Current Biology 18, 597–601, April 22, 2008 ©2008 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2008.03.058

Report

Fragment-Based Learning of Visual Object Categories

Jay Hegdé,[1,*] Evgeniy Bart,[2] and Daniel Kersten[1]
[1]Department of Psychology
University of Minnesota
Minneapolis, Minnesota 55455
[2]Department of Electrical Engineering
California Institute of Technology
Pasadena, California 91125

Summary

When we perceive a visual object, we implicitly or explicitly associate it with a category we know [1–3]. It is known that the visual system can use local, informative image fragments of a given object, rather than the whole object, to classify it into a familiar category [4–8]. How we acquire informative fragments has remained unclear. Here, we show that human observers acquire informative fragments during the initial learning of categories. We created new, but naturalistic, classes of visual objects by using a novel "virtual phylogenesis" (VP) algorithm that simulates key aspects of how biological categories evolve. Subjects were trained to distinguish two of these classes by using whole exemplar objects, not fragments. We hypothesized that if the visual system learns informative object fragments during category learning, then subjects must be able to perform the newly learned categorization by using only the fragments as opposed to whole objects. We found that subjects were able to successfully perform the classification task by using each of the informative fragments by itself, but not by using any of the comparable, but uninformative, fragments. Our results not only reveal that novel categories can be learned by discovering informative fragments but also introduce and illustrate the use of VP as a versatile tool for category-learning research.

Results

Using VP to Create Shape Classes

The VP algorithm generates naturalistic object categories by emulating biological phylogenesis (see Supplemental Data available online). With VP, we created three classes of novel objects, classes A, B, and C and used 200 exemplars from each (Figure 1). Note that the three classes are very similar to each other, so that distinguishing among them is nontrivial (see below and Figure S1). Moreover, no two objects, including objects within a given category, were exactly alike, so that distinguishing among them required learning the relevant statistical properties of the objects and ignoring the irrelevant variations. Finally, note that the differences between categories arose spontaneously and randomly during VP, rather than as a result of externally imposed rules.

Extracting Informative Fragments

We isolated ten fragments ("Main" fragments, Figures 2A and 2B) that were highly informative for distinguishing class A from class B (the main task in experiment 1, see Supplemental Data for details). We also isolated ten "Control" fragments (Figures 2C and 2D) and ten "IPControl" fragments (Figure S2) that were uninformative for the main task but visually comparable to the main fragments. The mutual information (MI) value of a given fragment quantifies the information it conveys about a given category. The higher the fragment's MI, the more useful the fragment is for categorization. The MI values of all fragments used in this study are listed in Supplemental Data.

Testing the Informativeness of Individual Fragments

The experiments consisted of training the subjects on whole objects and then testing them on fragments. Because only whole objects, not fragments, were used during training, subjects were not aware of the fragments or required to learn them. After the subjects were trained in the task, we tested the extent to which subjects were able to perform the classification task by using the fragments, each presented individually (see Figure 3 and Supplemental Data). We hypothesized that if the subjects learned informative object fragments during the training, then the subjects must be able to perform the categorization task by using the individual main fragments, but not the control fragments.

The observed performance closely matched these predictions. Figure 4A shows the average performance of six subjects using the main fragments. Subjects performed significantly above chance with each of the fragments (binomial tests, $p < 0.05$ in each case). Moreover, with one exception (see below), the performance of each individual subject with each main fragment was indistinguishable from his/her performance with whole objects during the final two training sessions (binomial tests, $p > 0.05$, data not shown). The only exception to this was the performance of one subject with main fragment #9, for which she classified the object containing the fragment as A in only 1/16 (6.25%) of the trials (also see below). Altogether, these results indicate that the subjects were able to categorize the objects on the basis of each of the fragments alone and that the performance with the fragments was generally indistinguishable from the performance of the subjects with the whole object.

By contrast, subjects were unable to perform the task above chance levels by using any of the control or IPControl fragments (Figures 4B and 4C; binomial tests, $p > 0.05$). That is, subjects were about equally likely to classify an object as belonging to class A or class B on the basis of a given control or IPControl fragment. Thus, although all three types of fragments belonged to class A, only the main fragments were likely to be assigned to class A.

To ensure that above results were not a function of a fortuitous designation of object classes, we performed experiment 2 in which we repeated the design of experiment 1, but with a different set of class designations, whereby the main task was to distinguish class C from class B (see Figure S4). A different set of four subjects participated in this experiment. The results of this experiment were similar to those in experiment 1 (Figure S5).

Additional analyses indicated the performance showed no improvement during the testing phase of the experiment,

*Correspondence: hegde@umn.edu

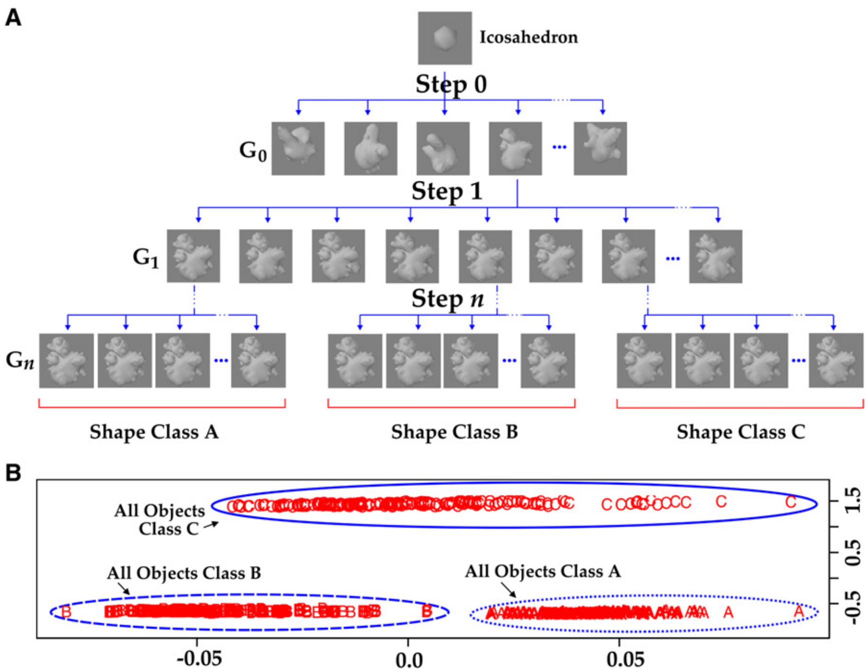

Figure 1. Generating Naturalistic Shape Classes by "Virtual Phylogenesis"

(A) The VP algorithm emulates biological evolution in that in both cases, novel objects and object classes emerge as heritable variations accumulate selectively. In the present study, we used a class of novel objects called "digital embryos," which develop from a given parent object through simulated embryonic developmental processes [17]. At each generation G_n, selected embryos procreate, leading to generation G_{n+1}. The progeny inherit the shape characteristics of their parent but accrue random shape variations of their own as they develop. Thus, children of a given parent constitute a shape class. In the present study, embryos were grown for four generations with the VP algorithm, starting from a single common ancestor, an icosahedron. Three shape classes (A, B, and C) were chosen at generation n = 4, each with ~1500 "siblings." Note that the entire object-generation process operated completely independently of the fragment-selection process or any other classification scheme. For larger images of exemplar objects from each class and for a demonstration that the categorization task is nontrivial and cannot be performed without learning the relevant classes, see Figure S1.

(B) A metric multidimensional scaling (MDS) plot of the 600 objects, 200 each from class A, B, and C, used in this study. Pixel-wise correlations of gray-level values were used as the input to MDS. Each data point represents one individual object, and the plotting symbol (A, B, or C) denotes the class to which the object belonged. MDS plots the data points so as to cluster similar data points together and disperse dissimilar data points from each other (for details, see [9, 18]). The values on either axis denote the class distance measures used by the MDS. Note that the two axes have different scales. The objects of the three classes formed three nonoverlapping clusters (ellipses), so that each cluster contained all the objects, and only the objects, of a given class.

indicating that the subjects learned the fragments during the training phase, i.e., before the testing began (see Figure S6).

Necessity of Prior Training

In additional experiments, subjects were tested with informative fragments without having learned the categories beforehand (i.e., with the training phase omitted). Six subjects were used, five of whom also participated in experiment 1 above and one who participated in experiment 2. All subjects performed at chance levels (binomial tests, $p > 0.05$; Figure 4D). The performance was also indistinguishable from chance when the testing was preceded by training with similar, but task-irrelevant object categories (Figure 4E). This confirms that the categorization task required learning and in particular

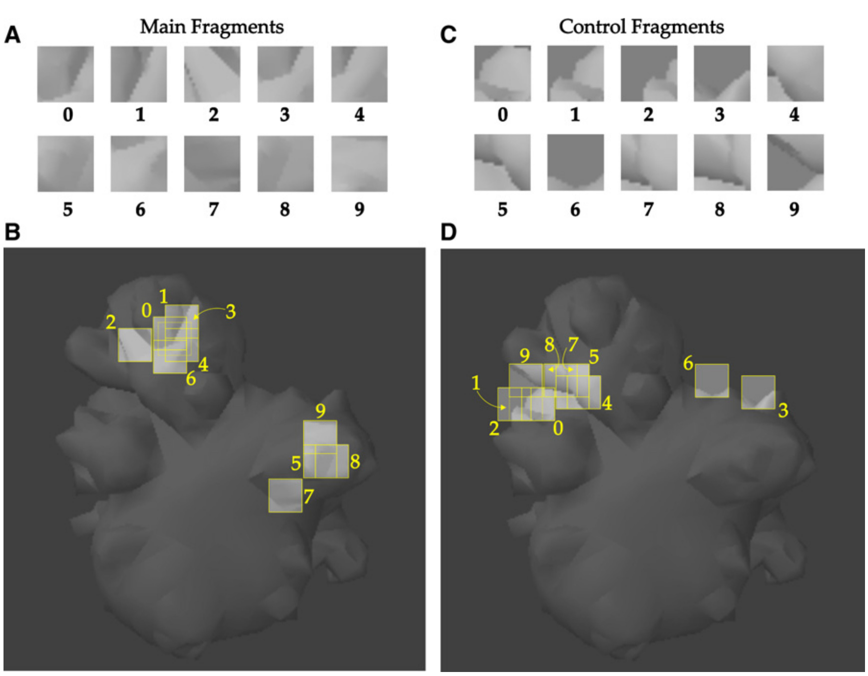

Figure 2. Informative Object Fragments

(A) Main fragments, which are 20 × 20 pixel fragments of objects from class A that are useful for distinguishing class A from class B (main task).

