A developmental dissociation of view-dependent and view-invariant object recognition in adolescence

Martin Jüttner[a*], Alexander Müller[b], Ingo Rentschler[b]

Reference for citations:

Jüttner, M., Müller A. & Rentschler, I. (2006). A developmental dissociation of view-dependent and view-invariant object recognition in adolescence. *Behavioural Brain Research* 175, 420-424.

[a]*School of Life and Health Sciences - Psychology, Aston University, Aston Triangle, Birmingham B4 7ET, UK*

[b]*Institut für Medizinische Psychologie, Universität München, Goethestr. 31, 80336 München, Germany*

[*]Corresponding author. Tel. +44 121 2044065; fax +44 121 3593257.

E-mail address: m.juttner@aston.ac.uk

Abstract

Spatial generalization skills in school children aged 8-16 were studied with regard to unfamiliar objects that had been previously learned in a cross-modal priming and learning paradigm. We observed a developmental dissociation with younger children recognizing objects only from previously learnt perspectives whereas older children generalized acquired object knowledge to new viewpoints as well. Haptic and - to a lesser extent - visual priming improved spatial generalization in all but the youngest children. The data supports the idea of dissociable, view-dependent and view-invariant object representations with different developmental trajectories that are subject to modulatory effects of priming. Late-developing areas in the parietal or the prefrontal cortex may account for the retarded onset of view-invariant object recognition.

Keywords: Object recognition; Representation; Development; Adolescence; Learning; Viewpoint invariance; Cross-modal; Priming

Theories of human object recognition generally assume that image features extracted from single, static two-dimensional (2D) views are related to some form of internalised object representation. Concerning the quality of such representations there are two dominant views. On the one hand, it has been postulated that objects are mentally represented by three-dimensional (3D) configural descriptions in terms of parts or components (e.g. [1,2,3]). This class of theories generally predicts spatial generalisation, i.e. observers should be able to recognize objects, which have been previously learnt from a limited number of viewpoints only, also from novel, i.e. previously unseen perspectives, as long as the objects can be decomposed into a unique configuration of parts that do not occlude each other. On the other hand, a wide range of so-called image-based models have been proposed that generally assume that 3D objects are represented in terms of multiple, viewer centred 2D views (e.g. [4,5,6]). These theories generally predict a pronounced viewpoint dependency of object recognition unless the number of views is sufficiently large to enable the visual system to interpolate between them.

While both types of representations often have been discussed as mutually exclusive alternatives, more recent evidence from behavioural [7,8,9], neuropsychological [10,11,12,13] and neuroimaging [14] studies suggests that the two formats might co-exist in the brain, and that the visual system draws upon them in a task-dependent manner. Multiple representations are also plausible given the fact that knowledge of object geometry has an intrinsic multimodal sensory quality as it may be acquired both by vision and touch [15,16,17,18]. Experiments on cross-modal object learning show that even a short haptic exploration of objects (haptic priming) has long-lasting facilitating effects on subsequent visual learning that are stronger than in condition of a purely visual exploration (visual priming) [19]. This facilitation is particularly pronounced for objects with complex part structure, suggesting that haptic exploration might generate substantial information on object shape in a format that is readily accessible by the visual modality.

We have shown in previous work that the ability of a cross-modal transfer of configural information in object recognition does not develop before late adolescence, i.e. at an age of 13-14 years [20]. The present study aimed to further substantiate the link between the development of cross-modal learning and the representations underlying object recognition. For this purpose we tested spatial generalization skills in 60 school children aged 8-16 with regard to unfamiliar, artificial objects that had been previously learned in a cross-modal priming and learning paradigm. It was hypothesised that effects of age and priming on the recruitment of multiple object representations should manifest themselves in different degrees of recognition invariance with regard to novel object views that had not been encountered during the learning phase of the experiment.

