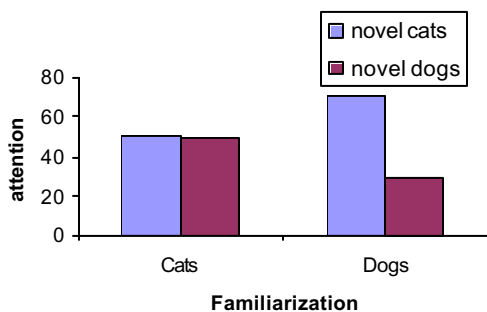


Figure 3b: Corresponding results for 34 month old infants

Figure 3a shows the model's generalization error to novel exemplars of cats and dogs as a function of whether they were trained on cats (the broad category) or on dogs (the narrow category). Networks trained (i.e., familiarized) with cats show very little difference in error (hence predict little difference in

infant looking times) when tested with a novel cat or a dog. In contrast, networks originally trained with dogs show significantly more error ($F(1, 198)=13.4$, $p<0.0005$) when tested with a novel cat than when tested with a novel dog (suggesting a preference for looking at a novel cat vs. a novel dog). Figure 3b shows the corresponding attentional asymmetry in 3-4 month old infants, as reported in French et al. (2001).

These simulation results using Gabor-filtered spatial frequency data allow us to conclude that the use of this type of spatial frequency data produces a reasonable fit to data. Most importantly, this result allows us to circumvent the thorny issue of using a particular set of "high-level" feature measurements (ear length, eye separation, etc.) to characterize the images used in the simulations.

Simulation 2. Improved categorization with reduced visual acuity

Does the autoencoder model of infant categorization (Mareschal & French, 1997; Mareschal et al., 2000) show improved categorization performance (at least on the dog/cat basic-level category images used in French et al., 2001) when "reduced acuity" input is used compared to "full acuity" input? The answer is that categorization performance is, indeed, enhanced, as we will show below.

To reiterate, the key idea of this simulation, which at first blush seems rather counter-intuitive: categorization performance for basic-level categories (Rosch et al., 1976) should be better *without* high spatial frequency information. This information is rather akin to noise in the input since, while it does indeed add information to the signal, it is not needed for accurate basic-level categorization. This extraneous information thus makes it more difficult (for the infant or for the network) to make use of the lower spatial frequency information that is, in fact, essential to basic-level categorization.

We used the same network as in Simulation 1, with an identical parameter set. We first ran the network (100 independent runs) with input data that contained all of the spatial-frequency information in the images. We then ran the network again (100 independent runs) with input data from which most of the high-spatial frequency information had been removed, as described above. The network was trained for 250 epochs on the training stimuli, as in Simulation 1.

As can be seen in Figure 4, whether the network was trained on Cats or Dogs, whether it was tested on novel dogs or novel cats, its categorization performance is significantly poorer when the input signal also contains high spatial frequency information compared to input with the high spatial frequencies removed.

It is also important to note that in the reduced visual acuity condition, we continue to see a significant difference in error (corresponding to

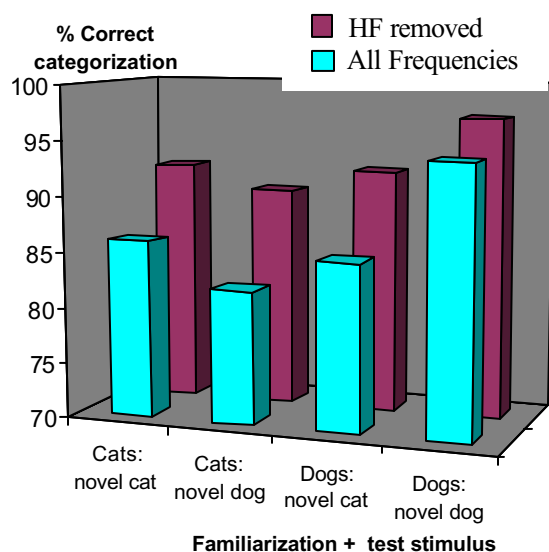


Figure 4. More information is not always better information, at least for basic-level categorization. The addition of high spatial frequency information makes correct basic-level categorization *more* difficult for the network.

attention in infants) when the network is trained first on dogs (in these experiments, the narrow category) and then sees a new cat, compared to when the network is first trained on cats (the broad category) and then sees a new dog.