(B) Location of the main fragments, overlaid on a typical object from class A. Fragment borders are outlined in yellow for clarity.

(C) Control fragments, which are fragments of objects that are not useful for the main task from class A (see Supplemental Data for details).

(D) Location of the control fragments.

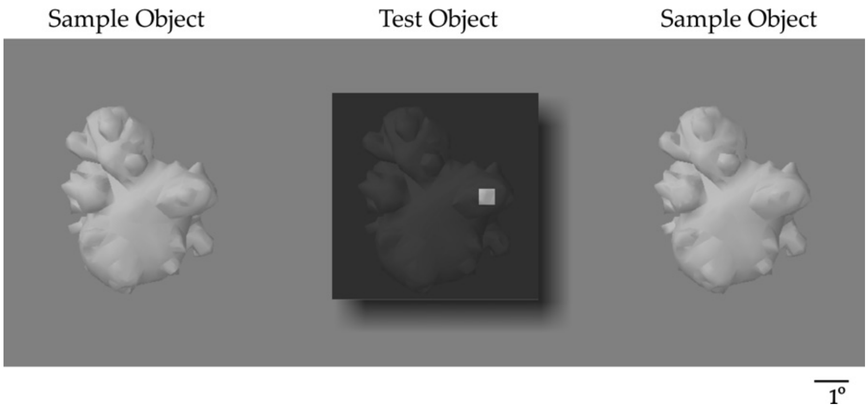

Figure 3. The Testing Paradigm

A test object (center) and two sample objects, one from each class (left and right), were simultaneously shown. The test object was occluded by a translucent surface with a hole, such that only the given object fragment was visible, unoccluded, through the hole, and the location of the fragment relative to the overall object was apparent through the translucent occluder. Subjects had to classify the object into the class exemplified by the sample object on the left or right on the sole basis of the fragment visible through the hole. Subjects were informed that only the fragment, but not the darkened remainder of the test object, was useful for the task. See Supplemental Data for details. The fragment shown in this figure is the same as fragment 5 in Figures 2A and 2B.

that the subjects could not perform the task during the testing phase by simply comparing the given fragment to the two whole objects in the display.

Learning Fragments Was Not Necessary

It is clear from the metric multidimensional scaling (MDS) plot of the three classes in Figure 1B that exemplars form three nonoverlapping clusters, each corresponding to one of the classes. The three classes are obviously linearly separable in this plot, as evident from the fact that one can draw a straight line separating any class from the other two. The fact that the projection found by MDS is linear [9] means that the original images are also linearly separable (in the pixel space). Therefore, subjects could have learned to separate the categories with complete images and did not have to learn object fragments.

Discussion

Our study is novel in two important ways. First, it reveals that informative fragments are learned during category learning. Second, it illustrates VP as a potentially powerful new tool for category-learning research.

Fragment Learning as a Part of Category Learning

Our results indicate that subjects learn informative, intermediate-complexity fragments as a matter of course when they learn new object categories, even when they were not explicitly required to learn the fragments. In other words, fragment learning was incidental to category learning. This result is significant because it straightforwardly links category learning with categorization, in that informative fragments play a role in both.

The performance of the subjects was a function of the task relevance of the fragments because subjects did not consistently associate task-irrelevant fragments to learned categories, even when the fragments were otherwise visually interesting or were informative for distinguishing the objects from another class. Together, these results reveal, for the first time, that humans selectively learn informative fragments as a part of category learning. Note that it would not have been possible to elucidate this by testing fragments from familiar categories (e.g., faces or cars; q.v. [7, 8].) because objects of

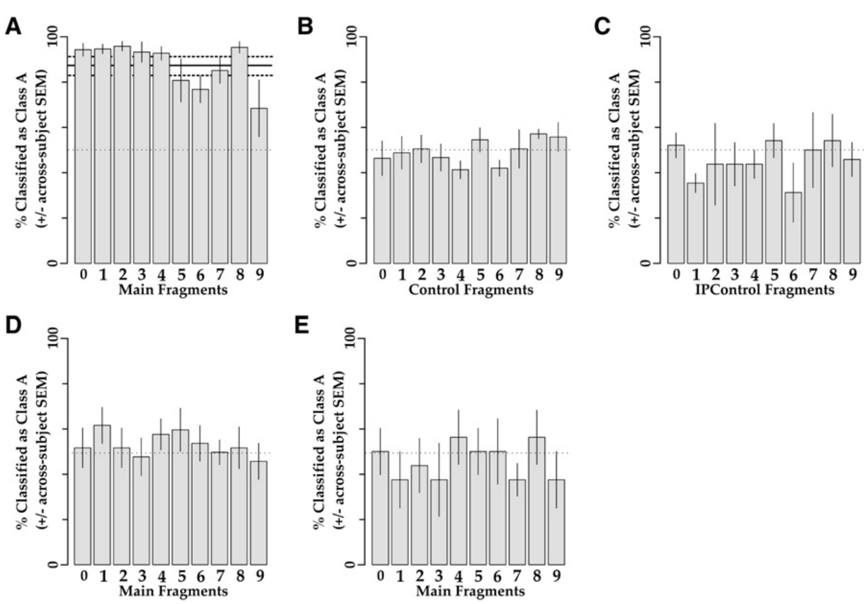

Figure 4. Classification Performance Using Fragments

In each panel, each bar shows the average percentage (±SEM) of trials in which the subjects classified a given fragment as belonging to class A. The thin dotted line denotes 50%, or chance level performance. The thick black lines in the background in (A) denote the mean (solid line) and the SEM (dashed lines) of the subjects using whole objects during the last two sessions of training.

(A and B) Performance in experiment 1a (six subjects) with main fragments (A) and control fragments (B).

(C) Performance in experiment 1b with IPControl fragments (three of the six subjects). The IPControl fragments are shown in Figure S2. The performance with main fragments from experiment 1b is shown in Figure S3.

(D) Performance with main fragments without prior training. Subjects were tested with the same paradigm as above, but without any prior training in the categorization task. The data are averaged from six subjects.

(E) Testing with irrelevant training. Data are shown from one subject. The subject was trained in a similar, but irrelevant, categorization task and tested with the main fragments with the same paradigm as above.

these categories are frequently seen occluded, so that fragment learning could be attributed to the necessity for overcoming occlusions.

In previous studies of novel category learning, the algorithm for generating novel objects depended on the algorithm for classifying them into categories [1, 2], whereas the two were independent in our case, as they are in nature. To the extent that our stimuli and the experimental conditions reflected category learning under natural conditions, our results indicate that such incidental learning of fragments may be a common principle of learning of natural object categories (see below).

Subjects' performance with task-relevant fragments was comparable to performance with the whole objects, suggesting that the learning of each of the fragments could, in principle, account for all or most of the category acquisition. Moreover, subjects performed close to perfect with most individual task-relevant fragments, indicating that the subjects were able to acquire most of the information conveyed by the individual task-relevant fragments (all of which had MIs at or near 1, see Supplemental Data).

Some models of perceptual learning, most notably the reverse-hierarchy theory [10], have suggested that subjects learn local features only when more global features do not suffice. In brief, reverse-hierarchy theory posits that learning takes place in spatially global-to-local fashion, such that the visual system initially learns large-scale features relevant to the task and "resorts" to finer-scale features when the large-scale features do not suffice. In our case, it was clearly not computationally necessary to learn the fragments because the tasks could be performed on the basis of whole objects (see Figure 1B). One reason why subjects nonetheless learned the fragments may be that the fragments were highly informative about the task in our case. Another, mutually nonexclusive possibility is that fragments represented the optimal spatial scale for learning in this case because the individual fragments were small enough to fit in the fovea, whereas it would have necessitated integration of information across multiple fixations to perform the task at level of the whole object. Further experiments are needed to resolve these issues.

Implications for the Mechanisms of Category Learning

Two previous studies, Harel et al. [7] and Lerner et al. [8], have examined the extent to which informative fragments support categorization of objects into familiar categories. Both showed that the ability of subjects to decide whether a given fragment was a part of a familiar object (e.g., a car or a face [7]) correlated with the MI of the fragment. Our study differed from these previous studies in several key respects, three of which are particularly worth noting. First, by using novel stimuli classes, we were able to study category learning, rather than just categorization. Second, because we controlled subject training, our fragments were extracted from the same set of images used by subjects during category learning. Third, we eliminated the possibility that the subjects might have learned the fragments out of necessity (e.g., to cope with occlusions) by ensuring that (1) the training images were completely unoccluded and (2) the classes were linearly separable, so that the categorization tasks could be performed on the basis of whole objects.

Our experiments did not test whether new categories can be learned solely from informative fragments. This is because our goal was to study learning under natural viewing conditions. In general, views strictly confined to informative fragments are highly unlikely under natural viewing conditions. Our result that subjects learned informative fragments even when presented with whole objects is therefore of greater relevance to natural vision.

Usefulness of VP in Categorization Research

Apart from the fact that the VP algorithm represents a novel method of creating object categories (c.f., "Greebles" [3, 11]), the resulting categories have several desirable features for the study of categorization and category learning. First, the categories have measurable, but randomly arising, within-class shape variations (c.f., [12, 13]). In most of the earlier studies using object categories created by compositing shape primitives, there tends to be little or no within-class variation (for reviews, see [14–16]). However, in natural scenes, two exemplars of a given category are seldom identical. Second, if necessary, both within-class variants and between-class variants in VP can be artificially selected to fit desired distributions (although we did not impose any such distributions in the present study). This means that the categories can be generated on the basis of, or independently of, an a priori classification algorithm, as desired. Third, VP can be used to generate a hierarchy of categories, directly analogous to the phylogenetic hierarchy of categories of biological objects in nature, so that VP can be a useful tool for exploring our hierarchical understanding of natural objects [1–3, 13, 16]. Finally, note that although we used "digital embryos" as the substrate for VP in the present study (Figure 1A), any virtual object, biological or otherwise, real-world or novel, can be used as a VP substrate and the algorithm can be readily modified to simulate a more complex phylogenetic process (e.g., convergent evolution, in which different taxa, such as whales and fish, come to resemble similar visual categories). Altogether, VP represents a powerful and versatile tool for generating naturalistic categories.