A stimulus set of three molecule-like learning objects (Fig. 1) was employed. The unfamiliarity of the stimuli ensured that the learning process was completely under experimental control and minimized confounding effects due to naming and conceptual world knowledge. Each object was composed of four spheres. Three spheres formed an isosceles triangle, while the fourth was placed perpendicular above the centre of one of the base spheres (Fig. 1A, top). Because of this construction principle the three objects only differed in the spatial configuration of their constituent parts, resulting in two types of objects: a non-symmetric object (object 1) and two objects that were mirror-symmetric to each other (object 2 and object 3). All objects were generated both as physical models and as virtual models. Physical models were constructed from polystyrene balls each measuring 6 cm in diameter. Virtual models were generated by means of a 3D graphics software package (Open Inventor, Silicon Graphics Inc., USA) and displayed as perspective 2D projections on a SGI O2 workstation (Silicon Graphics Inc., USA). From the virtual object models two sets of views were generated. The first set, used to train the participants, was obtained by sampling the viewing sphere in 60 deg steps (Fig. 1B). Views redundant due to object symmetry were eliminated, resulting in 22 views in total (6 views for object 1, 8 views for both object 2 and object 3; cf. Fig. 1A, bottom). The second set of views, used to test spatial generalization, was obtained by sampling the viewing sphere in an analogous manner in 30 deg steps, resulting in a further 83 views (21 for object 1, 31 for both object 2 and object 3). Because of the sampling interval, 19 of the 83 test views were old views (i.e., they were identical to those used for training), whereas the others were new views for the subjects. At the viewing distance of 1 m the object views subtended a visual angle of 1.5 deg.

The experiment consisted of three parts: priming phase, supervised visual learning and generalisation test. During the priming phase the subjects either explored the physical object with their hands being blindfolded (*haptic priming* condition), or they explored the virtual models visually by rotating them successively around the three axis on the computer screen (*visual priming* condition). No instruction other than the invitation to familiarize themselves with the objects was given in either condition. Both haptic and visual priming lasted for 5 minutes and were immediately followed by the second part of the experiment, the supervised learning. As a baseline for the two priming conditions a neutral (*control*) condition was included, where subjects entered the supervised learning directly, i.e. without prior exploration of the objects.

The supervised-learning procedure was partitioned into learning units (see [20] for details), each consisting of a learning phase and a test phase. During the learning phase each of the 22 views of the learning set was presented once for 250ms and in random order, followed by the corresponding object label ("1", "2" or "3") displayed for 1 s. During the test phase, each learning view was presented once and assigned to one of the three objects by the observer. Upon completion of the test phase, participants received feedback concerning the percent correct value of their responses. Each subject performed a series of 12 learning units, which were completed within 50 minutes. For the learning unit with the maximum percent correct value in this series, the classification frequencies of each object were transformed into *d'* scores [21], in order to obtain a stable, bias-free sensitivity measure d'_{max} characterizing the learning status of the subject.

The final part of the experiment assessed spatial generalisation. Here each of the 83 test views was presented three times for 250 ms and in random order. Participants had to assign each view to one of the objects 1-3 by pressing the corresponding key of the computer keyboard. For old views and novel views, classification frequencies were separately derived for each object and transformed into *d'* scores.

Participants were assigned into four age groups (8-9 years, 10-11 years, 13-14 years, 15-16 years) and to one experimental condition defined by the type of priming (haptic, visual or control) used in the first part of the experiment. Each of the four age groups had equal number of participants in each condition.

A preliminary three-way ANOVA (object type x age group x priming condition) revealed, for both the data on learning and generalization, significant main effects of object type, age group and priming condition. The only significant interaction was between priming condition and age. To further explore this interaction, the data of the non-symmetric object 1 and the data of the two mirror-symmetric objects 2 and 3 were analysed in separate two-way ANOVAs (age group x priming condition) in the following.

Figure 2A displays learning performance in terms of the average

d'_{max} values in each priming condition and age range. The data are split into two columns according to object type, indexing the classification of the non-symmetric object 1 and the mean of the two (mirror-)symmetric objects 2 and 3, respectively. The data shows a distinct increase of learning performance with age (non-symmetric object: $F(3,48)=47.87$, $p<0.001$); symmetric objects: $F(3,108)=32.04$, $p<0.001$)) that is modulated by priming. In case of the non-symmetric object, the main effect of priming condition was close to significance ($F(2,48)=2,90$, $p=0.06$) and there was a significant interaction of priming condition and age (($F(6,48)=2.29$, $p<0.05$). For the symmetric objects the main effect of priming condition was significant ($F(2,108)=2.54$, $p<0.05$) whereas the interaction of priming condition and age was not ($F(6,108)=1.62$, $p=0.14$).