Basic-level categories and incrementally increasing cognitive load

It is important to note that reduced acuity should improve categorization learning in the case of basic-level categories, but not subordinate-level categories. To see why this would be true we need to refer to Rosch et al.'s (1976) definition of "basic-level" categories. This level of categorization is the level for which the ratio of between-category variance to within-category variance is the highest. In other words, between-category variance is high with respect to within-category variance, which is generally relatively low. Within-category variance increases as fine-grained details of category exemplars increase. But these finer details are revealed only by the high spatial frequencies. For this reason, a decreased visual acuity that consists of partially or completely removing high-spatial frequency information, will decrease within-category variance and leave between-category variance largely unchanged. This would improve the learning of basic level categories, but would make it difficult, if not impossible, for 34 month old infants to learn categories that depend on high spatial frequency information. This applies, in particular, to subordinate-level categories.

Having *already learned* a certain number of basic-level categories under conditions of reduced visual acuity, when the high spatial frequency

apparatus does begin to come on-line at around 7 to 8 months of age (Kellman & Arterberry, 1998), the infants will be in a better position to then do more refined (i.e., subordinate-level) category learning. Thus, rather than having to confront all of the information associated with a particular category at once, the limitations of visual acuity of the infants' immature visual system first helps the infant to distinguish broader categories. Once these have been learned (or partially learned), then their visual/cognitive apparatus is then ready to build on this knowledge by incorporating the fine-grained details, perceived through high spatial frequency perception, that characterize subordinate expert-level categorization. The overall results of the simulations are thus consistent with a differentiation-driven view of early category development (Quinn & Johnson, 1997, 2000).

Furthermore, these results are entirely consistent with Archambault, Gosselin, & Schyns (2000), who showed that basic-level categorization seems to be more resistant to changes in viewing distances than that of subordinate-level categorization. This is because of the fact that as an object recedes from the viewer, information about details (i.e., high spatial frequency information) is lost, whereas low-spatial frequency information is not. Since basic-level categorization is largely based on the latter, we would expect more resistance to change of this type of categorization compared to subordinate-level categorization, where features are, indeed, essential.

A Prediction of the Model

A simple prediction emerges from these results. By manipulating the amount of high-frequency information in test images, it should be possible to vary infants' responses to these items after familiarization on a standard set of basic-level categories. So, for example, consider the Dog/Cat stimuli from the experiment by Quinn et al. (1993), in which the Dog category largely subsumes the Cat category. Under normal circumstances when infants are familiarized with cats, then shown a novel dog and a novel cat, they devote significantly more attention to the novel dog than to the cat. But were we to choose a novel dog and a novel cat whose differences were based largely on high spatial frequency information, we would expect the previously observed novelty preference to disappear, even if for us, adults, the two animals were quite different, one clearly being a dog, the other, clearly a cat.

Conclusion

In an extension of work done by Mareschal & French (1997), Mareschal et al. (2000) and French et al. (2001), we have been able to show that there is no need to use feature-based characterizations of the stimuli presented to the encoder network. Autoencoder results using Gabor-filtered input

corresponding approximately to the set of frequencies that the human visual system is known to use also produce a good approximation to categorization results in infants. We have also modeled a rather counter-intuitive learning advantage for basic-level categories that arises from reduced acuity input. Finally, based on the results of our autoencoder model of infant categorization and on the results we obtained using reduced acuity input, we have suggested experiments that might be performed on infants to further examine the validity of this model.

Acknowledgments

This work was supported in part by Grant HPRN-CT-2000-00065 from the European Commission to R. French, ESRC (UK) Grant R000239112 to D. Mareschal, and by Grant BCS0093600 from the National Science Foundation to P. C. Quinn.

