

Similarity relations in visual search predict rapid visual categorization

Krithika Mohan

Indian Institute of Science Education and Research,
Pune, India
Centre for Neuroscience, Indian Institute of Science,
Bangalore, India



S. P. Arun

Centre for Neuroscience, Indian Institute of Science,
Bangalore, India



How do we perform rapid visual categorization? It is widely thought that categorization involves evaluating the similarity of an object to other category items, but the underlying features and similarity relations remain unknown. Here, we hypothesized that categorization performance is based on perceived similarity relations between items within and outside the category. To this end, we measured the categorization performance of human subjects on three diverse visual categories (animals, vehicles, and tools) and across three hierarchical levels (superordinate, basic, and subordinate levels among animals). For the same subjects, we measured their perceived pair-wise similarities between objects using a visual search task. Regardless of category and hierarchical level, we found that the time taken to categorize an object could be predicted using its similarity to members within and outside its category. We were able to account for several classic categorization phenomena, such as (a) the longer times required to reject category membership; (b) the longer times to categorize atypical objects; and (c) differences in performance across tasks and across hierarchical levels. These categorization times were also accounted for by a model that extracts coarse structure from an image. The striking agreement observed between categorization and visual search suggests that these two disparate tasks depend on a shared coarse object representation.

Keywords: visual search, shape perception, object recognition

Citation: Mohan, K., & Arun, S. P. (2012). Similarity relations in visual search predict rapid visual categorization. *Journal of Vision*, 12(11):19, 1–24, <http://www.journalofvision.org/content/12/11/19>, doi:10.1167/12.11.19.

Introduction

Categorization is a fundamental cognitive process that involves evaluating the similarity of an object with other category members (Goldstone, 1994; Margolis & Laurence, 1999; Smith, Patalano, & Jonides, 1998). What are the underlying features and similarity computations? One influential approach has been to study categorization of visual objects varying along prespecified features (Freedman, Riesenhuber, Poggio, & Miller, 2001; Maddox, Ashby, & Gottlob, 1998; McKinley & Nosofsky, 1996; Minda & Smith, 2002; Sigala & Logothetis, 2002; Stewart & Morin, 2007). However, these results only confirm that categorization is based on the manipulated features. In contrast, other studies have characterized how visual object categorization in natural tasks is affected by various manipulations. For instance, animal categorization is unaffected by color (Delorme, Richard, & Fabre-Thorpe, 2000), by removal of certain spatial frequencies (Harel & Bentin, 2009; Morrison & Schyns, 2001; Nandakumar & Malik, 2009), can be performed on

silhouettes (Quinn, Eimas, & Tarr, 2001), and does not depend on Fourier power in the image (Girard & Koenig-Robert, 2011; Joubert, Rousselet, Fabre-Thorpe, & Fize, 2009; Wichmann, Braun, & Gegenfurtner, 2006; Wichmann, Drewes, Rosas, & Gegenfurtner, 2010). Although these results constrain the information used for categorization, they do not explicitly define the underlying features. It is also not clear whether these features are purely visual or influenced by verbal or semantic factors (Goldstone, 1994). In fact, it is a common assumption that animal or vehicle categorization is a high-level task involving visual as well as semantic representations (Li, VanRullen, Koch, & Perona, 2002; Peelen, Fei-Fei, & Kastner, 2009; Rousselet, Fabre-Thorpe, & Thorpe, 2002).

Any candidate feature representation must account for two important observations regarding categorization. First, categorization tasks vary in difficulty: humans are fastest to categorize an object at the superordinate level (e.g., animal), slower at the basic level (e.g., dog), and slowest at the subordinate level (e.g., Labrador) (Large, Kiss, & McMullen, 2004;

Macé, Joubert, Nespoulous, & Fabre-Thorpe, 2009; Mack, Wong, Gauthier, Tanaka, & Palmeri, 2009). The fact that superordinate categorization is fastest contradicts the classic basic level advantage (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), but this discrepancy arises from the key press responses used in recent studies compared to the verbal naming used in the classic studies (see Discussion). A second influential observation is that humans take longer to categorize atypical objects than typical ones (McCloskey & Glucksberg, 1978; Rosch et al., 1976; Rosch & Mervis, 1975). Accounts of both observations have invoked visual, verbal, and semantic representations but do not disentangle them (Goldstone, 1994; Macé et al., 2009; Margolis & Laurence, 1999; Rosch et al., 1976; Smith et al., 1998).

Here we hypothesized that visual categorization depends on visual similarity relations between objects within and outside the category. We performed four separate experiments to investigate whether similarity relations in visual search can predict rapid visual categorization. In each experiment, subjects performed a visual categorization task and a visual search task (Figure 1). In the visual categorization task, they were asked to categorize a briefly flashed object as belonging to a particular category or not. In the visual search task using the same images, they had to identify an oddball target among distractors and were given no explicit instruction regarding the identity of the targets. We took the search times for each pair of images as a measure of perceived similarity between these images. We then investigated whether these perceived similarity relations could predict categorization performance.