Inspection of Figure 2A suggests that, consistent with previous results [20], haptic priming facilitates learning of older children but not that of children in the youngest age group. An additional a posteriori contrast analysis showed a significant ($p<0.05$) advantage of haptic priming relative to the control condition in age groups 9-10 (non-symmetric and symmetric objects) and 15-16 (non-symmetric objects only), which turned into a significant disadvantage for children aged 8-9 (non-symmetric and symmetric objects).

Spatial generalization for the two types of objects is summarised in Figure 2B. For each priming condition and age group, shaded bars represent the mean classification scores of new views, whereas open bars refer to the classification of old views. Because recognition performance for old views and new views necessarily depends on learning status, i.e. the score achieved in the learning task preceding the generalisation test, an analysis of covariance (ANCOVA) was conducted using the d'_{max} scores obtained during learning (Fig. 1A) as covariate. Taking into account learning status in this way rendered the effect of age non-significant for the recognition scores of the non-symmetric object (old views: $F(3,47)=1.38$, $p=0.26$; new views: $F(3,47)=0.88$, $p=0.46$). In contrast, for the symmetric objects age significantly affected the recognition of new views ($F(3,107)=3.96$, $p<0.01$) but not of old views ($F(3,107)=2.03$, $p=0.12$). Priming effects were significant for the recognition of new views (non-symmetric object: $F(2,47)=4.27$, $p<0.05$; symmetric objects: $F(2,107)=9.92$, $p<0.001$); they were only approaching significance for the recognition of old views of the non-symmetric object ($F(2,47)=3.09$, $p=0.06$) and were non-significant for the symmetric objects ($F(2,107)=0.11$, $p=0.90$). Interactions of age and priming condition were non-significant for the non-symmetric object; for the symmetric objects they were significant for new views ($F(6,107)=3.21$, $p<0.01$) but not old views ($F(6,107)=0.18$; $p=0.84$).

These results indicate a differential effect of age and priming on the recognition of new as opposed to old views. For the recognition of new views, an additional a posteriori contrast analysis showed a significant ($p<0.05$) advantage of haptic priming relative to the control condition in age group 9-10 (non-symmetric objects only), 13-14 (non-symmetric and symmetric objects) and 15-16(non-symmetric and symmetric objects). This advantage was absent for children aged 8-9 (non-symmetric and symmetric objects). For the recognition of old views, all contrasts were non-significant.

Figure 2B shows that the recognition of old views improves faster with increasing age than that of new views and parallels the developmental trajectory of learning status, shown by the d'_{max} scores in Figure 2A. The performance difference between old and new views indicates a pronounced viewpoint dependency of recognition. However, this difference is strongly modulated by priming. Both visual and haptic priming distinctly improve classification of the new views, thereby reducing the old-new discrepancy up to the point where performance with regard to the two types of view becomes equivalent - the hallmark of viewpoint independence. Haptic priming proved more effective than visual priming in advancing the development of viewpoint independence: In case of the non-symmetric object a near independence is observed in children aged 9-10 for haptic priming, but only in children aged 13-14 for visual priming. In case of the symmetric objects, only haptic priming enables viewpoint-independence in the oldest children (15-16 years). In contrast, children in the control condition only show a marginal improvement of performance towards viewpoint independence in case of the non-symmetric object, whereas recognition of the symmetric objects remains clearly view-dependent throughout the tested age range.