References

- Archambault, A., Gosselin, F., & Schyns, P. (2000). A natural bias for the basic level? *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, NJ:LEA, 585-590.
- Banks, M.S., & Salapatek, P. (1981). Infant pattern vision: A new approach based on the contrast sensitivity function. *Journal of Experimental Child Psychology*, 31, 1-45.
- Charlesworth, W. R. (1969). The role of surprise in cognitive development. In D. Elkind & J. Flavell (Eds.), *Studies in cognitive development. Essays in honor of Jean Piaget*, pp. 257-314, Oxford, UK: Oxford University Press.
- Cohen, L. B. (1973). A two-process model of infant visual attention. *Merrill-Palmer Quarterly*, 19, 157-180.
- Courage M.L., Adams R.J. (1995). Infant peripheral vision: the development of monocular visual acuity in the first 3 months of postnatal life. *Vision research*, 36, 1207-1215.
- De Valois, R.L., De Valois K.K. (1988). *Spatial Vision*. Oxford University Press. New York.
- Dobson, V., & Teller, D. Y. (1978). Visual acuity in human infants: A review and comparison of behavioral and electrophysiological studies. *Vision Research*, 18, 1469-1483.
- French R. M., Mermillod M., Quinn P. C. & Mareschal D. (2001). Reversing category exclusivities in infant perceptual categorization: simulations and data. *Proceedings of the 23th Annual Cognitive Science Society Conference*, LEA, 307-312.
- Jones, J.P. & Palmer L.A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J Neurophysiol.* 58(6), 1187-211.
- Jones, J.P., Stepnoski A. & Palmer L.A. (1987). The two-dimensional spectral structure of simple receptive fields in cat striate cortex. *J Neurophysiol.* 58(6): p. 1212-32.
- Kellman, P. J., & Arterberry, M. E. (1998). *The cradle of knowledge: Development of perception in infancy*. Cambridge, MA: MIT Press.
- Mareschal, D., & French, R. M. (1997). A connectionist account of interference effects in early infant memory and categorization. In *Proceedings of the nineteenth annual conference of the Cognitive Science Society* (pp. 484-489). London: Erlbaum.
- Mareschal, D., French, R., & Quinn, P. (2000). A connectionist account of asymmetric category learning in early infancy. *Developmental Psychology*, 36, 635-645.
- Quinn, P. C. & Johnson (2000). Global before basic object categorization in connectionist networks and 2 month-old infants. *Infancy*, 1(1), 31-46.
- Quinn, P. C., & Eimas, P. D. (1996). Perceptual cues that permit categorical differentiation of animal species by infants. *Journal of Experimental Child Psychology*, 63, 189-211.
- Quinn, P. C., & Johnson, M. H. (1997). The emergence of perceptual category representations in young infants: A connectionist analysis. *Journal of Experimental Child Psychology*, 66, 236-263.
- Quinn, P., Eimas, P., & Rosenkrantz, S. (1993). Evidence for representations of perceptually similar natural categories by 3 and 4-month-old infants. *Perception*, 22, 463-475.
- Rosch E., Mervis C.B., Gray D.G., Johnson D.M. & Boyes-Braem P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-439.
- Sakai K. & Tanaka S. (1999). Spatial pooling in the second-order spatial structure of cortical complex cells. *Vision Research*, 40, 855-871.
- Solokov, E. N. (1963). *Perception and the conditioned reflex*. Hillsdale, NJ: LEA.
- Spencer, J., Quinn, P. C., Johnson, M. H., & Karmiloff-Smith, A. (1997). Heads you win, tails you lose: Evidence for young infants categorizing mammals by head and facial attributes (Special Issue: Perceptual Development). *Early Development and Parenting*, 6, 113-126.
- Tootell, R. B., Silverman, M. S., De Valois, R. L. (1981). Spatial frequency columns in primary visual cortex, *Science*, 214, 813-815.
- Turkewitz G., Kenny P. A. (1982). Limitations on input as a basis for neural organization and perceptual development: a preliminary theoretical statement *Developmental Psychobiology*, 15(4), 357-368.
- Turkewitz G., Kenny P. A. (1985). The role of developmental limitations of sensory input on sensory/perceptual organization. *Journal of Developmental and Behavioral Pediatrics: JDBP*, 6(5), 302-306.