

Development of configural 3D object recognition

Ingo Rentschler^{a*}, Martin Jüttner^b, Erol Osman^a, Alexander Müller^a, Terry Caelli^c

^a *Institute of Medical Psychology, University of Munich, Munich, Germany*

^b *Neuroscience Research Institute, Aston University, Birmingham, U.K.*

^c *Department of Computing Science, University of Alberta, Edmonton/Alberta, Canada*

Behavioural Brain Research, accepted for publication June 3rd, 2003

Manuscript content: 14 pages (including 2 figures)

Reference for citations:

Rentschler, I., Jüttner, M., Osman, E., Müller, A. & Caelli, T. (2004).
Development of configural 3D object recognition. *Behavioural Brain Research*
149, 107-111.

*Corresponding author: Goethestrasse 31, 80336 München, Germany; Tel. +49 (0) 89 5996
641; fax +49 (0) 89 7499 6873; e-mail address: ingo@imp.med.uni-muenchen.de

Authors' full names and complete addresses:

Ingo Rentschler, Erol Osman, Alexander Müller, Institute of Medical Psychology, University of Munich, Goethestrasse 31, 80336 München, Germany (telephone, fax, and e-mail on title page)

Martin Jüttner, Neuroscience Research Institute, School of Life & Health Sciences – Psychology, Aston University, Aston Triangle, Birmingham B4 7ET, U.K.

Terry Caelli, Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada T6G 2H1

Abstract

There is evidence for the late development in humans of configural face and animal recognition. We show that the recognition of artificial three-dimensional objects from part configurations develops similarly late. We also demonstrate that the cross-modal integration of object information reinforces the development of configural recognition more than the intra-modal integration does. Multimodal object representations in the brain may therefore play a role in configural object recognition.

Key words: object recognition, representation, configural recognition, mirror symmetry, category learning, cross-modal, multimodal, cognitive development

For retrieving the spatial structure of three-dimensional (3D) objects from single static two-dimensional (2D) views humans require additional information. It has been suggested that this is accomplished by relating image features to visual representations in the brain [2, 4]. Yet object knowledge retains an intrinsic multimodal quality as it may be acquired both by vision and by touch. The haptic modality can be assumed to yield object information as rich as that provided by the visual modality [15], and haptics and vision tend to complement one another in the type of information used to represent the object [17, 24]. This indicates that 3D object recognition may benefit from the multimodal integration of sensory information.

Evidence for the existence of multimodal object representations in the brain comes from neurophysiological research including functional magnetic resonance imaging (fMRI) and transcranial magnetic stimulation (TMS) techniques. TMS experiments suggest that the visual cortex is also closely involved in the tactile discrimination of object orientation and shape [33]. fMRI studies probing cross-modal priming in object recognition have revealed an activation not only in the somatosensory cortex but also in regions of the occipital cortex traditionally associated with visual processing, thus pointing at overlapping haptic and visual representations [1, 14].

From developmental psychology it would seem that multimodal object recognition occurs already in infants [20]. The finding however that infants can use a variety of cues to perceive objects [22, 31] does not answer the questions as to when and “how these cues (i.e., knowledge sources) are combined in the process of constructing coherent object representations” [ref. 27, p. 32]. Experiments on visual animal recognition [10] indicate that recognition abilities in young children (under the age of 10 years) draw much upon properties of isolated object parts, whereas comparable performance levels with regard to the processing of part configurations, or global form, are reached only in adolescence (around 15-16 years).

A similar development from the processing of parts to that of part configurations has been postulated for face recognition [5, 6, 8].

Taken together these observations suggest that the *configural* recognition of artificial 3D objects may occur as late as face and animal recognition do, and that multimodal object representations play a role in its development. To test this hypothesis we compared the developmental characteristics of the effects of haptic and visual prior knowledge on visually learning 3D objects differing in part configurations only. For achieving this goal, we employed a psychophysical paradigm based on two essentials. First, a learning set of unfamiliar complex objects containing mirror-symmetric (“left-right”) objects was used. Such objects cannot be identified on the basis of “diagnostic” features since they differ in higher-level configural properties only [28]. Second, a fixed procedure of visual category learning was used. Thus only characteristics of internalized object representations and not of (physical) stimulus redundancies could affect learning performance [25].