Our data thus demonstrates a developmental dissociation between view-dependent and view-invariant object recognition during adolescence. Whereas the unsurprising improvement of learning performance with increasing age is directly reflected in the gradual improvement of recognition performance for old views in the generalization test, performance with regard to new views shows a marked retardation. Mental rotation skills are unlikely to account for this retardation as such skills develop much earlier in life, and children as young as five have been shown to use them for tasks involving perspective taking or spatial comparisons [22,23]. Moreover, even when assuming that older children find mental rotation easier to use for object recognition purposes than younger children the “difficulty” of mental rotation may have two reasons: It could mean that in young children (1) the necessary 3D object representations to which mental transformations such as 3D rotations could be applied are not yet in place, or (2) the skill proper to apply mental transformations to existing representations is insufficiently developed. The distinct facilitation of recognition brought about by haptic or visual priming clearly favours the first alternative. When classifying new (but not old) views, children aged 8-9 who had received haptic priming performed significantly better than subjects aged 9-13 with no priming (control condition) for both the non-symmetric ($t(8)=2.63$, $p<0.05$) and the symmetric ($t(18)=2.1$, $p<0.05$) objects. Given our paradigm it would seem unlikely that a few minutes of haptic exploration should boost subjects’ mental rotation skills to a level that exceeds performance of control subjects who are 4 years older.

A more parsimonious explanation is that priming (haptic or visual) triggers the recruitment of existing representations that preserve the spatial structure of 3D objects and permit the use of mental transformations (such as rotations) to achieve viewpoint independence. The observed retardation could then be explained on the assumption that young children do not have representations allowing view-independent recognition, or that such representations do exist but that children are disinclined to use them. Our present results do not permit to distinguish between these two possibilities as they provide partial evidence for either interpretation: the modulating effect of priming (visual or haptic) suggests a strategic component, whereas the absence of generalization skills in very young children would seem to indicate a rigid developmental constraint.

While the exact relationship between these two factors remains an issue for future research, our data clearly supports the idea of a non-unitary representation of objects in the brain. Multiple formats have been postulated previously to account for view-dependent and view-independent benefits in object priming and learning, and the type of representation tapped by a particular task has been shown to depend on object class [12,24] and the engagement of attention [7,9]. The possibility of a selective impairment of viewpoint independence in brain-damaged patients (e.g. [10,11,12,13]) suggests dissociable neural subsystems for the different formats, and even a hemispheric specialisation for certain classes of stimuli depending on their preferred representation [10,12,25].

Evidence from brain imaging [14] shows that areas selective for objects and non-objects during priming tasks form clusters with view-dependent areas in the right fusiform cortex, whereas areas only selective for objects (but not non-objects) cluster together with view-invariant areas in the left fusiform cortex. This indicates that learning (i.e. the process of a ‘non-object’ becoming an ‘object’) may involve the transition from a purely view-specific format to a dual representation including both a view-dependent and view-independent component (see also [7]). A dual representation could also account for a slight bias in our generalization data favouring old views (Fig. 2B), as recognition of the latter could rely both on the view-dependent and view-invariant component, whereas the recognition of novel views would necessarily rely on the view-invariant representation only.

As for the developmental trajectory of the observed dissociation between the two formats, the delayed onset of viewpoint independence and the facilitating effect of visuomotor interaction may be accounted for by late-developing structures in the parietal or the prefrontal cortex. Within the former, the angular gyrus is known to be one of the last myelinating areas of the human cortex [26] and an important platform of multisensory associative integration. Lesions of this area may cause Gerstman’s syndrome, which includes finger agnosia, left-right confusions as well as agraphia and which has been characterized as an impairment in the manipulation of mental images [27]. It seems conceivable that the late development of this structure may impede the ability to relate mental representations of objects to the motor programs that guide haptic or visual object exploration in our two priming conditions. A crucial role in the perception-action cycle is also played by the prefrontal cortex (PFC), which controls the generation of complex sensorimotor sequences by connecting sensory areas with motor areas in the cerebral cortex [28]. PFC is also one of the last brain regions to mature [29,30] and its development between the ages of 7 and 16 is strongly predictive of delayed verbal and visuo-spatial memory functions [29,31]. The maturation of the PFC could therefore enable humans to integrate haptic object knowledge from self-generated exploratory activity with visual object representations acquired during supervised learning.

In conclusion, we have shown that a fixed procedure of visual supervised learning may result in object representations with distinctly different invariance properties, depending on whether the learning has been preceded by a brief priming phase of (visual or haptic) exploration. On the one hand, our results add to the growing body of evidence suggesting the dual use of view-dependent and view-invariant object representations in the brain. On the other hand, the age dependency of the observed priming effects suggests that such representations have different developmental trajectories that extend well into adolescence and imply a retardation of view-invariant relative to view-dependent object formats.