Supplemental Data

Additional Results, Experimental Procedures, seven figures, and two tables are available at http://www.current-biology.com/cgi/content/full/18/8/597/DC1/.

Acknowledgments

This work was supported by ONR grant N00014-05-1-0124 to D.K. and by NEI grant R01 EY017835. E.B. was a postdoctoral associate at the Institute of Mathematics and its Applications, University of Minnesota, during parts of this work. We thank Dr. Sharon Jansa, Dr. Norman Johnson, Dr. Paul Schrater and Mr. Peter Battaglia for helpful advice and discussions.

Received: January 30, 2008
Revised: March 20, 2008
Accepted: March 20, 2008
Published online: April 17, 2008

References

1. Mervis, C.B., and Rosch, E. (1981). Categorization of natural objects. Annu. Rev. Psychol. *32*, 89–115.
2. Smith, E.E., and Medin, D.L. (1981). Categories and Concepts (Cambridge, MA: Harvard University Press).
3. Palmeri, T.J., and Gauthier, I. (2004). Visual object understanding. Nat. Rev. Neurosci. *5*, 291–303, Erratum: (2004). Nat. Rev. Neurosci. 5, 510.
4. Ullman, S., Vidal-Naquet, M., and Sali, E. (2002). Visual features of intermediate complexity and their use in classification. Nat. Neurosci. *5*, 682–687.
5. Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. Trends Cogn. Sci. *11*, 58–64.

6. Bart, E., Byvatov, E., and Ullman, S. (2004). View-invariant recognition using corresponding object fragments. In Proc. ECCV, Part II (New York, NY: Springer-Verlag), pp. 152–165.
7. Harel, A., Ullman, S., Epshtein, B., and Bentin, S. (2007). Mutual information of image fragments predicts categorization in humans: Electrophysiological and behavioral evidence. Vision Res. *47*, 2010–2020.
8. Lerner, Y., Epshtein, B., Ullman, U., and Malach, R. (2007). Class information predicts activation by object fragments in human object areas. J. Cogn. Neurosci. *20*, 1–28.
9. Kruskal, J.B., and Wish, M. (1978). Multidimensional Scaling (Newbury Park, CA: Sage Publications).
10. Ahissar, M., and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. Trends Cogn. Sci. *8*, 457–464.
11. Palmeri, T.J. (1999). Learning categories at different hierarchical levels: A comparison of category learning models. Psychon. Bull. Rev. *6*, 495–503.
12. Sokal, R.R. (1983a). A phylogenetic analysis of the Caminalcules. I. The data base. Syst. Zool. *32*, 159–184.
13. Sokal, R.R. (1983b). A phylogenetic analysis of the Caminalcules. II. Estimating the true cladogram. Syst. Zool. *32*, 185–201.
14. Markman, A.B., and Ross, B.H. (2003). Category use and category learning. Psychol. Bull. *129*, 592–613.
15. Ashby, F.G., and Maddox, W.T. (2005). Human category learning. Annu. Rev. Psychol. *56*, 149–178.
16. Rosch, E. (1978). Principles of categorization. In Cognition and Categorization, E. Rosch and B.B. Lloyd, eds. (Hillsdale, NJ: Erlbaum).
17. Brady, M.J., and Kersten, D. (2003). Bootstrapped learning of novel objects. J. Vis. *3*, 413–422.
18. Duda, R.O., Hart, P.E., and Stork, D.G. (2001). Pattern classification, Second Edition (New York, NY: Wiley Interscience).

Supplemental Data

Fragment-Based Learning of Visual Object Categories

Jay Hegdé, Evgeniy Bart, and Daniel Kersten

Supplemental Results

The Categories Used Were Fine-Grained, and the Classification Task Was Nontrivial

One concern about the testing paradigm we used (Figure 3) is that the subjects can "cheat," i.e., do the task by comparing a given fragment to the corresponding parts of individual sample objects from the relevant categories. Figure S1 may be used to convince oneself that it is not possible to do the task reliably in this fashion. Choose a fragment of interest from a given object, but do not look up its class designation. Next choose one object each from class A and class B. Assign the fragment to either category by comparing the chosen fragment to the chosen objects. Repeat this several times for different objects and fragments and estimate your performance for each fragment. Empirical data show that, although the subjects could in principle adopt this strategy, in practice they do not do so (Figures 4D and 4E).

Fragments Were Learned during Training and Not Testing

Although the performance with the main fragments during the testing phase was comparable to the performance with whole objects during training, it is possible that at least some of the learning took place during the testing phase, especially because the subjects encountered the fragments repeatedly during the testing phase. This issue is germane to whether fragment learning accompanies category learning per se. It is unlikely that the subjects learned fragments during the testing, both because no feedback was provided during testing and because no more than 50% or 33% of the fragments (in experiment 1 or 2, respectively) were informative about the task.

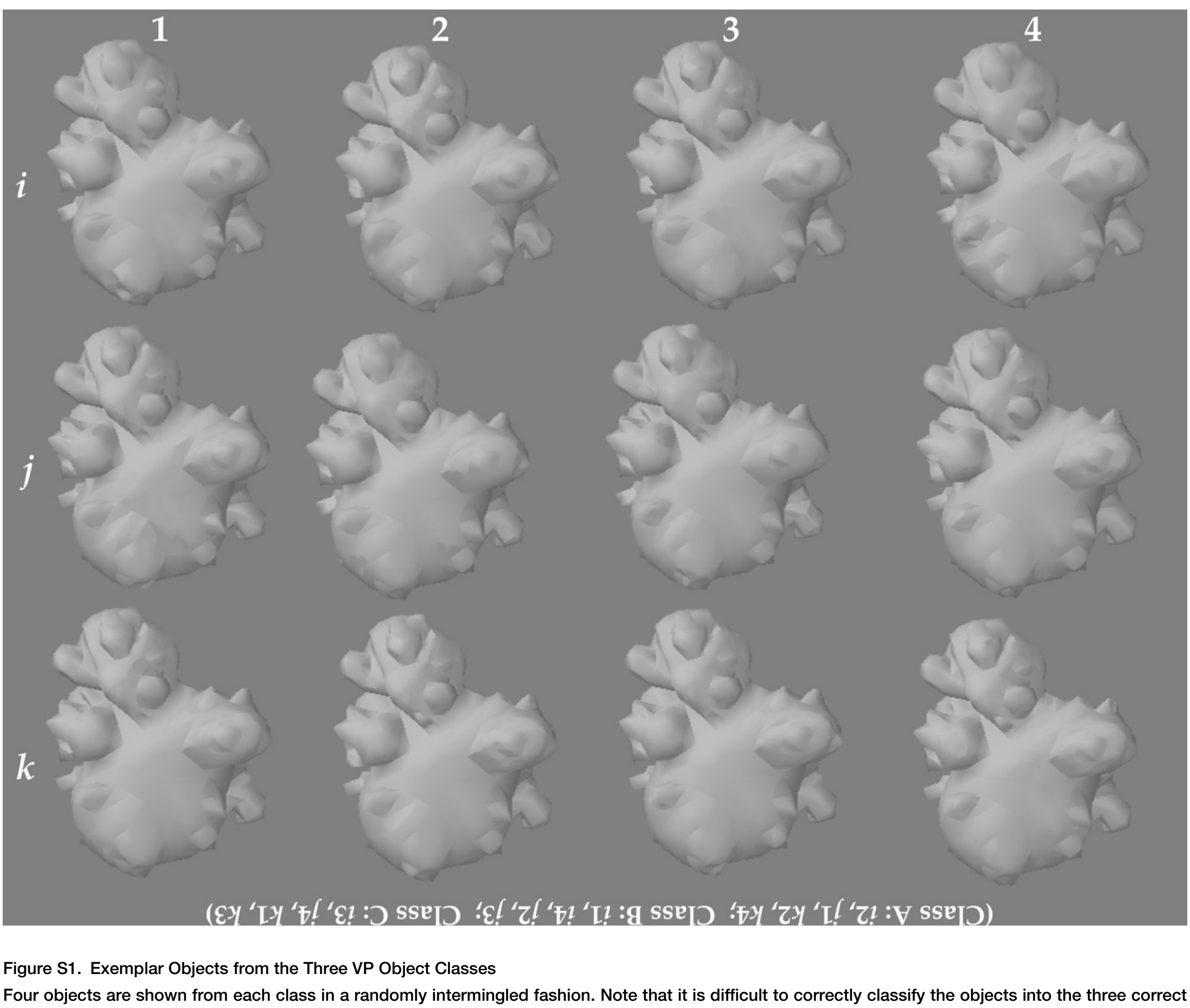

Figure S1. Exemplar Objects from the Three VP Object Classes

Four objects are shown from each class in a randomly intermingled fashion. Note that it is difficult to correctly classify the objects into the three correct classes without learning or knowing the classes. The class designations are shown at the bottom of the figure.

