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.

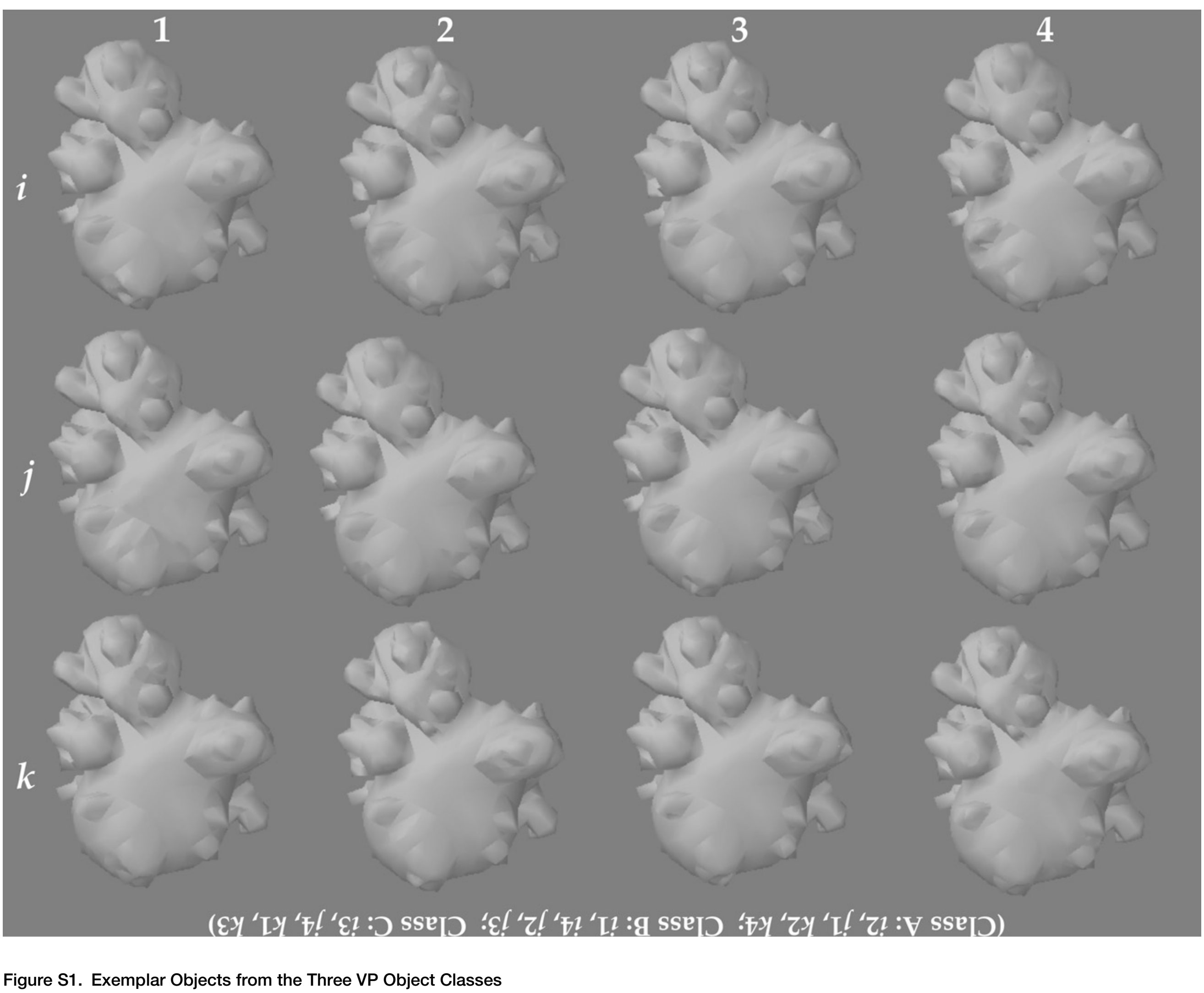

Figure S1. Exemplar Objects from the Three VP Object Classes

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.