Acknowledgement

This work was supported by the Heidehofstifung, Stuttgart, Germany, and by the Deutsche Forschungsgemeinschaft (DFG).

References

[1] Biederman I. Recognition by components: A theory of human image understanding. Psychol Rev, 1987; 4: 115-147.

[2] Hummel JE, Biederman I. Dynamic binding in a neural network for shape recognition. Psychol Rev, 1992;99: 480-517.

[3] Biederman I, Bar M. One-shot viewpoint invariance in matching novel objects. Vision Res, 1999;39: 2885-2899.

[4] Poggio T, Edelman S. A network that learns to recognize three-dimensional objects. Nature, 1990;343: 263-266.

[5] Bülthoff HH, Edelman S, Tarr MJ. How are three-dimensional objects represented in the brain? Cereb Cortex, 1995;5: 247-260.

[6] Riesenhuber M, Poggio T. Hierarchical models of object recognition in the cortex. Nat Neurosci, 1999;2: 1019-1025.

[7] Hummel JE . Complementary solutions to the binding problem in vision: Implications for shape perception and object recognition. Vis Cogn, 2001;8: 489-517.

[8] Foster DH, Gilson SJ. Recognizing novel three-dimensional objects by summing signals from parts and views. Proc R Soc Lond B, 2002;269: 1939-1947.

[9] Thoma V, Hummel JE, Davidoff J. Evidence for holistic representations of ignored images and analytic representations of attended images. J Exp Psychol Human, 2004;30: 257-267.

[10] Warrington EK, Taylor AM. Two categorical stages of object recognition. Perception, 1978;7: 695-705.

[11] Humphreys GW, Riddoch MJ. Routes of object constancy – implications from neurological impairments of object constancy. Q J Exp Psychol A, 1984;36: 385-415.

[12] Farah MJ. Visual agnosia: Disorders of object recognition and what they tell us about normal vision. Cambridge MA: MIT Press, 1990.

[13] Davidoff J, Warrington EK. The bare bones of object recognition: Implications of a case of object recognition impairment. Neuropsychologia, 1999;37: 279-292.

[14] Vuilleumier P, Henson RN, Driver J, Dolan RJ. Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. Nat Neurosci, 2002;5: 491-499.

[15] Klatzky RL, Lederman SJ. Identifying objects from a haptic glance. Percept Psychophys, 1995;57: 1111-1123.

[16] Newell FN, Ernst MO, Tjan BS and Bülthoff HH. View dependence in visual and haptic object recognition. Psychol Sci, 2001;12: 37-42.

[17] Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. Nature, 2002;415: 429-433.

[18] Grüsser OJ. Multimodal structure of the extrapersonal space. In: Hein A, Jeannerod, M, editors. Spatially oriented behaviour. New York: Springer, 1983, 327-352.

[19] Osman E, Pearce AR, Jüttner M, Rentschler I. Reconstructing mental

object representations: A machine vision approach to human visual recognition. Spatial Vision, 2000;13: 277-286.

[20] Rentschler I, Jüttner M, Osman E, Müller A, Caelli, T. Development of configural object recognition. Behav Brain Res, 2004;104: 107-111.

[21] Green DG, Swets JA. Signal Detection Theory and Psychophysics. Huntington: Krieger, 1974.

[22] Marmor GS. Development of kinetic images: When does the child first represent movement in mental images? Cognitive Psychol, 1975;7: 548-559.

[23] Estes D (1994) Young children's understanding of the mind: imagery, introspection, and some implications. J Appl Dev Psychol, 1994;15: 529-548.

[24] Marsolek CJ. Dissociable neural subsystems underlie abstract and specific object recognition. Psychol Sci, 1999;10: 111-118.

[25] Grüsser OJ, Landis T. Visual agnosia and other disturbances of visual perception and cognition. In: Cronly-Dillon JR, editor. Vision and Visual Dysfunction, vol 12. London: Macmillan Press, 1990.