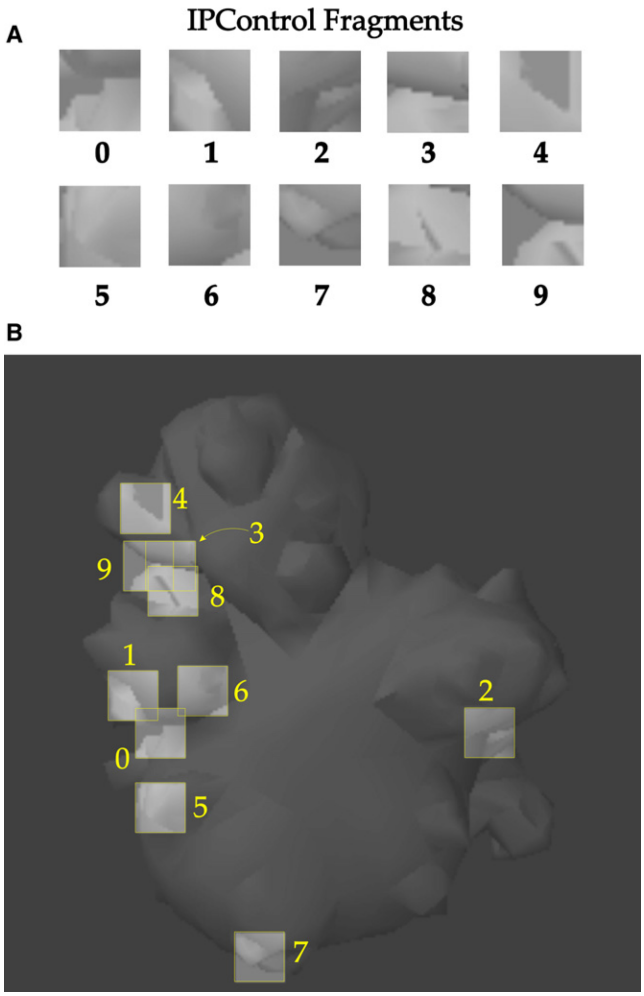

Figure S2. IPControl Fragments Used in Experiment 1b

IPControl fragments used in experiment 1b (A) and the location of the fragments (B). The fragments are overlaid on a typical object from class A. The behavioral data from these fragments are shown in Figure 4C. See text for details.

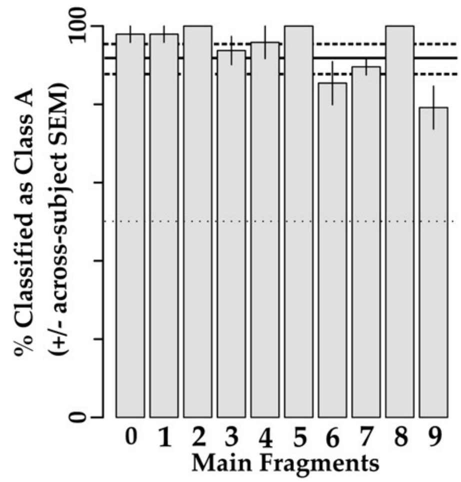

Figure S3. Performance with Main Fragments in Experiment 1b

Each bar shows the average percentage (±SEM) of trials in which the subjects classified a given fragment as belonging to class A. The thin dotted line denotes 50%, or chance level performance. The thick black lines in the background denote the mean (solid line) and the SEM (dashed lines) of the subjects with whole objects during the last two sessions of training.

Nonetheless, we examined the data for evidence of learning during the testing. Figure S6 shows the performance of the six subjects in the main task experiment 1a during the first and the last session of testing. The performance improved for no subject. Indeed, the performance showed a modest decrease overall, although the decrease was statistically insignificant (one-tailed Mann-Whitney test, $p > 0.05$). Performance with control and IPControl fragments, or the reaction times for all three fragment types, also showed no significant change during testing (not shown). Results from experiments 1b and 2 were qualitatively similar (not shown). Together, these results indicate that subjects had learned the informative fragments by the end of the training session, i.e., before the testing began.

The Classification Performance Is Highly Correlated with the Mutual Information of the Individual Fragments

To test the extent to which the classification performance is determined by the mutual information (MI) value of the main fragments, we isolated a different set of main fragments (not shown) with a range of low-to-high MI values (x axis). We then tested the categorization performance by using each of these fragments with the same training and testing procedure as above. This figure shows the average categorization performance (±SEM) of four subjects for each fragment. The thin dotted line denotes chance level performance. The performance was highly correlated with the MI values (correlation coefficient r, 0.86; $p < 0.05$), consistent with previous studies [S1, S2].

These results indicate that main fragments with high MI values can be expected to elicit correspondingly high performance. Therefore, the high performance elicited by the main fragments in experiments 1a, 1b, and 2 (Figures 4A, S4, and S5) is directly attributable to the fact these fragments had MI values at or near 1 (see Tables S1 and S2 below).

In some experimental contexts, performances at or near 100% are potentially problematic because they may reflect response saturation, thereby making it difficult to compare performances across the various conditions. However, high performance is not problematic in our context, in which the comparison of interest is between the main versus control fragments and not across the various main (or control) fragments. Indeed, high performance is advantageous in our context because the data from various main (or control) fragments amount to independent measurements of the corresponding category-learning effect.

Overlap among Fragments

In both experiments 1 and 2, the fragments overlapped with each other in some cases. For instance, the ten main fragments in experiment 1 occurred in four nonoverlapping clusters in two different regions of the embryo (top center and far right in Figure 2B). The largest of these clusters consisted of five fragments, 0, 1, 3, 4, and 6, with fragments with 1 and 6 mutually nonoverlapping. The second cluster consisted of fragment 2, which was close to, but did not overlap, the first cluster. The third cluster consisted of fragments 5, 8, and 9, and fourth cluster consisted of fragment 7 by itself. The ten control fragments in this experiment also showed comparable clustering (Figure 2D). We decided against excluding fragments on the sole basis of overlap, because they were judged

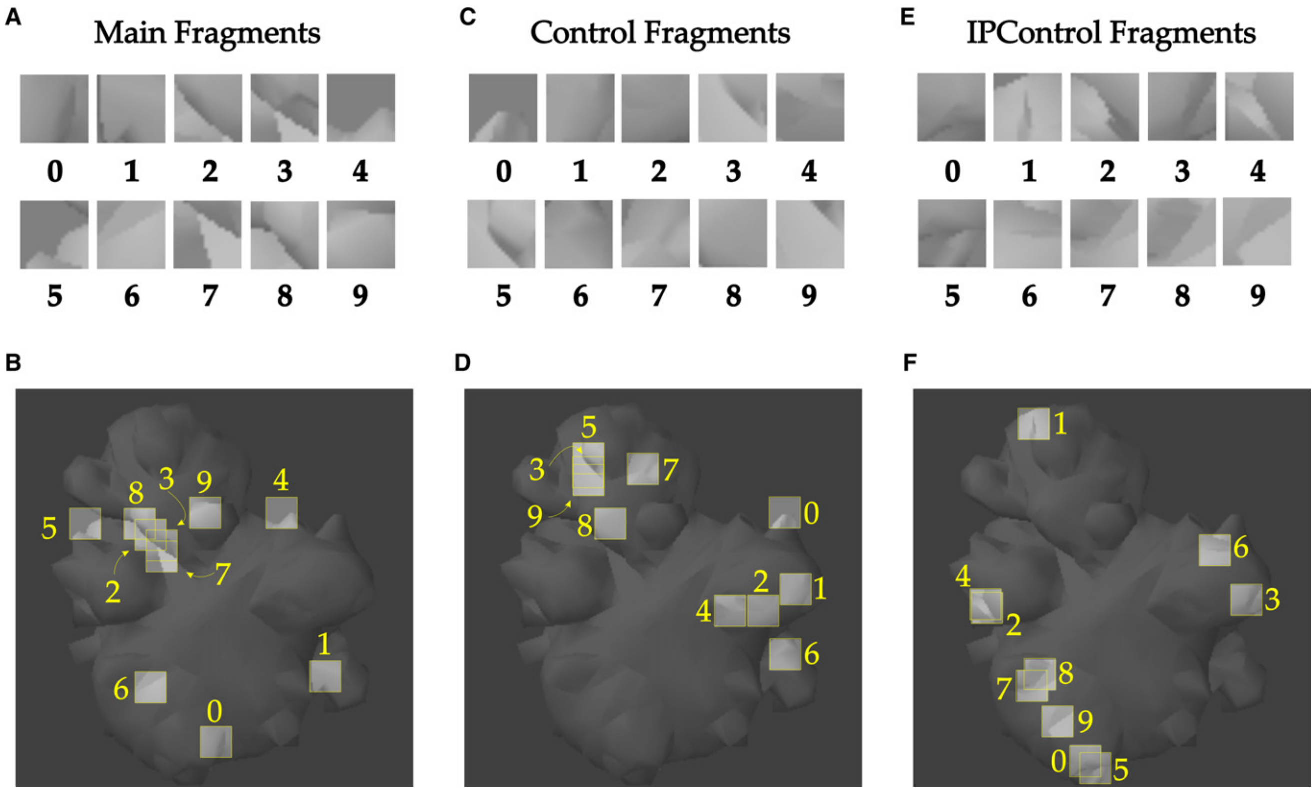

Figure S4. Fragments Used in Experiment 2

(A) Main fragments.
(B) Location of the main fragments, overlaid on a typical object from class C.
(C) Control fragments.
(D) Location of the control fragments.
(E) IPControl fragments.
(F) Location of the IPControl fragments. See text for details.

to be mutually dissimilar by an objective measure (see Supplemental Experimental Procedures). Moreover, our results hold even when only the data from nonoverlapping fragments are considered (see Figures 4A and 4B). This was also true for fragments from experiments 1b and 2 (see Figures 4C, S4, and S5; also see Figures S2 and S7).