In Experiment 1, subjects were asked (in separate blocks) to categorize an object as an animal, dog, or Labrador, and they then performed a single visual search block involving all pairs of objects used in the three tasks. Objects in this experiment were stereotyped

in their view. In Experiment 2, we set out to test whether our results would generalize across variations in view. Here, subjects performed an animal categorization task in which objects were presented in four different three-dimensional views. In Experiment 3, we tested whether the results would hold for another common category, namely vehicles. In Experiment 4, we tested whether the results would generalize to a category defined by its function rather than visual form, namely tools. In each experiment, we used search times to predict categorization times and obtained a striking correlation in all tasks. The data were also accounted for by a model based on extracting coarse structure from images. Taken together, these results suggest that both visual search and rapid visual categorization are driven by object representations sensitive to coarse object structure.

Experiment 1: Categorization across hierarchical levels

In Experiment 1, we tested the hypothesis that a single set of similarity relations (measured using visual search) could predict categorization performance at three distinct hierarchical levels—superordinate (animal), basic (dog), and subordinate (Labrador). To this end, subjects had to categorize a briefly presented object as an animal, dog, or Labrador in separate blocks.

Methods

Subjects: A total of 12 subjects, aged 20–30 years, with normal or corrected-to-normal vision were recruited for the experiments. All participants were naïve to the purpose of the experiments. Subjects gave written

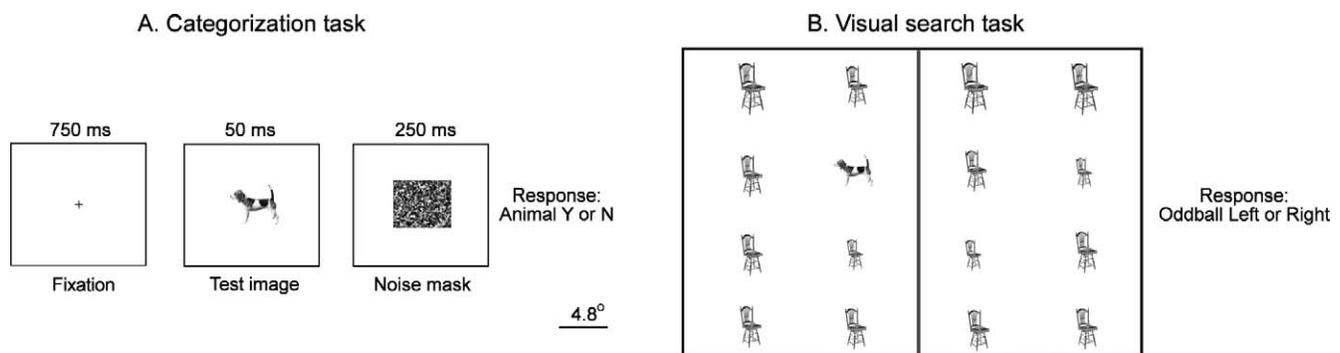


Figure 1. In each experiment, human subjects performed a visual categorization task (A) and a visual search task (B) on the same objects. In the visual categorization task, subjects saw a briefly presented image followed by a noise mask and had to indicate whether or not the object belonged to a particular category (in this example, animal). In the visual search task, subjects saw an array containing an oddball and had to indicate whether the oddball appeared on the left or right half of a vertical red line on the screen. No instructions were given as to the nature or category of the oddball target.

consent to a protocol approved by the Institutional Human Ethics Committee of the Indian Institute of Science.

Subjects were seated approximately 50 cm from a computer monitor that was under control of custom Matlab programs (Mathworks, Natick, MA) written in Psychtoolbox (Brainard, 1997). Each subject performed a categorization task and a visual search task. The order of these two tasks was counterbalanced across subjects. None of the subjects declared any expertise with dogs or Labradors, but we did not perform any detailed tests of their expertise.

Stimuli: The image set consisted of 48 gray-scale images, of which 24 were animals and 24 were nonanimals (12 man-made and 12 natural objects). The 24 animals consisted of 12 dogs and 12 animals that were not dogs. The 12 nondog animals consisted of six typical animals (all quadrupeds: cat, cow, elephant, llama, rhinoceros, and deer) and six atypical animals (two snakes, two birds, one monkey, and one kangaroo). The fact that these animals are indeed atypical was independently established using a typicality rating task (however, atypicality may depend on context; see below and also [Discussion](#)). Of the 12 dogs, there were six Labradors and six non-Labradors. All images were segmented from their original scene context and presented against a black background. Images were equated for brightness and rescaled such that their longer dimension was 140 pixels—this corresponded to a visual angle of 4.8°. Animal images were chosen to be profile views with the head pointed left—this was done to minimize influences of view angle on search times. The same images were used in both categorization and visual search tasks.

Categorization task: Subjects performed three categorization blocks—an animal task, a dog task, and a Labrador task. Block order was counterbalanced across subjects. Each block began with a preview of all objects to avoid confusion regarding the category (primarily for the Labrador task). Each trial began with a fixation cross that appeared for 750 ms, followed by the test object presented briefly for 50 ms, followed by a noise mask for 250 ms ([Figure 1A](#)). The subjects were instructed to press the “M” key to indicate that the object belonged to the target category being tested and to press “Z” otherwise. Subjects had to make a correct response within 2 s, failing which the trial repeated after a random number of other trials. The next trial began 500 ms after the subject made a response. Each image was presented eight times within a block. To investigate possible repetition or learning effects, we performed a post-hoc analysis in which we separated the eight responses to each image into the two halves. We found a modest decrease in reaction times between the first half and second half (average = 21 ms), but the main trend in categorization times (animal < dog <

Labrador) remained the same and attained statistical significance in both halves.