[26] Flechsig P. Anatomie des Menschlichen Gehirns und Rückenmarks auf Myelogenetischer Grundlage. Leipzig: Thieme, 1920.

[27] Mayer E, Martory M-D, Pegna AJ, Landis T, Delaville J, Annoni J-M. A pure case of Gerstmann syndrome with a subangular lesion. Brain, 1999;122:1107-1120.

[28] Fuster JM. The prefrontal cortex – an update: time is of the essence. Neuron, 2001;30: 319-333.

[29] Casey BJ, Giedd JN, Thomas KM. Structural and functional brain development and its relation to cognitive development. Biol Psychol, 2000;54: 241-257.

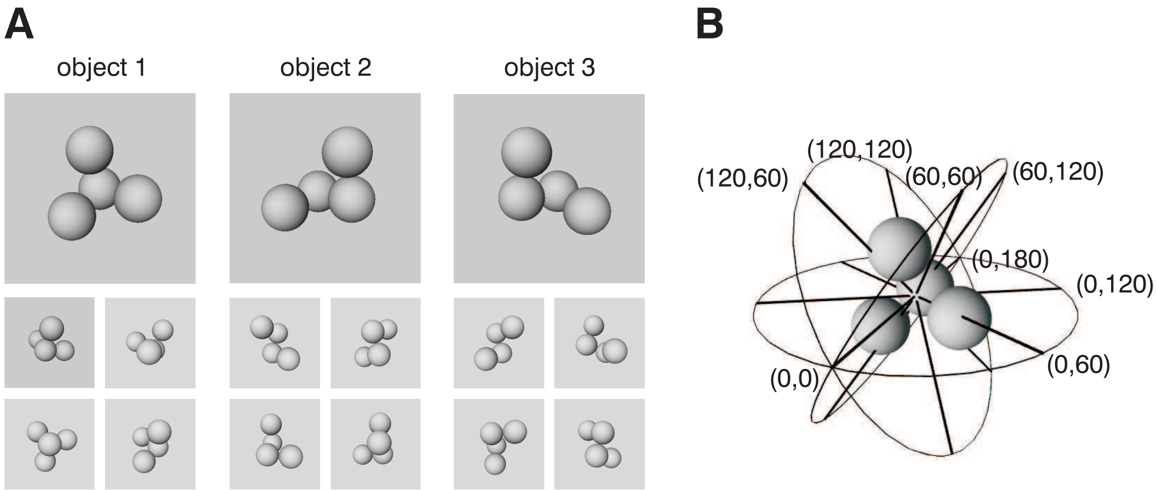
[30] Blanton RE, Levitt JG, Thompson PM, Narr KL, Capetillo-Cunliffe L, Singerman JG, McCracken JT, Toga AW. Mapping cortical asymmetry and complexity patterns in normal children. Psychiatry Res, 2001;107: 29-43.

[31] Sowell ER, Delis D, Stiles J, 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.