Supplemental Experimental Procedures

Using VP to Create Naturalistic Object Classes

We created novel, naturalistic categories by using the VP algorithm (Figure 1), which simulates key processes of biological evolution. Broadly speaking, in case of evolution, biological object categories arise when heritable random variations are differentially passed on to the next generation

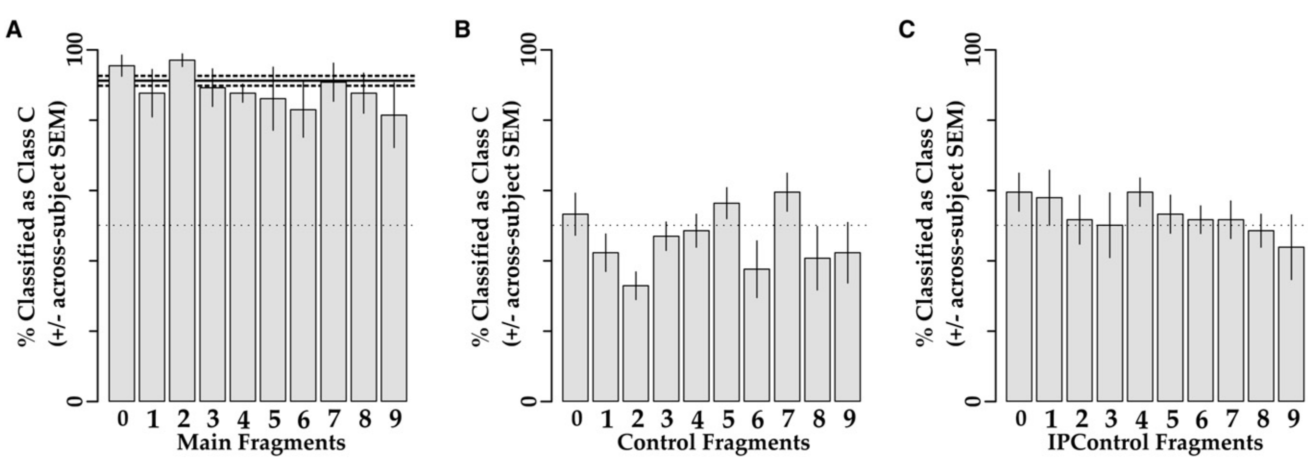

Figure S5. Performance in Experiment 2

In (A)–(C), each bar shows the average percentage (±SEM) of trials in which the subjects classified a given fragment as belonging to class C. The thin dotted line denotes 50%, or chance level performance. The thick black lines in the background in (A) denote the mean (solid line) and the SEM (dashed lines) of the subjects with whole objects during the last two sessions of training.

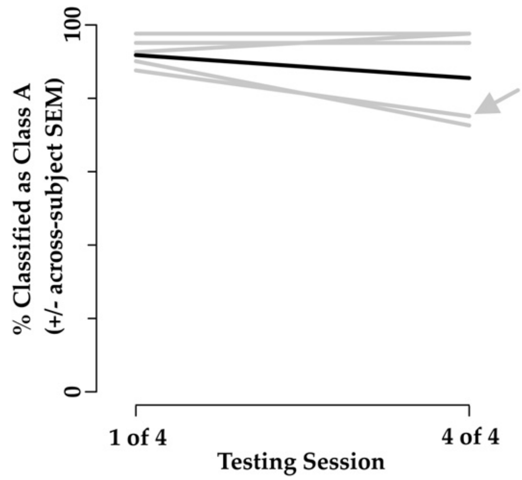

Figure S6. Performance in the Main Task over the Course of Testing in Experiment 1a

The performance of each subject (gray lines) during the first and the last (i.e., fourth) sessions of testing show no evidence of learning during testing. (The data from the intervening sessions are omitted for visual clarity.) The gray line denoted by the arrow represents overlapping data from two subjects. The thick black line denotes the average of all six subjects.

by processes such as natural selection, genetic drift, extinction, etc. (for a rigorous exposition, see [S3]). Equivalently for the present purposes, biological categories can also arise through externally imposed selection, such as in the breeding of farm animals and plants. In order to keep the origin of the categories as transparent as possible, the version of the VP algorithm used in this study only simulates the bare essentials of the phylogenetic process (see Discussion).

In the VP algorithm, shape variations among objects of a given generation arise randomly. All variations are heritable in principle in that each object starts as an exact replica of its parent and develops further on its own. Selection is externally imposed and consists of the fact that at each generation, only some of the objects are allowed to generate descendents. The children of a given parent constitute an object class (Figure 1A). We emphasize that the goal of the VP algorithm was not to develop a realistic simulation of evolution per se but rather to create naturalistic object categories by

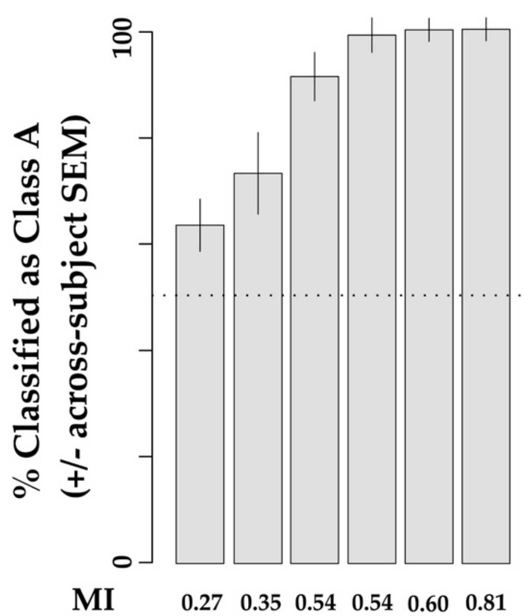

Figure S7. Categorization Performance as a Function of MI Values of the Main Fragments

Table S1. Mutual Information of Individual Fragments in Experiment 1

Fragment Type	# Belonged to Category	Categorization Task	MI
Main			
	0 A	A versus B (main task)	1.0
	1 A	A versus B (main task)	1.0
	2 A	A versus B (main task)	1.0
	3 A	A versus B (main task)	1.0
	4 A	A versus B (main task)	1.0
	5 A	A versus B (main task)	0.95
	6 A	A versus B (main task)	0.95
	7 A	A versus B (main task)	0.95
	8 A	A versus B (main task)	0.95
	9 A	A versus B (main task)	0.95
Control			
	0 A	A versus B (main task)	0
		A versus C (control task)	1.0
	1 A	A versus B (main task)	0
		A versus C (control task)	1.0
	2 A	A versus B (main task)	0
		A versus C (control task)	1.0
	3 A	A versus B (main task)	0
		A versus C (control task)	1.0
	4 A	A versus B (main task)	0
		A versus C (control task)	1.0
	5 A	A versus B (main task)	0.01
		A versus C (control task)	1.0
	6 A	A versus B (main task)	0
		A versus C (control task)	1.0
	7 A	A versus B (main task)	0.01
		A versus C (Control task)	1.0
	8 A	A versus B (main task)	0.03
		A versus C (control task)	1.0
	9 A	A versus B (main task)	0
		A versus C (control task)	1.0
IPControl			
	0 A	A versus B (main task)	0.05
	1 A	A versus B (main task)	0.05
	2 A	A versus B (main task)	0.05
	3 A	A versus B (main task)	0.05
	4 A	A versus B (main task)	0.06
	5 A	A versus B (main task)	0.06
	6 A	A versus B (main task)	0.06
	7 A	A versus B (main task)	0.07
	8 A	A versus B (main task)	0.07
	9 A	A versus B (main task)	0.07

simulation of morphological aspects of phylogenesis. For this reason, the VP algorithm bypasses many of the important complexities of biological evolution, such as the reshuffling of heritable characteristics through sexual reproduction and the fact that multicellular organisms typically develop from a single-cell embryo. Moreover, selection is imposed externally in our case. Nonetheless, it is worth noting that the categories arise naturally in VP, by means of selective propagation of heritable variations.

VP algorithm can, in principle, use any virtual object as a substrate. In the present study, we used a previously described type of naturalistic objects called digital embryos [S4]. In brief, the digital-embryo algorithm can create a virtually endless variety of naturalistic 3D shapes by simulating the natural processes of embryonic development, such as morphogen-mediated cell division, cell growth, and cell movement.

By using VP, we created three novel classes of digital embryo objects, classes A, B, and C, each containing ∼1500 objects. It is important to emphasize that the classes were generated without any regard to whether or how they could be classified and whether they contained any fragments useful for this classification.

We arbitrarily selected 200 embryos from each class for use in the experiments. Each 3D object was rendered without externally applied texture and with the same viewing and lighting parameters against a neutral gray background in the OpenGL graphics environment (www.opengl.org) with the software developed by Brady [S4] (also see http://www.psych.ndsu.

Table S2. Mutual Information of Individual Fragments in Experiment 2

Fragment Type	# Belonged to Category	Categorization Task	MI
Main			
	0 C	C versus A (Main task)	1.0
	1 C	C versus A (Main task)	1.0
	2 C	C versus A (Main task)	1.0
	3 C	C versus A (Main task)	1.0
	4 C	C versus A (Main task)	1.0
	5 C	C versus A (Main task)	1.0
	6 C	C versus A (Main task)	1.0
	7 C	C versus A (Main task)	1.0
	8 C	C versus A (Main task)	1.0
	9 C	C versus A (Main task)	1.0
Control			
	0 C	C versus A (Main task)	0.07
		C versus B (Control task)	0.8
	1 C	C versus A (Main task)	0.08
		C versus B (Control task)	0.71
	2 C	C versus A (Main task)	0.09
		C versus B (Control task)	0.7
	3 C	C versus A (Main task)	0.19
		C versus B (Control task)	0.95
	4 C	C versus A (Main task)	0.2
		C versus B (Control task)	0.71
	5 C	C versus A (Main task)	0.2
		C versus B (Control task)	0.71
	6 C	C versus A (Main task)	0.2
		C versus B (Control task)	0.78
	7 C	C versus A (Main task)	0.21
		C versus B (Control task)	0.86
	8 C	C versus A (Main task)	0.22
		C versus B (Control task)	0.71
	9 C	C versus A (Main task)	0.23
		C versus B (Control task)	0.74
IPControl			
	0 C	C versus A (Main task)	0.05
	1 C	C versus A (Main task)	0.05
	2 C	C versus A (Main task)	0.05
	3 C	C versus A (Main task)	0.05
	4 C	C versus A (Main task)	0.06
	5 C	C versus A (Main task)	0.06
	6 C	C versus A (Main task)	0.06
	7 C	C versus A (Main task)	0.06
	8 C	C versus A (Main task)	0.06
	9 C	C versus A (Main task)	0.06

nodak.edu/brady/downloads.html) and modified extensively by the authors. The images were stored as 8-bit, 256 × 256 pixel grayscale bitmaps.