A total of 48 objects were presented in the animal task—24 animals (18 typical and six atypical) and 24 nonanimals (12 natural and 12 man-made objects). The atypical animals consisted of two birds (seagull and pigeon), two snakes (viper and cobra), a monkey, and a kangaroo—these were rated in an independent experiment as being the six most atypical animals. In the dog task, there were 24 objects—12 dogs and 12 nondogs. Of the 12 nondogs, there were six animals—three typical (cat, llama, and rhinoceros), three atypical (bird, monkey, and snake), and six nonanimals (flower, shoe, squash, stone, leaf, and bottle gourd). In the Labrador task, we tested six Labradors and six non-Labradors. The six non-Labradors comprised two dogs (dachsbracke and beagle), two typical animals (cat and llama), and two nonanimals (flower and shoe).

In a separate experiment, we confirmed that the results were qualitatively similar when the relative proportion of nondogs was kept identical in the animal and dog tasks. Specifically, since there were 12 nondog animals and 24 inanimate objects in the animal task, we kept the relative proportion the same in the dog task by selecting four nondog animals and eight inanimate objects as distractors in the dog task. We found that differences in categorization performance were qualitatively similar (data not shown; but see Macé et al., 2009 for a similar observation).

Visual search task: Subjects performed a single visual search task involving the set of images used in all three categorization tasks. The experiment began with a motor reaction measurement block in which a white circle appeared on the left or right side of the screen, and the subject had to indicate with a key press the side on which the target appeared (Z for left, M for right). This was followed by practice visual search trials involving random objects (not belonging to the image set). Each visual search trial began with a fixation cross that appeared for 500 ms, followed by a 4 × 4 array of items consisting of one oddball target among a field of otherwise identical distractors with a red vertical line in the middle of the screen (to facilitate left/right judgments). The search array was displayed until the subject made a response with a maximum of 5 s, after which the trial was marked as an error trial and repeated later on ([Figure 1B](#)). We varied the size of the distractors relative to the target in order to prevent low-order cues such as image size (or alignment of items along rows or columns) from influencing visual search. Specifically, among the 16 items in the 4 × 4 array, one target and seven distractors measured 80% of the object size used in the categorization experiment (i.e., 3.84° along the longer dimension), and four distractors each measured 60% and 100% of the original size.

Subjects were asked to respond as quickly and accurately as possible with a key press (Z or M) indicating which side (left or right) of the screen the oddball target appeared. They were given no instructions regarding the nature or category of the oddball target. We tested all 1,128 possible pairs of images (48 choose 2) in this manner. For each image pair (A, B), subjects performed four correct trials—the target could be either A or B and the target could be on the left or right. The location of the target was random in each trial and search displays corresponding to the image pairs appeared in random order. Incorrect trials were repeated after a random number of other trials. Subjects performed this task across two sessions lasting roughly 30 min each with a break in between. For each subject, image pairs were randomly assigned to the two sessions. *Typicality rating task:* We performed an additional experiment on 12 independent subjects (the same subjects who performed Experiments 3 and 4) to assess the degree to which they considered the animal images as typical examples of the animal category. Specifically, these subjects were asked to rate each animal in the animal task on a scale of 1–5, where 1 implied that the animal was a bad example of the category, and 5 indicated that it was a good example. We note, however, that atypicality can depend on context—a pigeon rated as atypical compared to the many four-legged animals in this set would clearly be considered typical among a set of birds (see Discussion).

Within-category and between-category similarity

We reasoned that the time required to categorize an object would depend on (a) its similarity to items of its own category (denoted as within-category similarity, or CRT) and (b) its similarity to items outside its category (denoted as between-category similarity, or NRT). For each object, we calculated its CRT as the average search time (across subjects and repetitions) required during visual search to find the object among members of its own category (or vice-versa; e.g., the beagle among all other animals, or all animals among beagles). NRT was also calculated likewise as the time required to find the object among items outside its own category (e.g., a beagle among cucumbers or vice-versa). When testing a model such as coarse footprint or aspect ratio, we calculated the pair-wise similarity between every pair of objects (see below). These pair-wise similarity ratings were then used to calculate CRT and NRT as before.

Coarse footprint

To instantiate similarity driven by coarse object structure, we used a simple image model described previously (Sripati & Olson, 2010). To calculate the

similarity in coarse structure between two images, we first shifted and scaled the images to a constant frame (while preserving the aspect ratio), and then low-pass filtered them using a Gaussian blur function. Next, images were normalized by dividing the intensity of each pixel in the image by the total intensity. A difference image was then created by pixel-wise subtraction of the normalized images. The coarse footprint index was calculated by adding the absolute values of the pixels in the difference image. To convert this coarse footprint index (which is a measure of dissimilarity) into a similarity measure akin to reaction times, we took its reciprocal and used it to calculate the within- and between-category similarity measures. The standard deviation of the Gaussian blur was varied to obtain the best match with the data.

Results

Does visual categorization vary in difficulty?