We studied object learning in 30 school children aged 8-14 years and a control group of 10 adults. The experiments employed a set of three artificial objects that were novel for the subjects. Thus learning was completely under experimental control, and confounding effects due to naming and conceptual world knowledge were minimized. The objects were each composed of four spheres. Three spheres formed an isosceles triangle, while the fourth was placed perpendicularly above the centre of one of the base spheres. Thus one unique object 1 and two left-right objects 2 and 3 were obtained (Figure 1A, left). The objects were generated both as physical models and as virtual models. Physical models were constructed of polystyrene balls measuring 6 cm in diameter. Virtual models were generated and displayed as perspective 2D projections on a computer graphics system (O2 workstation, Silicon Graphics Inc., USA).

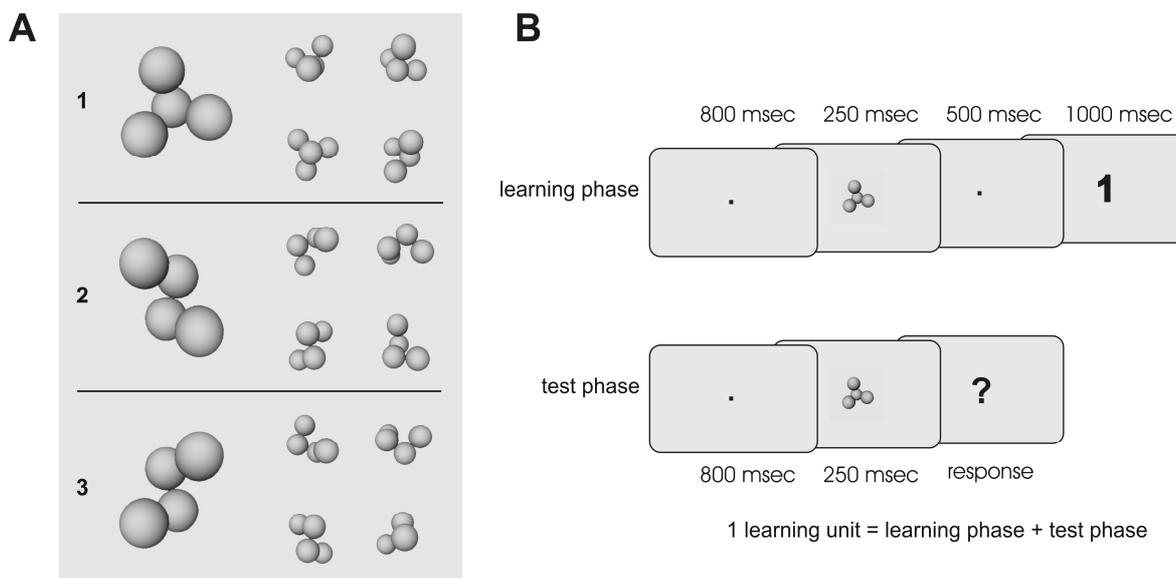


Figure 1. Learning the categorization of three 3D objects from 2D views. **(A)** Left: Objects consisted of four spheres, with three of them forming an isosceles triangle and the fourth being placed perpendicularly above the centre of one of the base spheres. Objects 2 and 3 were left-right versions of each other. Objects were generated both as physical and as computer graphic models. Right: Samples of the 22 views used for category learning. **(B)** Supervised category learning. The learning procedure was partitioned into “learning units” consisting of a training phase and a test phase.

For the acquisition of *haptic* prior knowledge the blindfolded subjects explored with their hands the physical models and compared them part-by-part. They were encouraged to do so by moving and rotating the objects in space without restrictions. For gathering *visual* prior knowledge subjects could “grasp” the virtual models on the computer screen by means of the computer mouse and deliberately change their viewing angle. Thus they could inspect the objects from any desired viewing direction. For both conditions, the pre-exploration lasted 5 min and was immediately followed by visual category learning.

During category learning the subjects were trained in a supervised-learning paradigm [29] to classify a set of 22 views of the previously explored objects. The views were obtained from the virtual object models by sampling the viewing sphere in 60° steps (Figure 1A, right)

[25]. The learning procedure was partitioned into a series of learning units, each consisting of a learning phase and a test phase (Figure 1B). During the learning phase each view of the learning set was presented once for 250 ms and in random order, followed by the corresponding object label displayed for 1s. During testing, each view was presented once and then assigned by the observer to one of the three objects. Each learning unit thus resulted in a (22 x 3) - dimensional classification matrix. The learning procedure lasted one hour within which each subject completed on average 13 learning units.

