When Touch Forms Vision - Object Recognition as a Function of Polysensory Prior Knowledge

Martin Jüttner
Neuroscience Research Institute, School of Life & Health Sciences Aston University, UK
m.juttner@aston.ac.uk

Erol Osman

Ingo Rentschler
Institute of Medical Psychology University of Munich, Germany

Abstract

We investigated how various forms of prior knowledge influence learning speed and recognition performance of previously unfamiliar three-dimensional (3D) objects. The learning objects were three 'molecule' models each of which was composed of four spheres. Prior knowledge was varied in terms of the sensory modality (visual versus haptic versus visuohaptic) employed during a so-called exploration phase. Following this exploration subjects were trained in a visual learning task with a fixed set of two-dimensional (2D) views of the learning objects. We found a significant effect of sensory modality employed during the exploration phase on both learning rate and recognition performance in the subsequent visual learning task. Specifically, a short period of haptic or visuohaptic exploration reduced learning duration by about 60% relative to conditions with either none or visual-only exploration. Computer simulations on the basis of the behavioural data suggest that prior haptic exploration stimulates the evolution of object representations which are characterized by an increased differentiation between attribute values and a pronounced structural encoding.

1. Introduction

Concerning the quality of internal representations underlying human object recognition there are two dominant views. On the one hand, it has been postulated that objects are mentally represented by three-dimensional (3D), object centred, part-based descriptions [1,2,3]. On the other hand, more recent studies have provided evidence for the notion that 3D objects are in fact represented in terms of multiple, viewer centred, two-dimensional (2D) views, among which the visual system interpolates if necessary [4,5,6].

Experimentally, both hypotheses have been mainly tested in studies where the test object was presented from different perspectives and where the change of the error rate or response latency has been measured in identification tasks as a function of viewing angle. However, it has been shown that the two alternative explanations are not readily distinguished within this paradigm. First, the dependency from viewpoint is itself dependent on object familiarity [7], demonstrating the necessity to take into account learning processes. Second, a closer inspection of the apparent complementary approaches shows that the assumed viewpoint-invariance of the 3D hypothesis holds only under certain conditions [8]. Conversely, representations in terms of multiple 2D views may become quasi-independent from viewpoint, if the number of views is sufficiently large, or if the interpolation mechanism between views becomes more efficient due to training.

A further complication arises from the fact that it is still unclear to what extent depth information is used for building internal representations of 3D shapes. There is evidence that the influence of binocular disparity (shown to be strongest contributor to depth information in recognition) declines with increasing familiarity [9]. In contrast, experiments in the field of haptic object recognition demonstrate that the identity of familiar objects may be established very quickly and seems to be mediated mainly by 3D structural information [10,11]. The latter result would suggest a participation of polymodal sensory systems in the ontogenesis of mental object representations.

In this study we have investigated, within a supervised learning paradigm, how prior knowledge of objects in various sensory modalities influences learning speed, recognition performance and the ability of spatial generalization. The learning objects were molecule-like models made up of four spheres. Prior knowledge was varied in terms of the sensory modality (visual versus haptic versus visuohaptic) employed during a so-called exploration phase. Following this exploration subjects were trained in a visual learning task with a fixed set of two-dimensional (2D) views of the learning objects. Learning speed was

measured as the number of training cycles that were necessary to reach a given criterion concerning the classification of the learning views. In a subsequent generalization test, recognition was assessed by the ability to correctly assign a set of novel 2D views of the previously learned objects. To gain further insight into the nature of the mental object representations acquired during learning additional computer simulations were conducted. The simulations employed a recognition model based on relational evidence theory [12] which allows to characterize object representations in terms of components and relational rules.

2. Method

The experiments employed a set of three objects. Each object was composed of four spheres, with three of them forming an isosceles triangle and the fourth being placed perpendicular above the centre of one of the base spheres rendering the objects similar to 'molecule' models. The objects were generated both as virtual models and as physical models. Virtual models were constructed and displayed on a SGI O2 workstation using the Open Inventor software package. The physical models were constructed of styrofoam balls. From the virtual models, 2D views were generated as perspective projections of the objects onto the screen plane of the computer display. Two sets of views were generated, a training set comprising 22 images and a test set comprising 83 images. The two sets were obtained by sampling the viewing sphere in 60 deg steps (training set) and 30 deg steps (test set), respectively, and by eliminating all views that were redundant due to object symmetry. At the viewing distance of 1m the images appeared under a visual angle of 1.5 deg.

The experiment was divided into three parts, a so-called exploration phase, a supervised-learning phase and generalisation test. During the exploration phase the participants were allowed to familiarize themselves with the objects. The subjects were divided into three groups that differed concerning the sensory modality employed during the exploration: visual, haptic, visuohaptic. Haptic and visuohaptic exploration was done by grasping and manipulating the physical object models. In the purely haptic case the subjects were blindfolded. Visual exploration was mediated by actively rotating the virtual object models on the screen by means of the computer mouse. In each case, the exploration phase lasted for 3 minutes and was followed immediately by the visual-learning phase. In addition, there was a fourth (control) group where subjects started immediately with the visual learning without any prior exploration.