Subjects were highly consistent in both tasks, as evidenced by the correlation in average reaction times between two independent groups of subjects ($r = 0.74$, $p = 2 \times 10^{-6}$ across 84 objects in the three categorization tasks, and $r = 0.90$, $p = 0$ across 1,128 object pairs in the search task). Thus, the underlying strategies and features used to perform categorization did not differ between subjects. Subjects were fastest to categorize an object as an animal, slower to categorize an object as a dog, and slowest to categorize it as a Labrador (Figure 2A; mean reaction times in the animal, dog, and Labrador tasks were 687, 707, and 784 ms, respectively). This effect was significant as determined using an ANOVA on the reaction time (RT) with subject and task as factors ($p < 0.0001$). Post-hoc analysis of data from individual subjects revealed that although the animal < dog effect was fairly robust and present in 8 of 12 subjects, the other two effects (dog < Labrador and animal < Labrador) were present in all subjects and attained significance in all but two subjects. This effect persisted even upon consideration of the six Labrador images used in all three tasks (mean RTs: 644, 679, and 776 ms, respectively; $p < 0.0001$ for main effect of task, ANOVA) or the 12 images in the Labrador task that were common to all three tasks (mean RTs: 661, 711, and 784 ms, respectively; $p = 5 \times 10^{-29}$, ANOVA). Thus the effect was due to differences in the tasks rather than due to differences in the objects used. Although the set size varied across the three tasks in our study, similar results have been reported using equal set sizes (Macé et al., 2009)—thus the effect is unlikely to be due to different set sizes. Finally, we confirmed that the reaction times do not reflect a speed-accuracy tradeoff: subjects were not only fastest but also most accurate in the animal task, slower and less accurate in the dog task, and slowest and least accurate