Learning performance is summarised in Figure 2A in terms of a bias-free sensitivity measure d' [13] derived from the classification frequencies. Observer performance is shown in the initial and late phase of learning (learning units $N = 1$, and $N = 13$, respectively). The data are split into a row indexing the classification of the “non-symmetric” object 1 and a row indexing the mean classification of the “symmetric” objects 2 and 3. They show that children of grade 8 (13-14 years), and the adults of the control group, with haptic prior knowledge performed better than those with visual prior knowledge. This trend was reversed for subjects of grade 3 (8-9 years), thus leading to a significant interaction between the factors age and modality used during the pre-exploration both in initial and in late learning ($N= 1$: $F(3,104) = 3.87$, $p < 0.05$; $N = 13$: $F(3,104) = 4.80$, $p < 0.01$). Post-hoc comparisons revealed that children of grade 8 performed significantly better in the haptic relative to the visual condition both in initial and late learning of the non-symmetric object ($N=1$: $t(8) = 2.38$, $p < 0.05$; $N = 13$: $t(8) = 3.38$, $p < 0.05$) and in late learning for the two symmetric objects ($t(18) = 2.11$, $p < 0.05$). In contrast, children of grade 3 performed significantly better in the visual relative to the haptic condition for the non-symmetric object ($N = 1$: $t(8) = 4.17$, $p < 0.01$; $N=13$: $t(8) = 2.45$, $p < 0.05$) and in late learning ($N = 13$) of the two symmetric objects [$t(18) = 4.06$, $p <$