Rationale for Using VP

The existing studies of informative fragments, although important, have some significant limitations, all related to how informative fragments operate. By their very nature, informative fragments are parts of specific exemplar objects of a category, and not generic prototypes [S5, S6]. That is, a given fragment containing an eye is not a general model of what an eye "looks like," but is extracted from a specific bitmap of a specific face. This property of informative fragments has important computational advantages [S5, S6]. But it also means that, for a familiar category, the object exemplars from which the fragments are isolated computationally are not the same as those from which the categories were first learned by the subjects. Thus, fragments from familiar categories do not address the issue of how categorization (i.e., assigning an object to a familiar category) is related to category learning (i.e., acquiring a previously unfamiliar category). For instance, do we learn informative fragments during category learning, i.e., is fragment learning a part of category learning? Do we learn fragments only when it is necessary to do so (e.g., when the object of interest is partially occluded) or incidentally as a part of category learning?

Object categories that address the aforementioned issues must meet the following four criteria: First, the categories must be new to the subject, so that they need to be learned. Second, the new categories must be sufficiently different from the familiar categories, so that subjects cannot learn the new categories as variations of the familiar ones (e.g., SUVs as variants of cars). Third, to ensure that behavioral and computational results can be directly compared with each other, the fragments should be extracted from the same images as those used by subjects for category learning. This makes using highly familiar categories, such as faces or cars, undesirable because subjects are often exposed to uncontrolled instances of these categories in everyday life. Finally, the categories should still capture regularities inherent in natural object classes so as to approach conditions representative of natural category learning. No currently available method of creating object categories meets all of these criteria, whereas the VP algorithm meets them all.

Experiments

Two independent experiments were carried out with the same set of three classes. The two experiments differed in which category was distinguished from which (A versus C or B versus C, see below). Each experiment consisted of isolating the fragments, training the subjects, and subsequently testing them, all with the same given set of objects.

Extracting Fragments for Experiment 1

For this experiment, the "Main" task was defined as distinguishing objects of class A from objects of class B. Ten informative fragments supporting the main task were isolated ("Main" fragments). Each main fragment was a small 20 × 20 pixel ($0.53^\circ \times 0.53^\circ$) subimage of a class A object. We used small fragments because larger fragments were found to contain smaller informative subfragments, even when the fragment on the whole was uninformative. This is undesirable because subjects can potentially restrict their attention just to the informative subpart of an uninformative fragment.

The fragments were selected, on the basis of their MI, out of many candidate fragments. All 20 × 20 pixel fragments on a dense grid (with step size of 7 pixels) were considered. This resulted in more than 500 candidate fragments per image, or a total of about 115,000 candidate fragments for the 200 images. MI of each fragment for the main task was calculated. The fragment with the highest MI was selected, and the set of candidate fragments was pruned on the basis of visual similarity (see below) to this selected fragment. The process was repeated until a total of ten main fragments (Figure 2A) were selected.

Visual similarity was evaluated with the correlation coefficient of pixel values. To detect small overlaps between fragments, we also allowed the correlated fragments to move with respect to one another. Candidate fragments with visual similarity greater than 0.8 were considered too similar to a selected fragment and were removed. This constraint reduced shape redundancy across the selected fragments.

Main fragments are useful for performing the main task. Therefore, we expect human subjects to preferentially use these fragments during this task. For assessment of the degree of this preference, noninformative fragments need to be selected as a basis for comparison.

A naive approach would be to select fragments as above but with minimal, rather than maximal, MI. A disadvantage of this approach is that it tends to select visually uninteresting fragments. For example, image patches that are uniform or almost uniform in intensity have very low MI, so that several of these would typically be selected by the naive approach. Such fragments would indeed be uninformative, but for a trivial reason. So that the comparison is fair, it is desirable to avoid selecting such fragments.

We introduce two principled methods of selecting interesting but uninformative fragments for comparison. First, we introduce a "Control" task, which is to discriminate class A from class C. Ten fragments that are uninformative for the main task were selected, subject to the constraint that they have high MI for the control task ("Control" fragments). As before, these were selected from a pool of candidate fragments—all 20 × 20 pixel fragments of a class A object on a dense grid. First, all candidate fragments with MI for the control task less than 0.7 were removed (recall that the MI can vary between 0 and 1 in our case). Next, fragments uninformative for the main task were selected with the procedure described above, but fragments with minimal (rather than maximal) MI were chosen. The intuition behind this method is that visually uninteresting fragments are expected to be uninformative for any task. For example, the uniformly gray patches from the background provide information for neither the main nor the control task. the constraint of having high control task MI therefore ruled out such patches. Indeed, the resulting control fragments (Figure 2C) have significant visual content.

We also isolated ten additional fragments by using an interest-point detector ("IPControl" fragments). Interest-point detectors select areas of an image that have significant visual content, such as corners or intersections [S7] or high entropy [S8]. Such detectors are heavily used in computer vision (for a review, see [S9]). In our experiments, we used the popular Harris interest-point detector [S10, S11]. First, we detect all interest points in an image (typically, there are 300–600 per image). Because these points are by definition visually interesting, we then simply proceed to select ten fragments with low MI for the main task (as before, subject to the constraint of being visually dissimilar to one another) (see Figure S2).

Compared to control fragments, IPControl fragments explore the set of uninformative fragments more fully because the criterion for selection is based more directly on local visual content. By contrast, control fragments are constrained to be informative for an auxiliary task (the control task), and this criterion will certainly miss those visually interesting fragments that happen to be uninformative for the control task. On the other hand, the IPControl fragments may be uninformative for a trivial reason. Interest-point detector rules out the most trivial cases (such as patches of uniform intensity) but may still pass other uninteresting content (for example, a patch containing high-spatial-frequency random noise). Control fragments do not run that risk because they are guaranteed to be informative for some other task (the control task) and therefore are useful for categorization.

To summarize, we selected a total of 30 fragments for experiment 1. All of these are subimages of the main class objects. Out of these fragments, ten are informative for the main task, and 20 are uninformative.

Extracting Fragments for Experiment 2

The goal of our experiments was to determine whether human subjects learn to use informative fragments in categorization. However, experiment 1 described above only involves a single categorization task (the main task). To ensure the results are not specific to this particular set of categories, it is desirable to evaluate performance on a different set of categories. In experiment 2, we used the same three object classes (A, B, and C) but redefined their roles. To this end, we designated the main task as distinguishing objects of class C from objects of class A, and the control task was designated as distinguishing class C from class B. We then selected 30 additional fragments with the procedure described above, but with the new class designations.

Training in the Categorization Task

Subjects

All psychophysical procedures used in this study were reviewed and approved in advance by the University of Minnesota Institutional Review Board. Ten healthy adult volunteers that had normal (or corrected-to-normal) vision participated in this study. All subjects provided informed consent prior to the study and were compensated for their participation. Six subjects (three females) participated in experiment 1, and four different subjects (three females) participated in experiment 2.

Training Paradigm

Subjects in a given experiment were trained in the main task appropriate for that experiment. Subjects received no training in the control task and were not aware of existence of a third class (class C in experiment 1, class A in experiment 2). The reason is that all fragments used in the experiments were evaluated only with respect to MI in the main task, whereas the control task played only an auxiliary role.

During each training trial for experiment 1, two sample objects and a test object ($6.7^\circ \times 6.7^\circ$ each) were presented simultaneously 9° (center-to-center distance) apart. One of the sample objects was drawn randomly from class A, and the other was drawn randomly from class B. The class membership of the sample objects was indicated on the subject's screen, and the relative locations of the objects from the two classes were randomly switched across trials. Depending on the trial, the test object was drawn either from class A or from class B but was never the same object as either of the sample objects in a given trial. By using a key press, the subject had to classify the test object into class A or class B on the basis of the sample objects. After the subject made his/her report, the correct classification was shown on screen, so that the subject could to re-examine the three objects in light of the feedback. The subject was allowed unlimited time both to make the initial report and to review the subsequent feedback, so as to approximate natural viewing conditions as closely as practicable. The subject used another key press to proceed to the next trial. A given subject was considered trained if he or she performed significantly above 75% accuracy (i.e., at $p < 0.002$ by binomial test) for at least two consecutive blocks of 40 trials each. Subjects trained for a median of eight blocks (i.e., a total of 320 trials) before reaching this asymptotic level of performance. Because there were 200 embryos in each class (see above), this means that during the training phase, the subject saw each given embryo and average of 1.6×.

The training procedure for experiment 2 was identical to that for experiment 1, except that the class designations were different, as described above.

Testing the Fragments

During the testing phase, the subjects performed the classification task on the sole basis of a given fragment (Figure 3; see below). The subjects were not told anything about the fragments, except that they were derived from the type of objects they had seen during the training phase.

During the testing phase of experiment 1, we generated the test object by compositing the fragment of interest on an object drawn randomly from class A or class B (i.e., by graphically overlaying the given fragment over the given background object). The composite object was shown to the subject behind a rectangular translucent occluder with a hole, so that only the fragment ($0.53^\circ \times 0.53^\circ$) was visible through the hole, unhindered in its proper position on the object, whereas the rest of the object appeared as a faded "background" (see Figure 3). This design helped ensure that the subjects saw the fragment in its proper context. This is more advantageous in our context than presenting a given fragment by itself without the context because it minimizes the possibility that subject may have to use task-irrelevant semantic and spatial (e.g., configural) cues (e.g., left eye) to help perform the task.

Two sample objects, one drawn from each class, were shown on either side of the test object as during the training phase, although the class membership was not indicated for any of the three objects. The sample objects were provided to help ensure that (1) the task tested object categorization and not fragment categorization and (2) task required only implicit perceptual learning and not declarative (or explicit) association between a fragment with a category.

We confirmed that the subject could not use the sample objects to do a simple pixel-wise comparison between the fragment and the relevant regions of the sample objects, since the subjects were unable to perform the task with the same testing paradigm without first learning the correct categories (see Figures 4D and 4E). For a demonstration of this effect, the reader should choose a fragment of interest in Figure S1 and try categorizing it by comparing it to a whole object each from class A and class B.