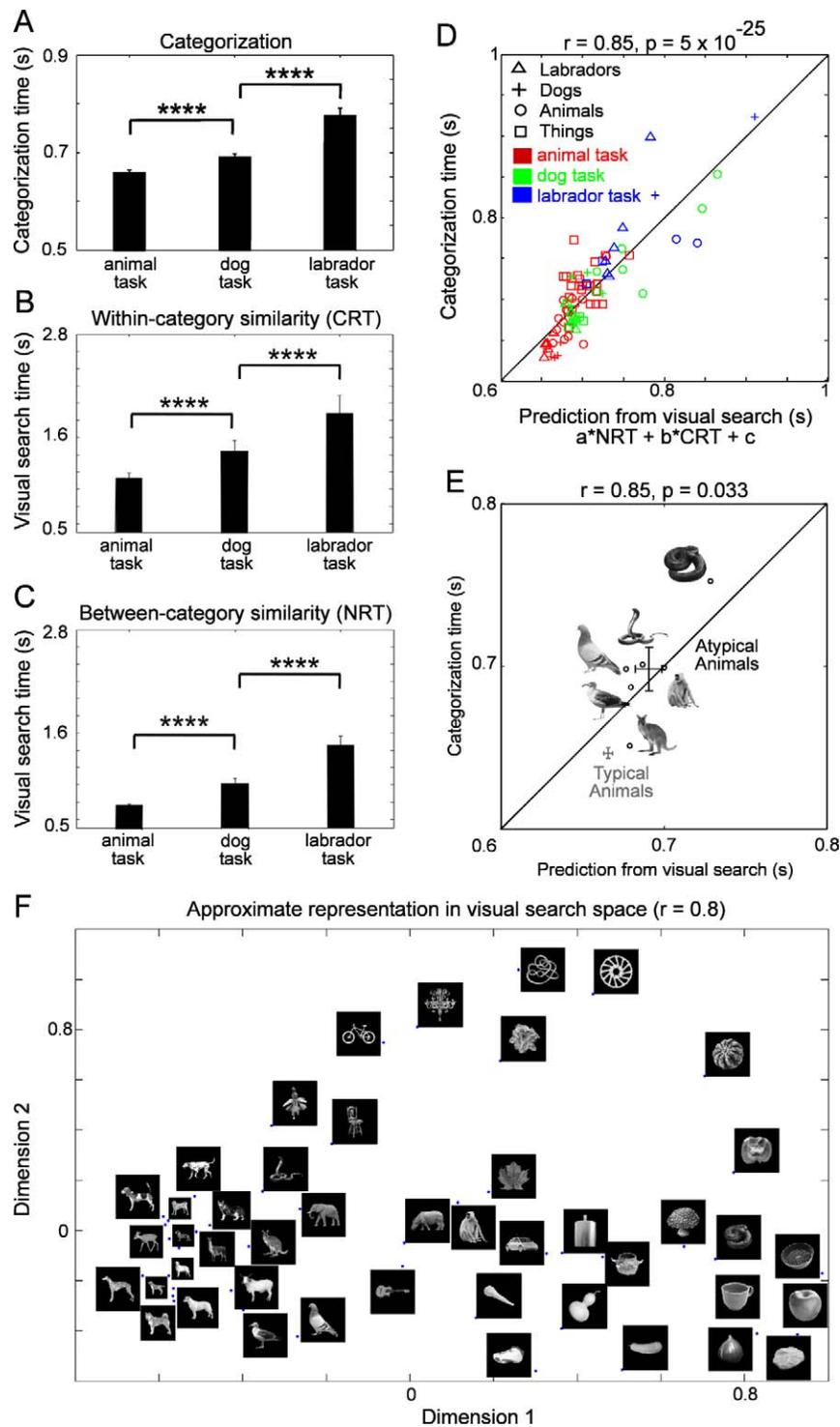


Figure 2. Categorization and visual search times in Experiment 1. (A) Average categorization times in the animal, dog, and Labrador tasks. (B) Within-category search times: times taken to search for animals among animals, dogs among dogs, and Labradors among Labradors. Error bars represent standard errors of the mean. (C) Between-category search times: times taken to search for animals among nonanimals, dogs among nondogs, and Labradors among non-Labradors. (D) Plot of categorization reaction time (averaged across subjects) for each item against the prediction based on average within- and between-category search times. *Triangles* represent Labradors, *plus* symbols represent non-Labrador dogs, *circles* represent other animals, and *squares* indicate inanimate objects. The color of each symbol indicates the task: *red* represents the animal task, *green* represents the dog task, and *blue* represents the Labrador task. All trends reached a high level of statistical significance (**** represents $p < 0.00005$). (E) Categorization times plotted against predicted times for atypical animals (two birds, two snakes, one monkey, and one kangaroo). *Circles* indicate average predicted and observed times

→

in the Labrador task (mean accuracy in the animal, dog, and Labrador tasks was 94.6%, 92.3%, and 92%, respectively; $p < 0.05$, ANOVA). We conclude that visual categorization in humans varies systematically with task: it is easiest to categorize objects as animals, harder to categorize them as dogs, and hardest to categorize them as Labradors.

Can visual search similarity account for categorization?

One possible explanation for the observed differences in categorization difficulty is that the three tasks differ in hierarchical level (superordinate vs. basic vs. subordinate levels). A simpler explanation (requiring no assumption about category hierarchy) is that the tasks vary in difficulty because of differences in the similarity relations between members within and outside each category. We set out to investigate whether this simpler account might explain categorization.

We considered two specific hypotheses regarding how similarity relations might influence categorization: (1) an object might be easy to categorize if it is similar to members of its own category; (2) alternatively, it might be easy to categorize if it is dissimilar to members outside its category. However, these hypotheses cannot be directly evaluated without a quantitative measure of similarity. In order to measure the similarity between two images A and B, we measured the average search times for subjects to search for A among Bs or vice-versa (Duncan & Humphreys, 1989). We can therefore rephrase these two hypotheses as: (1) categorization time would decrease with increasing similarity (i.e., search time) between the object and members of its own category (denoted as within-category similarity, or CRT); and (2) categorization times would increase with increasing similarity (i.e., search time) between the object and members outside its category (denoted as between-category similarity, or NRT).

We found that within- and between-category search times exhibited a trend similar to the categorization times (Figure 2B and C). Average within-category search times increased from animals to dogs to Labradors, contrary to the prediction of Hypothesis 1, but nonetheless covaried with average categorization times (Figure 2B; animal mean = 1177 ms, dog mean = 1505 ms, Labrador mean = 1987 ms, $p = 0$, ANOVA).

Thus at a coarse level, the within-category search times increase across tasks, contrary to the trend expected from Hypothesis 1. However, this may be due to differences in the task, and the expected negative correlation might nonetheless be present among objects within each task. To investigate this issue, we calculated the correlation between categorization times and within-category search times across objects in each task. As predicted by Hypothesis 1, categorization times were negatively correlated with the within-category search times for the 48 objects in the animal task ($r = -0.81$, $p = 2.9 \times 10^{-12}$), were negatively correlated and approached statistical significance for the 24 objects in the dog task ($r = -0.33$, $p = 0.11$), and were not correlated for the 12 objects in the Labrador task ($r = 0.05$, $p = 0.87$). These differences in the statistical significance presumably arose from the different sample sizes. Thus, at least for objects within a task, there is a tendency for categorization time to decrease with within-object similarity.

We then investigated whether between-category search times vary across objects used in the three tasks. Consistent with the second hypothesis, the average between-category search times (NRT) increased from animals to dogs to Labradors (Figure 2C; animal mean = 776 ms, dog mean = 955 ms, Labrador mean = 1356 ms, $p = 0$, ANOVA). A detailed correlation analysis yielded an overall positive and significant correlation across 84 objects in the three tasks ($r = 0.76$, $p = 6.2 \times 10^{-17}$). These correlations were also positive when considered separately for objects in each task ($r = 0.25$, $p = 0.08$ in the animal task; $r = 0.8$, $p = 2.5 \times 10^{-6}$ in the dog task; $r = 0.75$, $p = 0.005$ in the Labrador task). Thus, the time to categorize an object tends to decrease as it becomes increasingly dissimilar to members outside its category.