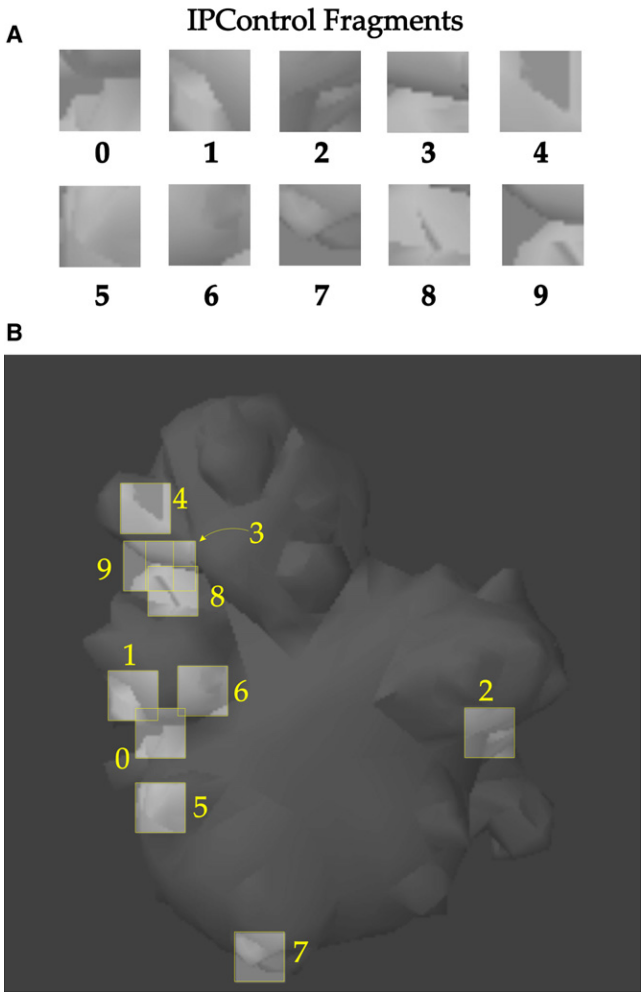

Figure S2. IPControl Fragments Used in Experiment 1b

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.

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].

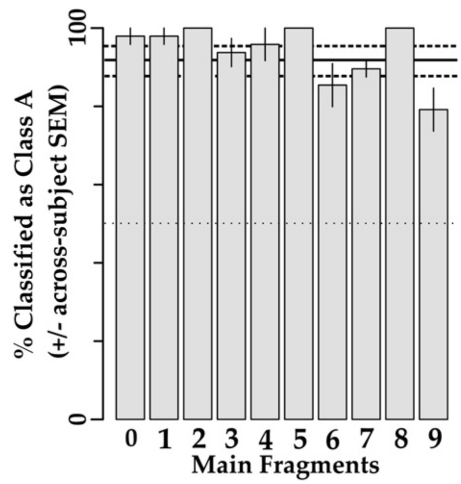

Figure S3. Performance with Main Fragments in Experiment 1b

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.