Subjects had to classify, by using a key press, the test object into the class represented by either sample object on the sole basis of the given fragment of the test object. Subjects were told that the faded background portion of the test object (i.e., the portion visible behind the translucent occluder) was randomly drawn, so that they would not be able to perform the task above chance levels with the background object. No feedback was provided. To help ensure that the testing conditions reflected categorization under natural conditions as closely as possible, we allowed subjects free eye movements and unlimited time to make their responses. The average response time of the subjects was 5.30 s ± 0.16 SEM (not shown) and was indistinguishable from the corresponding response times during the last two blocks of the training phase (ANOVA, unbalanced design; $p > 0.05$).

The trials for the various main and control fragments were randomly interleaved. For each fragment, the performance of each subject was measured over a total of 16 trials spread over four sessions of four trials each.

Testing for experiment 1 was carried out in two stages. During the first stage (experiment 1a), the main and the control fragments were tested with randomly interleaved trials for all six subjects in this experiment. During the second stage (experiment 1b), the main and the IPControl fragments were similarly tested for three of the six subjects.

The testing procedure for experiment 2 was identical to that for experiment 1, with two exceptions. First, the class designations were different, as described above. Second, main, control, and IPControl fragments were all tested together with randomly interleaved trials.

Supplemental References

S1. Harel, A., Ullman, S., Epshtein, B., and Bentin, S. (2007). Mutual information of image fragments predicts categorization in humans: Electrophysiological and behavioral evidence. Vision Res. *47*, 2010–2020.

S2. Lerner, Y., Epshtein, B., Ullman, U., and Malach, R. (2007). Class information predicts activation by object fragments in human object areas. J. Cogn. Neurosci. *20*, 1–28.

S3. Futuyma, D.J. (2005). Evolution (Sunderland, MA: Sinauer Associates).

S4. Brady, M.J., and Kersten, D. (2003). Bootstrapped learning of novel objects. J. Vis. *3*, 413–422.

S5. Ullman, S., Vidal-Naquet, M., and Sali, E. (2002). Visual features of intermediate complexity and their use in classification. Nat. Neurosci. *5*, 682–687.

S6. Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. Trends Cogn. Sci. *11*, 58–64.

S7. Harris, C.G., and Stephens, M.J. (1988). A combined corner and edge detector. In Proc. Fourth. Alvey Vision Conf. pp. 147–151.

S8. Kadir, T., and Brady, M. (2001). Scale, saliency and image description. Int. J. Comput. Vis. *45*, 83–105.

S9. Mikolajczyk, K., and Tuytelaars, T. (2005). A comparison of affine region detectors. Int. J. Comput. Vis. *65*, 43–72.

S10. Mikolajczyk, K., and Schmid, C. (2002). An affine invariant interest point detector. Proc. European Conf. Computer Vision *2350*, 128–142.

S11. Duda, R.O., Hart, P.E., and Stork, D.G. (2001). Pattern classification, Second Edition (New York: Wiley Interscience).

7. Rideout, E.J., Billeter, J.C., and Goodwin, S.F. (2007). The sex-determination genes *fruitless* and *doublesex* specify a neural substrate required for courtship song. Curr. Biol. *17*, 1473–1478.
8. Lee, T., and Luo, L. (2001). Mosaic analysis with a repressible cell marker (MARCM) for *Drosophila* neural development. Trends Neurosci. *24*, 251–254.
9. Wong, A.M., Wang, J.W., and Axel, R. (2002). Spatial representation of the glomerular map in the *Drosophila* protocerebrum. Cell *109*, 229–241.
10. Datta, S.R., Vasconcelos, M.L., Ruta, V., Luo, S., Wong, A., Demir, E., Flores, J., Balonze, K., Dickson, B.J., and Axel, R. (2008). The *Drosophila* pheromone cVA activates a sexually dimorphic neural circuit. Nature *452*, 473–477.
11. Patterson, G.H., and Lippincott-Schwartz, J. (2002). A photoactivatable GFP for selective photolabelling of proteins and cells. Science *297*, 1873–1877.
12. Ferveur, J.F. (2005). Cuticular hydrocarbons: their evolution and roles in *Drosophila* pheromonal communication. Behav. Genet. *35*, 279–295.
13. von Schilcher, F. (1976). The role of auditory stimuli in the courtship of *Drosophila melanogaster*. Anim. Behav. *24*, 18–26.
14. Kurtovic, A., Widmer, A., and Dickson, B.J. (2007). A single class of olfactory neurons mediates behavioural responses to a *Drosophila* sex pheromone. Nature *446*, 542–546.
15. Kondoh, Y., Kaneshiro, K.Y., Kimura, K., and Yamamoto, D. (2003). Evolution of sexual dimorphism in the olfactory brain of Hawaiian *Drosophila*. Proc. Biol. Sci. *270*, 1005–1013.
16. Jefferis, G.S., Potter, C.J., Chan, A.M., Marin, E.C., Rohlfing, T., Maurer, C.R. Jr., and Luo, L. (2007). Comprehensive maps of *Drosophila* higher olfactory centers: spatially segregated fruit and pheromone representation. Cell *128*, 1187–1203.
17. Technau, G.M. (1984). Fiber number in the mushroom bodies of adult *Drosophila melanogaster* depends on age, sex and experience. J. Neurogenet. *1*, 113–126.

University of Glasgow, IBLS-Division of Molecular Genetics, Anderson College, 56 Dumbarton Road, Glasgow G11 6NU, UK.
*E-mail: s.goodwin@bio.gla.ac.uk

DOI: 10.1016/j.cub.2008.03.035

Visual Categorization: When Categories Fall to Pieces

We cannot help but categorize the visual world into objects like cats and faces. An intriguing new study shows that observers automatically discover informative fragments of visual objects during category learning.

Quoc C. Vuong

We see the world in discrete categories in order to recognize and interact appropriately with objects in our environment [1]. How do we learn visual object categories? Our intuition suggests that, through experience, we acquire features found in members of one category but not in those from another category. For example, cats have whiskers; human faces, on the other hand, normally do not. There is empirical support for this intuitive view [2,3].

But a fundamental problem with this intuition is image variability. Familiar objects from the same category can have an enormous range of appearance; they are often occluded by other objects; how they appear to us can further be confounded by viewing conditions such as variable illumination; and so on [2]. These factors converge to make it extremely difficult to learn generic features that are reliable for visual categorization.

In work published recently in *Current Biology*, Hegdé *et al.* [4] offer a compelling solution to this problem, but one that highlights the need for us to re-think the pieces that make up objects and object categories. Armed with a set of novel visual categories [5] and a statistical means to select features [6,7], these authors have demonstrated that observers automatically discover fragments — literally, bits and pieces of images — during category learning that are very effective for visual categorization. This provides a new and important link between visual category learning and visual categorization.

In this new study [4], observers classified a large number of unfamiliar objects into two categories. The objects were synthesized from a novel virtual phylogenesis algorithm which simulated the evolution of biological forms [5], so that category members captured natural variations of categories we are more familiar with. The examples in Figure 1 show that this classification task is far from trivial, even with whole objects (see supplemental Figure S1 in the paper for more examples).

Two main sets of image fragments were extracted from trained objects using the same statistical procedure. Observers then classified all fragments, just as they had done with whole objects. This sounds like an even more daunting task. Amazingly though, observers were as accurate with one set of fragments as they were

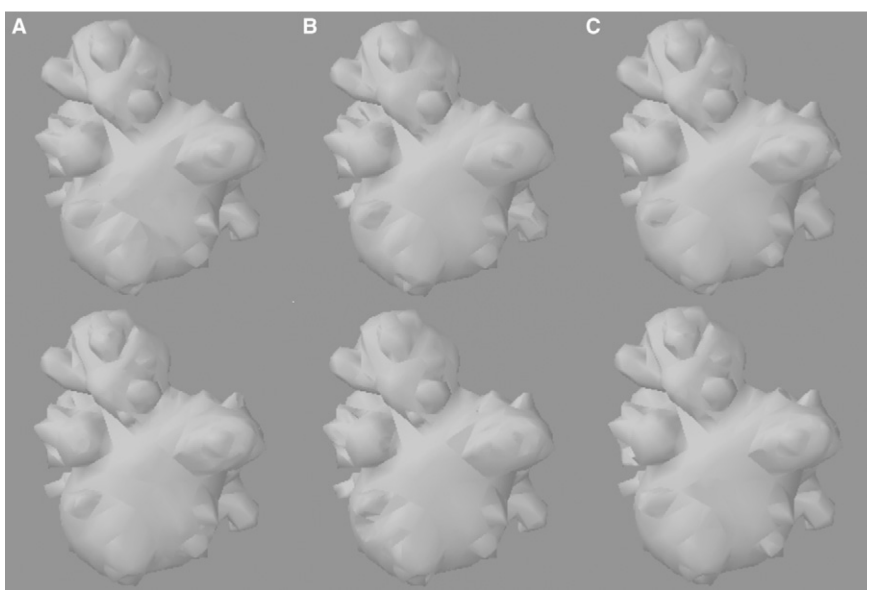

Figure 1. Example objects synthesized by virtual phylogenesis.
Observers were only trained on objects from two of the three categories A, B and C (from [4]).

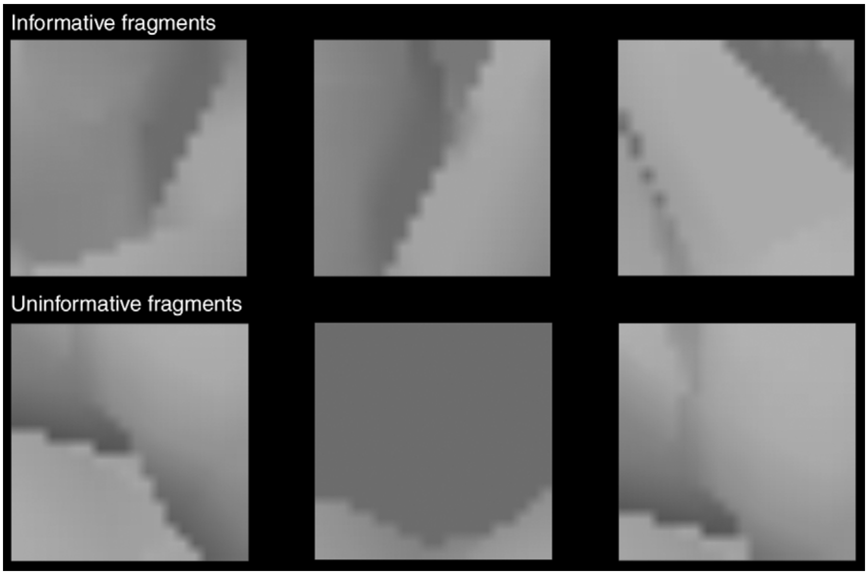

Figure 2. Examples of informative and uninformative fragments.