Although the above analyses indicate that categorization times covaried with within- and between-category search times, the correlations were not consistently significant. This could be either because categorization is unrelated to these search times, or alternatively, because categorization may be based on some combination of these measures rather than any one of them considered separately. To investigate this issue, we set up a model that uses both within- and between-category similarity to predict categorization times. To this end, we compared the average time taken by subjects to categorize each object with a linear sum

←
obtained for each atypical animal. The *gray* and *black* crosses represent the means obtained from typical and atypical animals respectively. Error bars represent standard errors of the mean. (F) Approximate representation of animals and things in visual search space, constructed using multidimensional scaling on visual search data. In this plot, distances between images are (approximately) inversely proportional to the average time taken by subjects to find one image among another in visual search. The correlation coefficient above the plot represents the degree to which distances in the two-dimensional plot capture the observed distances from visual search data. Some images are scaled down to accommodate them in the plot, and three others have been deleted to avoid clutter.

of the within- and between-category search times for that object (Figure 2D). The best-fitting weights of the linear sum were obtained using linear regression, and their magnitudes then revealed the relative contribution of within- and between-category similarities towards categorization. We observed a striking correspondence between the predicted and observed categorization times ($r = 0.85$, $p = 5 \times 10^{-25}$; Figure 2D). The best-fitting linear weights (Table 1) indicate that between-category similarity (weight = 0.16) has a four-fold greater influence on categorization times than within-category similarity (weight = -0.04), and their influences are reversed in sign as predicted by the two hypotheses. These weights and, consequently, model performance were largely similar even when the model was fit separately for each task or for each image type (Table 2), suggesting that the same underlying mechanisms may account for categorization performance across the animal, dog, and Labrador tasks. We therefore only report model performance using a single model fit across tasks and images.

To summarize, the time taken to categorize an image as an animal, dog, or Labrador is influenced strongly by its similarity to members outside the category and only weakly by its similarity to members within its category. The high degree of correlation is striking because (a) the data come from two seemingly disparate visual tasks—one involving scrutiny of single objects (categorization) and the other involving searching for an oddball target in an array (visual search)—which might have been driven by entirely different visual representations; (b) the underlying feature representations and/or strategies used by subjects could have differed widely; and (c) categorization could have

potentially been based on a much larger number of similarity relations than those tested here.

The striking correlation between categorization and search may arise simply from an overall difference in both categorization and search times between tasks or between stimuli, but with zero correlation within each group. To confirm that this was not the case, we calculated task-wise and stimulus-wise correlations between the observed and predicted categorization times (Table 2). All within-group correlations were positive and significant, indicating that search times consistently account for categorization times across tasks and across stimuli. However the magnitude of correlation varied slightly across tasks (from $r = 0.75$ to $r = 0.89$) and strongly between stimulus types ($r = 0.42$ to $r = 0.96$). In particular, the degree of fit of the model was strongest for animals, dogs, and Labradors ($r = 0.85$, $p = 1.5 \times 10^{-19}$ across these three groups) and was weaker but still significant for inanimate objects ($r = 0.42$, $p = 0.017$). We conclude that, in general, categorization times for individual objects can be predicted using similarity relations as measured using visual search.

To visualize the underlying similarity relations that contribute to categorization performance, we performed multidimensional scaling on the visual search data. For each pair of images, we took the distance (or dissimilarity) between them to be the reciprocal of the oddball search time. We then performed multidimensional scaling to find the configuration of points in two-dimensional space that correspond best with the observed distances. The best-fitting two-dimensional configuration is depicted in Figure 2F. This configuration is a reasonably faithful representation of the

Task	Item type	Correlation between categorization & search	Model coefficients RT = a × NRT + b × CRT + c		
			a	b	c
Animal/dog/Labrador in canonical views (Experiment 1)	All items	0.85****	0.16	−0.04	0.61
	Category	0.91****			
	Noncategory	0.85****			
Animals in oblique and profile views (Experiment 2)	All items	0.72****	0.08	−0.02	0.62
	Category	0.52*			
	Noncategory	0.56*			
Vehicles (Experiment 3)	All items	0.71****	0.17	−0.05	0.63
	Category	0.43*			
	Noncategory	0.48*			
Tools Experiment 4)	All items	0.62****	0.58	−0.13	0.44
	Category	0.70***			
	Noncategory	0.45*			

Table 1. Summary of categorization time predictions using visual search data. *Notes:* For each image, we calculated the within-category similarity (CRT) and between-category similarity (NRT) using visual search data. We then fit a model that uses a linear combination of NRT and CRT to account for categorization times. The resulting correlations and model coefficients are depicted above. Asterisks represent the statistical significance of the correlation (* $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$, **** $p < 0.00005$).

Model fitted on:	Correlation between categorization & search	Model coefficients $RT = a \times NRT + b \times CRT + c$		
		a	b	c
All tasks/items	0.85****	0.16	−0.04	0.61
Animal task	0.84****	0.09	−0.08	0.7
Dog task	0.89****	0.14	−0.03	0.61
Labrador task	0.77**	0.14	−0.02	0.61
Animals across tasks	0.91****	0.17	−0.02	0.55
Dogs across tasks	0.95****	0.21	−0.02	0.53
Labradors across tasks	0.96****	0.33	−0.13	0.63
Things across tasks	0.43*	0.11	−0.06	0.68

Table 2. Model performance on different images and tasks in [Experiment 1](#). *Notes:* For each image, we calculated the within-category similarity (CRT) and between-category similarity (NRT) using visual search data. We then fit the model either on all objects within a task (e.g., animals and nonanimals in the animal task) or for all objects of a given type (e.g., animal images in all three tasks) for various tasks and image types. The asterisk symbol beside each correlation coefficient indicates its statistical significance, with conventions as in [Table 1](#).

underlying search distances, as evidenced by the degree of fit between the observed distances and distances between images measured in the plot ($r = 0.8$, $p = 0$). The plot clearly reveals that (a) animals and nonanimals form reasonably distinct clusters; (b) nonanimals are more diverse than animals; and (c) atypical animals are dissimilar to the remaining typical animals. These similarity relations form the basis for the ability of the model to predict categorization times. To assess whether these similarity relations can predict category judgments, we trained a linear classifier on the two-dimensional coordinates obtained from multidimensional scaling. The performance of the classifier (92% correct) approached the accuracy of humans on the same task (95% correct; see [Table 4](#)). Thus, similarity relations in visual search account for animal categorization performance in humans.