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

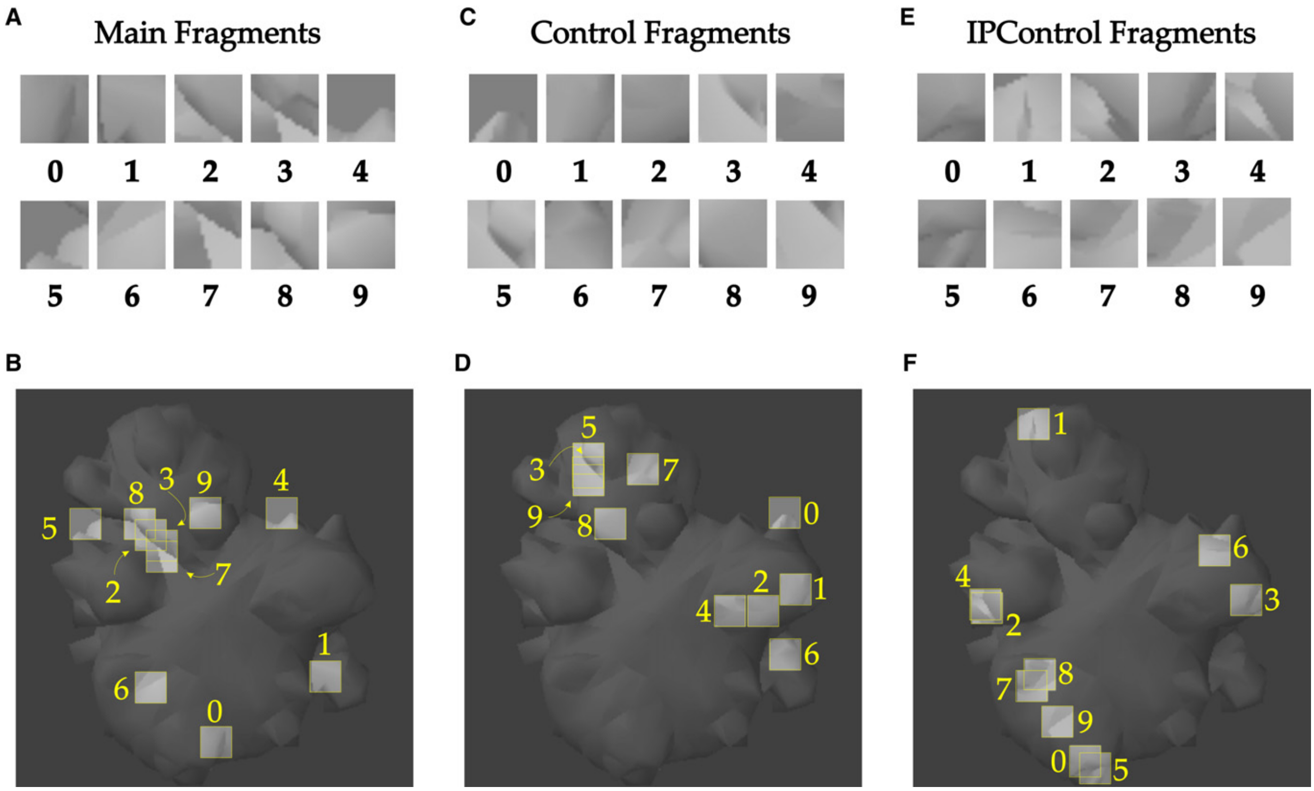

Figure S4. Fragments Used in Experiment 2

(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

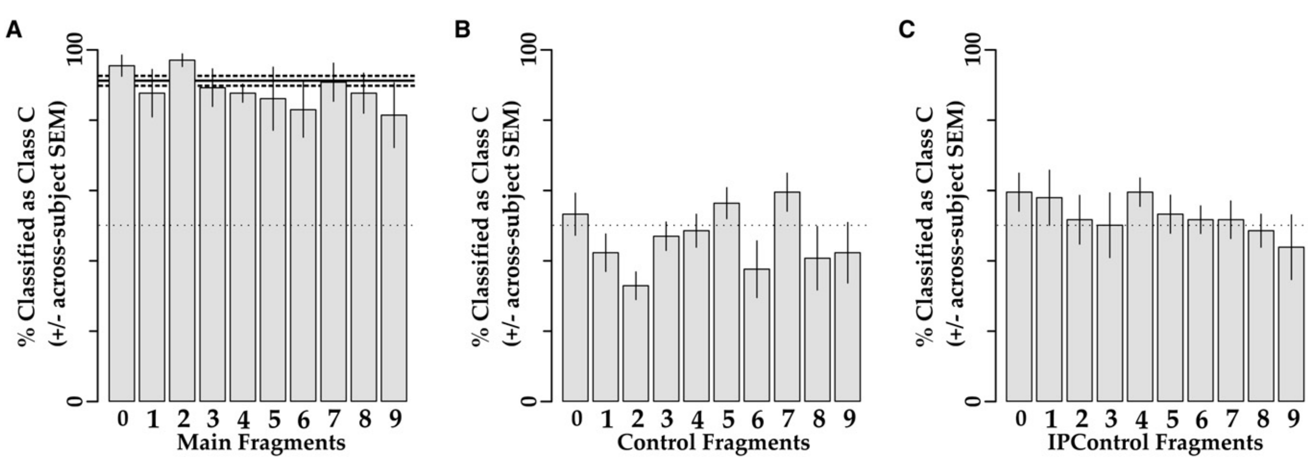

Figure S5. Performance in Experiment 2

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.

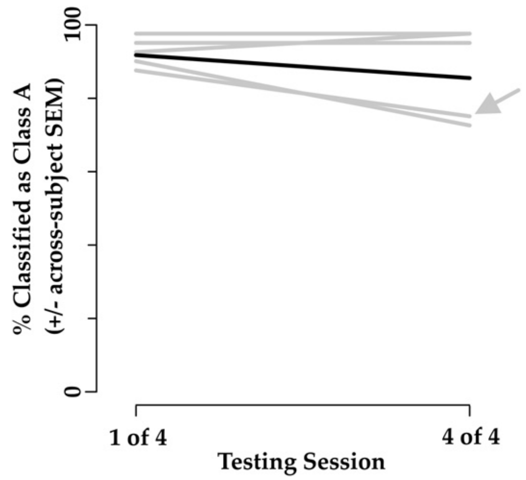

Figure S6. Performance in the Main Task over the Course of Testing in Experiment 1a

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

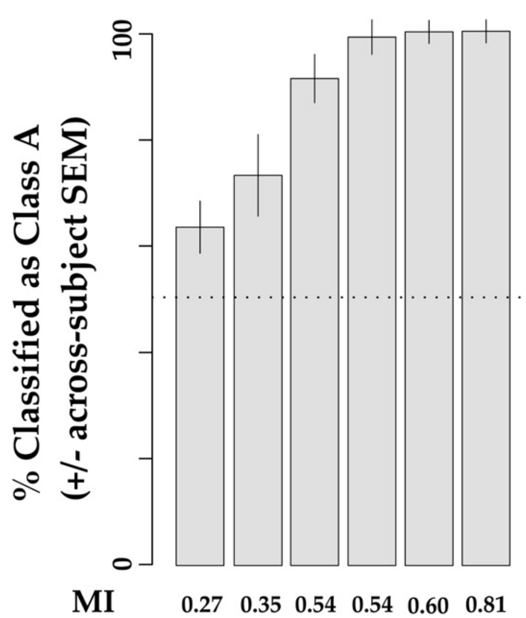

Figure S7. Categorization Performance as a Function of MI Values of the Main Fragments

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).