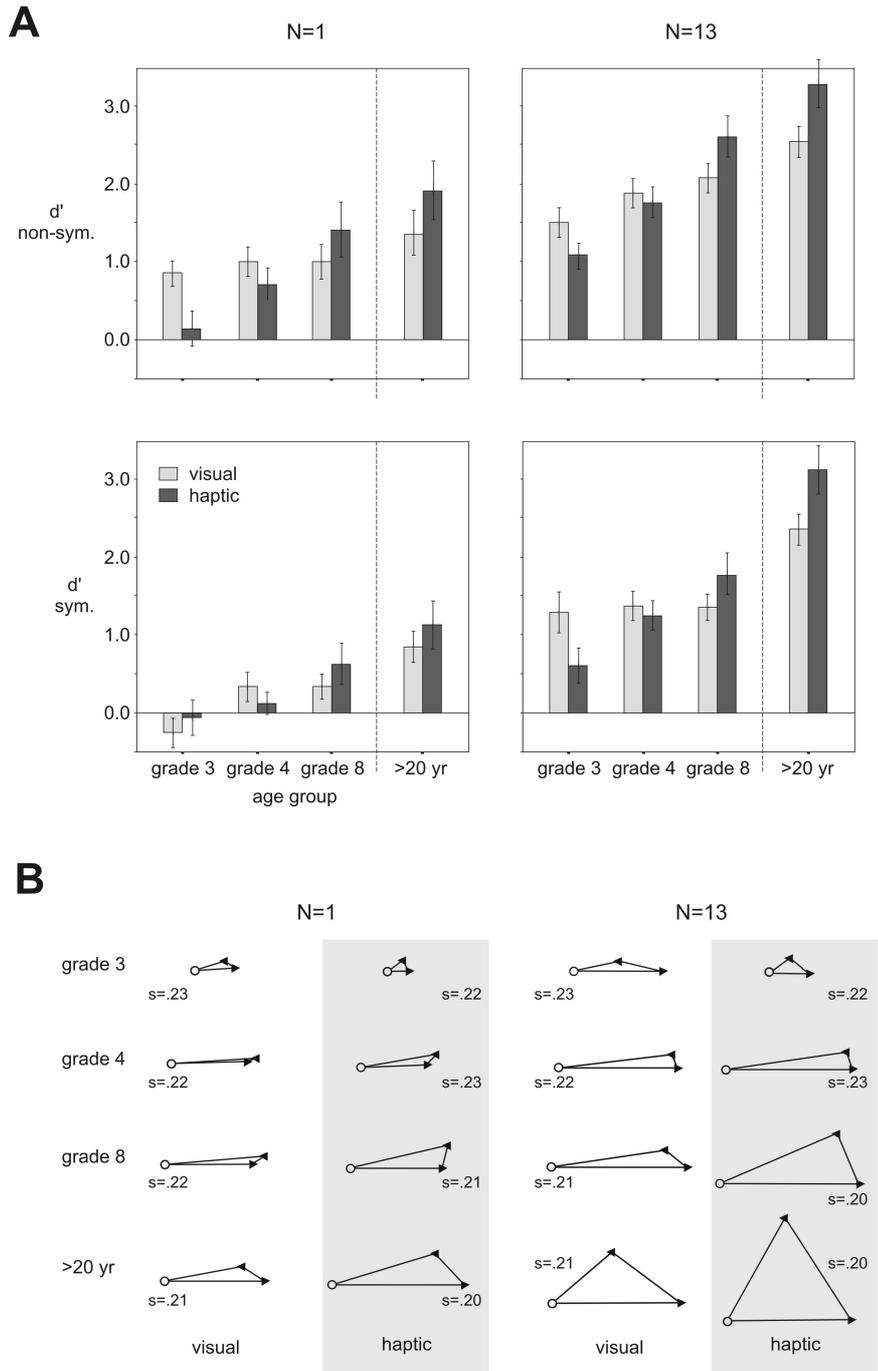


Figure 2. Effect of prior knowledge on learning 3D objects from 2D views. Dynamics of visual learning as reflected by performance in learning units N=1 and N=13. **(A)** Sensitivity measures d' for recognizing the non-symmetric object 1 and mean d' for the symmetric objects 2 and 3. Light bars: Prior knowledge from visually inspecting on the computer screen the 2D projections of virtual 3D object models that were rotated via the mouse. Dark bars: Prior knowledge from blindfolded subjects exploring the real 3D objects with their hands. School children in grade 3 (age 8-9 years), grade 4 (age 9-10 years), and grade 8 (age 13-14 years). Reference group of adults aged 20 – 45 years. 5 subjects at each condition; error bars S.E. (N=5). **(B)** Conceptual space representations derived by Multidimensional Scaling (MDS). Vertices of triangles represent object concepts (\circ : object 1; \blacktriangleright : object 2; \blacktriangleleft : object 3). Distances between vertices reflect perceived object similarities; variable s denotes stress values of MDS solutions.

0.01)]. Generally, the children performed significantly worse for the symmetric objects as compared to the non-symmetric object ($N = 1$: $F(1,104) = 18.63$, $p < 0.01$; $N = 13$: $F(1,104) = 24.46$, $p < 0.001$). However, the different levels of complexity induced by object symmetry did not affect the relationship of age and modality as the interaction age \times modality \times symmetry was non-significant ($N = 1$: $F(3,104) = 0.71$, $p = 0.51$; $N = 13$: $F(3,104) = 0.49$, $p = 0.65$).

The learning data in form of cumulated classification matrices were further analysed in terms of Multidimensional Scaling (MDS). For each pair of views a similarity measure was obtained from the classification frequencies. The resulting similarity matrices for the visual and haptic condition were conjointly subjected to weighed MDS. By combining views from same objects, the conceptual space could be visualized by triangular configurations with the distance of pairs of vertices reflecting the perceived similarity of class concepts. The configurations are shown in Figure 2B together with their stress values, which provide a measure of fit for the scaling solutions. The graphs reveal the pronounced distortions in the conceptual space introduced by the divergence between the two symmetric objects and the non-symmetric object. The distortion becomes particularly prominent in the conceptual space of children of grade 4 (age 9-10 years). For children of grade 8 (age 13-14 years), haptic prior knowledge enables a more veridical representation of the conceptual space than does visual prior knowledge, eventually leading to a perfectly balanced representation in the adults.

The MDS analysis of the learning data yielded two main results: First, at the complexity level of mirror-symmetries configural recognition did not develop until late childhood and adolescence. Second, cross-modal integration of object knowledge reinforced the development of configural 3D object recognition more than intra-modal integration did. This raises the question of whether the tardy development of cross-modal integration is due to a

retardation of (1) efficient haptic feature extraction, (2) the ability to build haptic object representations, or that of (3) the cross-modal binding of object information. As the development of tactual object exploration and recognition settles by the age of 5 - 6 years [21], the first two hypotheses are unlikely candidates for interpreting our findings. This conclusion is supported by our observation that there was no apparent age-dependent difference in haptic exploration behaviour between age groups. We therefore suggest that our results reflect the late development of the cross-modal binding of configural 3D object information.