How do similarity relations explain classic categorization phenomena?

We wondered whether well-known categorization phenomena can also be explained using similarity relations based on visual search. We account for three basic and well-known observations regarding visual categorization, as detailed below.

Can similarity relations explain differences in categorization difficulty?

The first and most widely reported phenomenon is that categorization increases in difficulty from superordinate to basic to subordinate levels (Macé et al., 2009; Mack et al., 2009). This was the case even in our data ([Figure 2A](#)). It was also true for our model: categorization times were smallest for animal categorization (mean = 687 ms), larger for dog categorization

(mean = 715 ms), and largest for Labrador categorization (mean = 770 ms), and all three differences were statistically significant ($p < 0.0007$, t -test). In the model, these differences arise because of differences in within- and between-category search times in the three different tasks. In other words, categorizing an object as a dog is hard because nondogs contain animals that share several features (e.g., legs, head) with dogs, increasing the average between-category similarity (NRT). This is consistent with the observation made that categorization times decrease when nondogs in a dog task are restricted solely to be inanimate objects (Macé et al., 2009). Because our model accounts for categorization performance across the superordinate, basic, and subordinate levels, we conclude that categorization difficulty may be explained on the basis of similarity relations alone without invoking category hierarchy.

Can similarity relations explain the longer time to reject category membership?

The second basic phenomenon is that humans are typically faster to confirm category membership than to reject it. This was the case even in our data: in the animal categorization task, subjects were faster to categorize an object as an animal than to reject it as a nonanimal (659 ms for animals, 716 ms for nonanimals, $p = 4 \times 10^{-51}$, ANOVA). We observed a similar trend in the model predictions (672 ms for animals, 701 ms for nonanimals, $p = 1.7 \times 10^{-6}$, t -test). Upon closer inspection, we found that within-category search times for animals (mean = 1476 ms) were significantly larger than the within-category search times for nonanimals (mean = 784 ms for nonanimals, $p = 2 \times 10^{-14}$, t -test). This is a straightforward consequence of the fact that nonanimals are more diverse and therefore more

Task	Item type	Correlation with categorization (<i>r</i>)	Model coefficients		
			RT = <i>a</i> × NRT + <i>b</i> × CRT + <i>c</i>		
Animal/dog/Labrador (Experiment 1)	All items	0.79****			
Optimal blur = 0.10	Category	0.68****	0.082	0.008	0.49
Visual search correlation = 0.41****	Noncategory	0.79****	0.097	−0.04	0.62
Animals with three-dimensional views (Experiment 2)	All items	0.64****			
Optimal blur = 0.10	Animals	0.41*	0.02	−0.02	0.66
Visual search correlation = 0.46****	Nonanimals	0.38#	0.02	0.02	0.61
Vehicles (Experiment 3)	All items	0.67****			
Optimal blur = 0.10	Vehicles	0.43*	−0.04	−0.01	0.77
Visual search correlation = 0.40****	Nonvehicles	0.33#	−0.008	0.03	0.71
Tools (Experiment 4)	All items	0.63****			
Optimal blur = 0.00	Tools	0.71***	0.83	−0.8	0.72
Visual search correlation = 0.25****	Nontools	0.47*	0.64	−0.34	0.7

Table 3. Categorization time predictions using coarse footprint. *Notes:* The coarse footprint model was used to calculate pair-wise similarities between objects. These pair-wise similarities were used to calculate the NRT and CRT for each object, which was then used to predict categorization times as described in Table 2. The resulting model fits and coefficients are depicted above, with conventions as before.

dissimilar to each other on average than are animals. Because of the way they are calculated, between-category similarity is simply the average time to search for animals among nonanimals or for nonanimals among animals. Thus, between-category similarity is identical for both animals and nonanimals. The longer categorization times for nonanimals in the model arise entirely because of smaller within-category search times that are penalized by negative weights by the model. Thus, humans are slower to reject category membership because noncategory items are more diverse.

Can similarity relations explain the categorization of atypical members?

The third well-known observation is that humans take longer to categorize atypical category members compared to typical category members. To independently assess typicality, we performed an additional experiment in which 10 subjects were asked to rate each animal image in the animal task on a scale of 1–5, where 1 indicated that the object was a poor example of the animal category, and 5 indicated that it was a good example. We then chose the six animals with the smallest typicality ratings—these were the viper (mean

Task	Item type	Human accuracy (%)	Classifier accuracy	Classifier accuracy	Classifier accuracy
			using visual search (%)	using coarse footprint (%)	using aspect ratio (%)
Animal	All items	95	92	89	63
	Category	95	92	83	66
	Noncategory	95	92	96	58
Animals with three-dimensional views	All items	98	94	83	50
	Animals	98	100	83	41
	Nonanimals	98	88	83	58
Vehicles	All items	93	88	81	63
	Vehicles	95	92	83	66
	Nonvehicles	92	83	79	58
Tools	All items	93	79	88	63
	Tools	93	83	83	54
	Nontools	94	75	92	71