The supervised learning procedure consisted of subsequent learning units (see [13]). In each unit the each view of the learning set was presented three times in random order. The exposure duration of each view was 250 msec, and each presentation was followed by the corresponding object label displayed for one second. The learning unit ended with a recognition test to assess the learning status of the subject. Here each view of the learning set was shown once for 250 msec and labelled by the observer. The sequence of learning units continued until the subject had reached the learning criterion of 90% correct responses in the recognition test. After having completed learning the observers continued with the generalisation test. Here all views of the test set were shown in random order. Each test view had to be assigned to the previously learned objects.

3. Results

The results show that learning duration, measured by the number of learning units necessary to reach the criterion, was significantly affected by sensory modality employed during the exploration phase. Specifically, the short haptic exploration phase reduced the learning duration by about 60% as compared to the control condition. Haptic exploration also distinctly improved the ability to recognize novel views of the previously learned objects in the generalisation test, from a level of about 60% to about 75% correct responses. The advantage in learning and generalisation was the same no matter whether the haptic exploration was accompanied by sight (in the visuohaptic condition) or not (in the haptic condition). However, a purely visual exploration of the objects (visual condition) yielded no significant advantage relative to the control condition, neither in terms of learning speed nor generalization.

To explore how prior knowledge acquired during the exploration phase specifically affected the ontogenesis of the object representations during the subsequent visual learning we conducted computer simulations on the basis of the confusion error data. The simulations employed CLARET, a recognition model based on relational evidence theory [12]. CLARET involves a process of rule induction, in which a taxonomy of rules is created to discriminate between objects. The rules refer to relational attributes, such as distance, angle or size ratio, that are used to describe object structure. Starting from an initial rule, defined by an attribute range which is satisfied by the relations of all existing objects, rule refinement proceeds in two ways: (1) By re-partitioning the attribute space via clustering, thus producing rules of increasing attribute specificity, and (2) by incorporating relational information in terms of rules which include other rules in their body. Thus the resulting rule hierarchy is characterized by two properties: attribute specificity and relational depth.

Since the behavioural data revealed a distinct dichotomy between subjects with haptic versus those with non-haptic prior object experience, the data of the subjects were divided

into two groups, haptic and non-haptic, respectively. The simulations yielded for the haptic group a high degree of differentiation between values along the feature dimensions, with an optimal fit around a value of 5 partitions. In contrast, observers with no prior experience tended to differentiate between attribute values much more coarsely, with an optimum partition number around 1-2. Concerning relational depth, only the fit for the haptic but not for the non-haptic group improved as more rules were included in rule generation. This warrants the conclusion that object rules in the former case extend over multiple object components whereas they remain confined to isolated component pairs in the latter.

4. Conclusion

We have shown that a prior haptic exploration of 3D objects leads to a substantial facilitation in visual learning and generalisation. Computer simulations suggest that theses facilitation may be due to a different representational format of the underlying mental object representations. Accordingly, prior object knowledge acquired via the haptic sense stimulates the development of representations that are characterized by a high degree of attribute differentiation and an enhanced use of relational information. The present results are preliminary in the sense that they do not yet allow conclusions concerning the dimensionality of mental representations of 3D objects. However, the observed dependency of visual learning on haptic information indicates that such object representations may have an intrinsic polymodal quality, thus questioning the validitiy of purely vision-based accounts of object recognition.

References

[1] D. Marr, and H.K. Nishihara, K., Representation and recognition of the spatial organization of three-dimensional shapes, *Proceedings of the Royal Society of London 200B,* 1978, 269-294.

[2] I. Biederman, I., Recognition-by-components: A theory of human image understanding. *Psychological Review 94,* 1987, 115-147.

[3] I. Biederman, Recognizing depth-rotated objects: A review of recent research and theory, *Spatial Vision 13,* 241-254.

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

[5] S. Edelman, Representation, similarity, and the chorus of prototypes, *Mind and Machines 5,* 1995, 45-68.

[6] M.J. Tarr, P. Williams, W.G. Hayward, and I. Gauthier, Three-dimensional object recognition is viewpoint dependent, *Nature Neuroscience 1,* 1998, 275-277.

[7] M.J. Tarr and S. Pinker, Mental rotation and orientation-dependence in shape recognition, *Cognitive Psychology 21,* 1989, 233-282.

[8] I. Biederman and P.C.Gerhardstein, Viewpoint-dependent mechanisms in visual object recognition, *Journal of Experimental Psychology, Human Perception and Performance 21,* 1506-1514.

[9] S. Edelman and H.H. Bülthoff, Orientation dependence in the recognition of familiar and novel views of 3D objects, *Vision Research 32,* 1992, 2385-4000.

[10] R.L. Klatzky and S.J. Lederman, Identifying objects from a haptic glance, *Perception & Psychophysics 57,* 1995, 1111-1123.

[11] R.L. Klatzky and S.J. Lederman, The haptic glance: A route to rapid object identification and manipulation. *Attention and Performance 17,* 1999, 165-196.

[12] A.R. Pearce and T. Caelli, Interactively matching hand-drawings using induction, *Computer Vision and Image Understanding 73,* 1999, 391-403.

[13] I. Rentschler, M. Jüttner, and T. Caelli, Probabilistic analysis of human supervised learning and classification, *Vision Research 34,* 1994, 669-687.