Lederman and Klatzky [17] suggested that adults identify form or global structural information early in the recognition process, whereas they discern object parts later. The age-related observations reported here would seem to contradict their claim. It should be noted however that, in the terminology of these authors, differences of “form” existed in our study between object 1 and each of the left-right objects 2 and 3. These differences were due to the bilateral symmetry that object 1 has in itself and that is lacking in objects 2 and 3. By contrast, objects 2 and 3 had the same symmetry properties in themselves but are mirror-symmetric versions of each other. The difference between objects 2 and 3 therefore reflected a higher level of configural complexity, or greater relational depth of part attributes [25], than did the differences between the latter objects and object 1. Hence the conclusion that the developmental trend we observed reflects age-related changes in the object-representation strategies proposed by Lederman and Klatzky [17].

More obvious is the correspondence between the late development of configural 3D object recognition observed here and the tardy development of image-based configural face [5, 6, 8] and animal recognition [10]. Whether there exists a causal relationship between these two types of developmental phenomena cannot be decided from the present study. Yet it is

tempting to speculate that the functional maturation of multimodal 3D object representations plays a role not only in configural 3D object recognition but in configural image processing in general.

The cross-modal reinforcement of configural 3D object recognition reported here is consistent with earlier concepts of active perception and cognition according to which a separation between cognition and action is difficult [23, 26, 32]. Two ways of interpreting related visuomotor interactions are conceivable. First, lesions in the left or right parietal lobes may cause constructional apraxia [3, 16]. This deficit has been attributed to the disconnection of the visual image of objects from the mental programs guiding a patient's movement as he/she constructs these objects from parts [16]. More specifically, a lesion of the left angular gyrus, a region within the inferior parietal lobule, may cause Gerstmann's syndrome [9, 19]. Finger agnosia, left-right confusion, agraphia and acalculia are the constituents of the latter deficit the common denominator of which may be an impairment in the manipulation of mental images [19]. Second, there is evidence that cross-temporal contingencies of new and complex sensorimotor sequences are integrated by connecting sensory areas with motor areas of the cerebral cortex under the control of the prefrontal cortex (PFC; ref. 12, p. 329). Both the parietal and the prefrontal types of function might subsume task demands as in the present study, where subjects haptically explored complex 3D objects by moving them freely in space and grasping them part-by-part. It is further of interest that the angular gyrus is among the last territories of the neocortex to develop [11] and so is the prefrontal cortex [7]. Thus it seems impossible to decide about these interpretations on the basis of the present study. It would be of interest however to see whether the observed developmental delay in cross-modal integration occurs for familiar stimulus material as well. If this were not the case it would seem to be more likely that the function of the prefrontal cortex underlies the present findings.

However this may be an apparent inconsistency in functional localization should not pass unnoticed. The brain imaging studies [1, 14] identified structures in the occipito-temporal cortex as the site of multimodal object representations, whereas we argued in favour of either parietal or prefrontal functions. We feel that this discrepancy in localization could be due to the fact that the former studies required their subjects to touch immobile objects, whereas our subjects explored the learning objects by manipulating them freely in space.

Conclusions: Configural 3D object recognition develops late and is not fully functional until adolescence. It is more strongly reinforced by the cross-modal integration of prior object knowledge as compared to the intra-modal integration of such knowledge. This suggests that multimodal object representations play a role in recognizing the configural characteristics of 3D objects. The integrative function of either the parietal or the frontal association areas of the cerebral cortex may underlie the formation of such representations.

Acknowledgement

We have benefited from the comments of an anonymous reviewer and discussions with Sabine Gross, Christoph Zetsche, and Mario Ferraro. Ms. Iona Sturmfels and Ms. S. Zeiler-Haas helped with conducting the experiments. This work was supported by the Stiftung für Bildung und Behindertenförderung, Stuttgart, Germany, and by Deutsche Forschungsgemeinschaft (DFG).

References

- [1] Amedi A, Malach R, Hendler T, Peled S, and Zohary E. Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neurosci*, 2001; 4: 324-330.
- [2] Biederman I. Recognition by components: A theory of human image understanding. *Psychol Rev* 1987; 4: 115-147.
- [3] Brown JW. *Aphasia, apraxia and agnosia*. Springfield, Ill.: Charles C. Thomas, 1972.
- [4] Bülthoff HH, and Edelman S. Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc Natl Acad Sci USA*, 1992; 89: 60-64.