The informative ones distinguished between two trained categories — say categories A and B from Figure 1 — whereas the uninformative ones distinguished between a trained (A) and untrained category (C).

with whole objects — nearly 100% correct! Surprisingly, the same observers struggled to perform above chance with the other set of fragments.

Figure 2 shows the remarkable lack of visual difference between informative and uninformative fragment sets. Clearly, observers do not acquire just any fragment during category learning. But what distinguishes informative fragments where performance is nearly at ceiling, from uninformative ones where performance is more or less at chance?

The answer lies in how the fragments are extracted from images. Following recent computational advances, Hegdé *et al.* [4] selected fragments which maximized their ability to distinguish categories, through a powerful statistic called mutual information [6,7]. This measure tells us how certain we can be about the presence of a category if a specific fragment is present in the image. For example, if a human eye is present in the image, then there is a good chance that a human face is also present in that image.

The informative fragments used by Hegdé *et al.* [4] distinguished between two trained categories (say categories A and B; see Figure 1). By comparison, the uninformative fragments distinguished between a trained and untrained category (say categories A and C). Observers never saw untrained category members.

Thus, both informative and uninformative fragments were of comparable visual complexity and both contained diagnostic information to distinguish categories, but only the informative ones were relevant for the observers' task. In fact, observers could not classify informative fragments prior to any training, which underscores the importance of category learning to discover the right pieces for the task.

There are thousands of possible fragments of an image, but only a fraction of them will reliably indicate that a particular category is present. Feature selection based on mutual information is a powerful framework to extract those fragments [6,7]. These are typically of intermediate complexity [6], balancing how likely the fragments will occur in an image and how indicative they are of a particular category.

For example, a fragment containing the eyes and a bit of the nose probably indicates that a face is present in the image but it is very unlikely to find such a large fragment in many different images. Conversely, a much smaller fragment containing just the hair line (so it looks like an edge) is likely to occur in many different images which do contain faces but may accidentally occur in images which do not contain faces.

This framework is very successful for familiar visual categories [7]. For example, the mutual information of familiar object fragments correlates with neural measures like visual evoked potentials [8] and haemodynamic brain responses [9]. So there is exciting new evidence that the brain may also extract fragments of intermediate complexity for everyday things.

One concern with using familiar objects is whether observers learn fragments out of necessity, as objects are often occluded, or whether fragment-based learning occurs automatically as a matter of course. Hegdé *et al.*'s [4] results clearly favour the latter, as observers learn novel whole objects. There was no need for them to discover fragments during training, but they did.

There is something to be said about Hegdé *et al.*'s [4] virtual phylogenesis algorithm for synthesizing objects. Like biological organisms, their objects evolve from a common ancestor. Objects from the same category inherit their common ancestor's shape characteristics but express individual shape variations. Indeed, this algorithm has a nice parallel to earlier work with an artificial taxonomy of 'caminalcules' used to study how taxonomists classify the evolutionary relationships between species [10].

Virtual phylogenesis gives rise to novel object categories with desirable properties: for example, objects have measurable natural within-class variations similar to biological organisms. It is also versatile: for example, objects can be structured into a hierarchy of categories, or other evolutionary mechanisms (such as sexual selection) can be incorporated into the algorithm. Importantly, it is a principled means to synthesize a large number of naturalistic objects without unknowingly pre-specifying the informative fragments studied.

The algorithm diverges from alternative methods of synthesizing novel objects, such as combining shape primitives [11,12] or clustering shapes on the basis of similarity [13]. Given its versatility, virtual phylogenesis is a significant addition to the repertoire of techniques for synthesizing objects that can be used for natural vision, machine learning, and even evolutionary taxonomy.

Hegdé *et al.*'s [4] findings provide strong support for a prominent computational model of object perception and categorization based on informative image fragments [6,7]. They also support observers' natural tendency to pick up statistical regularities in the visual input [14], which can develop as early as nine months [15]. Lastly, the results link visual category learning with visual categorization, in that informative fragments play a key role for both processes [4,6–9].

Category learning remains an important issue in visual cognition. There are ecological reasons for acquiring pieces of visual categories [1]; for example, to overcome very real problems like occlusions and image variability. The human visual system has evolved to automatically acquire informative fragments for visual categorization. Let's hope that we will likewise pick up the pieces.

References

1. Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., and Boyes-Braem, P. (1976). Basic objects in natural categories. Cogn. Psychol. *8*, 382–439.
2. Tarr, M.J., and Vuong, Q.C. (2002). Visual object recognition. In Steven's Handbook of Experimental Psychology (III edition), Volume 1: Sensation and Perception, H. Pashler and S. Yantis, eds. (New York: John Wiley and Sons, Inc.), pp. 287–314.
3. Palmeri, T.J., and Gauthier, I. (2004). Visual object understanding. Nat. Rev. Neurosci. *5*, 291–303.
4. Hegdé, J., Bart, E., and Kersten, D. (2008). Fragment-based learning of visual object categories. Curr. Biol. *18*, 597–601.
5. Brady, M.J., and Kersten, D. (2003). Bootstrapped learning of novel objects. J. Vision *3*, 413–422.
6. Ullman, S., Vidal-Naquet, M., and Sali, E. (2002). Visual features of intermediate complexity and their use in classification. Nat. Neurosci. *5*, 682–687.
7. Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. Trends Cogn. Sci. *11*, 58–64.
8. Harel, A., Ullman, S., Epshtein, B., and Bentin, S. (2007). Mutual information of image fragments predicts categorization in humans: Electrophysiological and behavioural evidence. Vision Res. *47*, 2010–2020.
9. Lerner, Y., Epshtein, B., Ullman, U., and Malach, R. (2008). Class information predicts activation by object fragments in human object areas. J. Cogn. Neurosci., in press.
10. Sokal, R.R. (1983). A phylogenetic analysis of Caminalcules. I. The database. Systematic Zool. *32*, 185–201.
11. Gauthier, I., Williams, P., Tarr, M.J., and Tanaka, J. (1998). Training "greeble" experts: A framework for studying expert object recognition processes. Vision Res. *38*, 2401–2428.
12. Biederman, I., and Gerhardstein, P.C. (1993). Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. J. Exp. Psychol. Hum. Percept. Perform. *19*, 1162–1182.
13. Edelman, S., and Duvdevani-Bar, S. (1997). A model of visual recognition and categorization. Philos. Trans. R. Soc. Lond. B *352*, 1191–1202.
14. Kersten, D., Mamassian, P., and Yuille, A. (2004). Object perception as Bayesian inference. Annu. Rev. Psychol. *55*, 271–304.
15. Fiser, J., and Aslin, R.N. (2002). Statistical learning of new visual feature combinations by infants. Proc. Natl. Acad. Sci. USA *99*, 15822–15826.

Institute of Neuroscience, Newcastle University, Newcastle upon Tyne NE2 4HH, UK.
E-mail: q.c.vuong@ncl.ac.uk

DOI: 10.1016/j.cub.2008.03.036

Cellular Evolution: What's in a Mitochondrion?

Mitochondria and their relatives constitute a wide range of organelles, only some of which function in aerobic respiration. Mitochondrial remnants from different anaerobic lineages show a striking degree of functional convergence.

Christopher J. Howe

For many years, the view was widely held that mitochondria originated when a primitive eukaryotic cell acquired through endosymbiosis a prokaryote capable of oxidative phosphorylation. Some of the endosymbiont's genes were lost, some were transferred to the nucleus, and a stable relationship was established that has lasted very successfully for well over a billion years. The fact that anaerobic eukaryotic lineages exist today — such as the gut-dwelling pathogen *Giardia* — was attractively consistent with this view of mitochondrial origin. These anaerobic eukaryotes appeared to lack mitochondria and according to molecular phylogenetic trees seemed to have diverged from other eukaryotes very early — presumably before the acquisition of mitochondria. This group became known as the Archezoa [1]. However, a discovery that would ultimately be crucial to the demise of the Archezoan concept had been made back in 1973 with the description of hydrogenosomes in anaerobic trichomonads [2]. Hydrogenosomes are now recognised as derived from mitochondria. They produce hydrogen and ATP and have been found in a range of anaerobic or almost anaerobic eukaryotes. Writing in *Current Biology*, Stechmann *et al.* [3] have now described another example of a mitochondria derived organelle that sheds light on their evolutionary fate.

A second development leading to the demise of the Archezoan concept was the recognition that the anaerobic, amitochondriate eukaryote *Entamoeba histolytica* contains nuclear genes for the mitochondrial proteins pyridine nucleotide transhydrogenase and the chaperonin cpn60 [4]. This discovery indicated that this supposedly amitochondriate organism had possessed mitochondria in the past and might even have retained a remnant of the organelle. Although the placement of *Entamoeba* among the Archezoa was controversial, other members of the Archezoa were soon shown also to harbour genes for proteins of mitochondrial origin and remnant mitochondrial compartments [5]. We now recognize that all eukaryotes probably have mitochondria, or their remnants, and indeed it arguably was the acquisition of the mitochondrion that marked the birth of the eukaryotes [6]. Furthermore, the phylogenetic position of Archezoa as early-diverging eukaryotes is also questionable [7,8].

A Diversity of Mitochondrial Forms

Mitochondrial remnants are known as hydrogenosomes or mitosomes, depending on their function. In general, organelles derived from mitochondria can be ordered on a spectrum based on their structure and function (Figure 1). Classical mitochondria, with their cristae as well as their electron transfer chain and F_1F_0 ATPase for oxidative phosphorylation in aerobic conditions, represent one end of the spectrum. Close to these are the mitochondria of some anaerobic metazoa, such as those of parasitic worms, which lack some components of the electron transfer chain [9].