Figure Captions

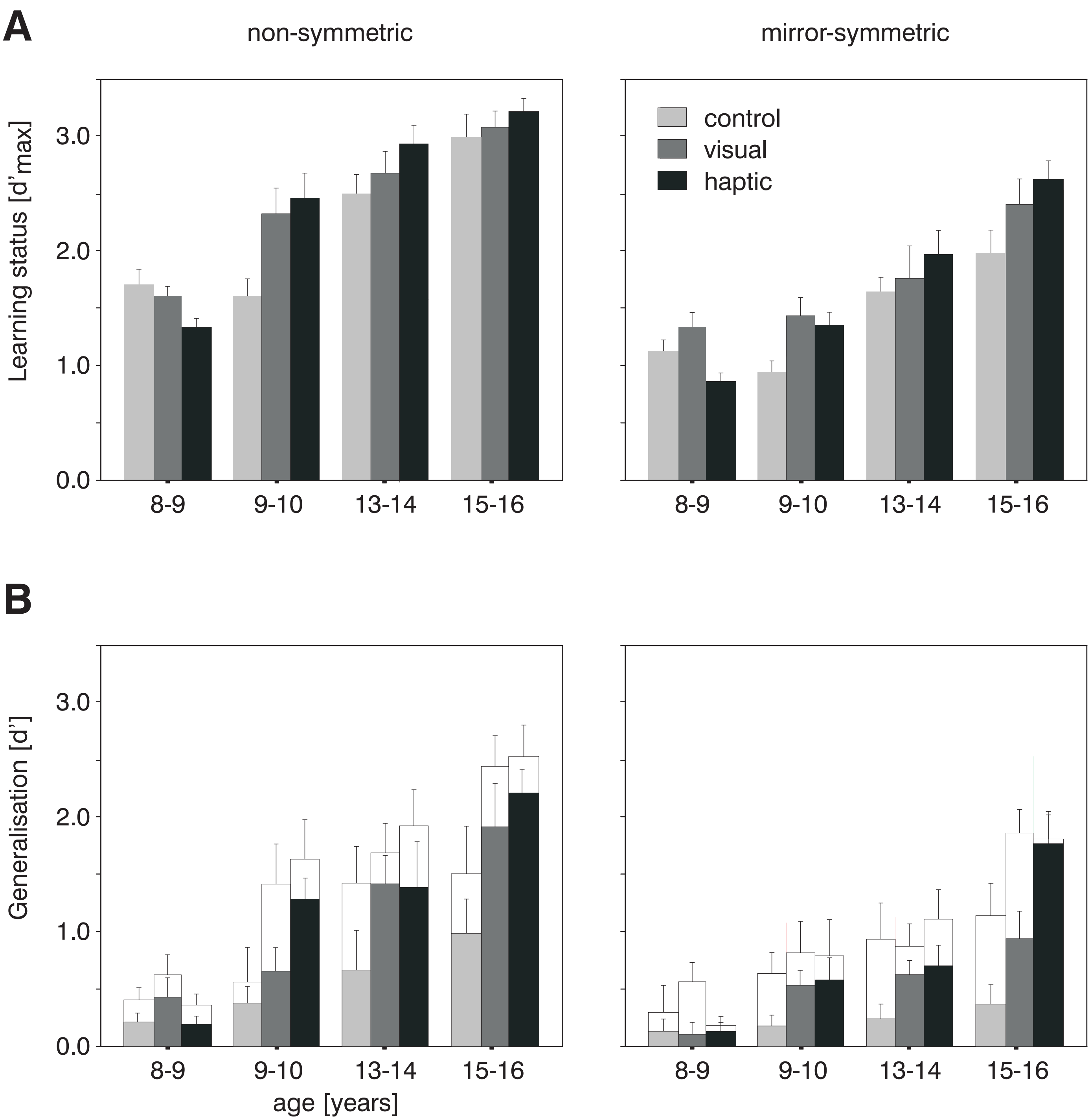
Figure 1. (**A**) Top: The three learning objects used in the experiments. Note that objects 2 and 3 are left/right mirror-symmetric versions of each other. The objects were generated both as real and as computer-graphic, virtual models. Bottom: Examples of learning views generated from the virtual models. **(B)** Sampling grid of the viewing sphere used to generate learning views. The sampling employed a step width of 60 deg plus a random angle around the virtual camera axis.

Figure 2. Effect of age and priming condition (haptic, visual, control) on learning status and generalisation performance. **(A)** Learning status, as reflected by the sensitivity measures d'_{max} for recognizing the non-symmetric object 1 and the mean d' for the (mirror-)symmetric objects 2 and 3. Individual d'_{max} scores were derived by transforming for each subject the classification frequencies of each object in the learning unit with the highest overall performance. **(B)** Spatial generalization for non-symmetric and symmetric objects. For each priming condition and age group, shaded bars represent the mean classification scores (d') of new views, whereas open bars refer to the classification of old views. Error bars S.E. (N=5).



A
object 1
object 2
object 3
B
(120,120)
(120,60)
(60,60)
(60,120)
(0,180)
(0,120)
(0,0)
(0,60)

Fig. 1



A
non-symmetric
mirror-symmetric
control
visual
haptic
Learning status [d' max]
3.0
2.0
1.0
0.0
8-9
9-10
13-14
15-16
B
Generalisation [d']
age [years]

Fig. 2