- [5] Carey S, and Diamond R. Are faces perceived as configurations more by adults than by children? *Visual Cog*, 1994; 1: 253-274.
- [6] Carey S, Diamond R, and Woods B. Development of face recognition: A maturational component. *Develop Psychol*, 1980; 16: 257-269.
- [7] Casey BJ, Giedd JN, and Thomas KM. Structural and functional brain development and its relation to cognitive development. *Biol Psychol*, 2000; 54: 241-257.
- [8] Chung MS, and Thomson DM. Development of face recognition. *Br J Psychol*, 1995; 86: 55-87.
- [9] Critchley M. *The parietal lobes*. London: Edward Arnold & Co., 1953.
- [10] Davidoff J, and Roberson D. Development of animal recognition: A difference between parts and wholes. *J Exper Child Psychol*, 200; 281: 217-234.
- [11] Flechsig P. *Meine myelogenetische Hirnlehre. Mit biographischer Einleitung*. Berlin: Springer, 1927.
- [12] Fuster JM. The prefrontal cortex – an update: time is of the essence. *Neuron*, 2001; 30: 319-333.
- [13] Green DG and Swets JA. *Signal Detection Theory and Psychophysics*. Huntington: Krieger, 1974.
- [14] James TW, Humphrey GK, Gati JS, Servos P, Menon RS, and Goodale MA. Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia*, 2002; 40: 1706-1714.
- [15] Klatzky RL, and Lederman SJ. The intelligent hand. In: Bower GH, editor. *The Psychology of Learning and Motivation*, vol 21. San Diego: Academic Press, 1987, 121-151.
- [16] Kleist K. *Gehirnpathologie*. Leipzig: O. W. Barth, 1934
- [17] Lederman SJ, and Klatzky RL. Hand movements: A window into haptic object recognition. *Cognitive Psychol*, 1987; 19: 342-368.
- [18] Lederman SJ, and Klatzky RL. Cognitive salience of haptic object properties: role of modality-encoding bias. *Perception*, 1996; 25: 983-998.
- [19] Mayer E, Martory M.-D., Pegna AJ, Landis, T, Delaville J, and Annoni J.-M. A pure case of Gerstmann syndrome with a subangular lesion. *Brain*, 1999; 122: 1107-1120.
- [20] Meltzoff AN, and Borton RW. Intermodal matching by human neonates. *Nature*, 1979; 282: 403 – 404.
- [21] Morrongiello BA, Humphrey GK, Timney B, Choi J, and Rocca PT. Tactual object exploration and recognition in blind and sighted children. *Perception*, 1994; 23: 833 – 848.
- [22] Needham A. Object recognition and object segregation in 4.5-month-old infants. *J Exper Child Psychol*, 2001; 78: 3-24.
- [23] Neisser U. *Cognition and reality. Principles and implications of cognitive psychology*. San Francisco: W. H. Freeman and Co., 1976.
- [24] Newell FN, Ernst MO, Tjan BS, and Bühlhoff HH. View dependence in visual and haptic object recognition. *Psychol. Science*, 2001; 12: 37-42.
- [25] Osman E, Pearce AR, Jüttner M, and Rentschler I. Reconstructing mental object representations: A machine vision approach to human visual recognition. *Spatial Vision*, 2000; 13: 277-286.
- [26] Piaget J, and Inhelder B. *L'image mentale chez l'enfant*. Paris: Presses Universitaires de France, 1966.

- [27] Quinn PC, and Bhatt RS. Object recognition and object segregation in infancy: Historical perspective, theoretical significance, “kinds” of knowledge, and relation to object categorization. *J Exper Child Psychol*, 2001; 78: 25-34.
- [28] Rentschler I. Symmetry-coded cells in visual cortex? *Nature*, 1985; 317: 581-582.
- [29] Rentschler I, Jüttner M, and Caelli T. Probabilistic analysis of human supervised learning and classification. *Vision Res*, 1994; 34: 669-687.
- [30] Sowell ER, Delis D, Stiles J, and Jernigan TL. Improved memory functioning and frontal lobe maturation between childhood and adolescence: A structural MRI study. *J Intern Neuropsychol Soc*, 2001; 7: 312-322.
- [31] Spelke ES. Perceptual knowledge of objects in infancy. In Mehler J, Garrett M, and Walker E, editors. *Perspectives on mental representation*. Hillsdale NJ: Erlbaum, 1982, 409-430.
- [32] von Weizsäcker V. *Der Gestaltkreis*. Berlin: Springer, 1940.
- [33] Zangaladze A, Epstein CM, Grafton ST, and Sathian K, Involvement of visual cortex in tactile discrimination of orientation. *Nature*, 1999; 401: 587-590.