Table 4. Categorization accuracy for humans and models. *Notes:* For human data, accuracy is calculated across trials and averaged across subjects. For visual search and coarse structure data, pair-wise similarity relations between images were projected into two-dimensional space using multidimensional scaling, and a linear classifier was trained on these coordinates. For each object the predicted category was obtained by training the classifier on all other objects. For aspect ratio data, the aspect ratio of each image was used as input to the linear classifier (the multidimensional scaling is redundant since its output will be identical to the input).

rating = 2.0), pigeon (3.5), seagull (3.6), cobra (3.6), kangaroo (4.1), and monkey (4.1). The average rating for these six animals (3.5) was smaller than the average rating for the remaining animals (4.7) and attained significance as assessed by a *t*-test ($p = 1.4 \times 10^{-6}$).

Because half the animals in our image set were dogs, we were concerned that typicality ratings may have been biased in favor of dogs. However, we found no evidence supporting this notion: typicality ratings for dogs (mean = 4.7) were similar to the ratings for nondog animals excluding the six atypicals (mean = 4.7). Furthermore, typicality ratings for the six atypical animals were significantly different from the remaining nondogs ($p = 0.004$, *t*-test). In the categorization task, however, subjects were significantly faster to categorize dogs (mean = 640 ms) than the other nondog typical animals (mean = 660 ms) in the animal task ($p = 0.001$, ANOVA). However, both dog and nondog categorization times were substantially faster than the times taken to categorize atypical animals (see below). Assessing whether these effects can be attributed to bias due to the image set used or to intrinsic differences will require careful manipulation of set context and is beyond the scope of this study. Nonetheless, our model automatically incorporates these effects in the form of many more image pairs containing dogs or typical animals compared to atypical animals, which in turn influence within- and between-category similarity.

Consistent with previous studies, subjects were slowest at categorizing atypical animals (mean = 699 ms) compared to other animals in general (typical mean = 647 ms, atypical mean = 699 ms, $p < 0.0001$, ANOVA). Typicality ratings were also negatively correlated with categorization times in the animal task ($r = -0.89$, $p = 6.3 \times 10^{-9}$), implying that on an image-by-image basis, animals considered to be more typical were categorized faster than the atypical ones. Importantly, we observed a similar trend in the between-category visual search times: among arrays containing nonanimals, subjects took longer to search for atypical animals compared to typical animals (typical mean = 758 ms, atypical mean = 795 ms, $p = 8 \times 10^{-5}$ for main effect of typicality, ANOVA on average search times for each item with subject and typicality as factors). Similarly, among arrays containing animals, subjects took longer to search for typical animals compared to an atypical animal (typical mean = 1590 ms, atypical mean = 1133 ms, $p = 9 \times 10^{-30}$, ANOVA). Thus, subjects considered atypical animals as more similar to inanimate objects and less similar to other typical animals. Observed categorization times for atypical animals were strongly correlated with model predictions based on between- and within-category search times (Figure 2E; $r = 0.85$, $p = 0.033$). We conclude that atypical animals take longer to categorize as animals

because of their greater similarity to nonanimals and their lower similarity to other animals.

Overall, our results suggest that differences in performance on visual categorization arising from category level (superordinate, basic, or subordinate), category membership (belonging or not belonging), and object typicality (atypical vs. typical) can be explained entirely by visual similarity alone.

Can coarse object similarity account for categorization?

Having established a close correspondence between categorization times and similarity as determined by visual search, we then asked whether categorization times for individual objects could be predicted directly from the image pixels. A positive outcome would be remarkable because it would relate animal categorization—a high level cognitive process—directly to image content without invoking high-order verbal or semantic influences. We reasoned that since animal categorization is unaffected by blurring (Nandakumar & Malik, 2009), it must depend on the coarse structure in an image. So does visual search for images differing in global arrangement (Sripati & Olson, 2010). In these studies, the coarse footprint of an image was formed by shifting and scaling it to a fixed frame, normalizing its brightness, and blurring it using a Gaussian function. In the present study, the normalizing operations are redundant because images were already equated for these factors. The difference in coarse structure for a pair of images was then calculated by computing the absolute pixel-by-pixel difference between the coarse footprints of the two images. The reciprocal of this coarse footprint difference was taken to be a similarity measure akin to reaction times in visual search.

We reasoned that differences in coarse footprint might account for similarity relations between images, which in turn might predict categorization times (Figure 3A). Indeed, we obtained a strong correlation between categorization times and the predictions from a model based on coarse content ($r = 0.68$, $p = 2 \times 10^{-12}$; Table 3). This correlation remained significant even upon considering the three categories separately ($r = 0.44$, $p = 0.001$ for the animal task; $r = 0.62$, $p = 0.001$ for the dog task; and $r = 0.79$, $p = 0.002$ for the Labrador task). We found a substantial improvement in model predictions when the model was fit separately for category and noncategory items (Figure 3B; $r = 0.79$, $p = 9 \times 10^{-19}$). This was true for all the categories. Accordingly we report the performance of the coarse footprint model upon fitting it separately on the category and noncategory items. Upon varying the coarseness of image content, we obtained the highest correlation when images were blurred by a Gaussian blur with standard deviation equal to 0.1 times the size

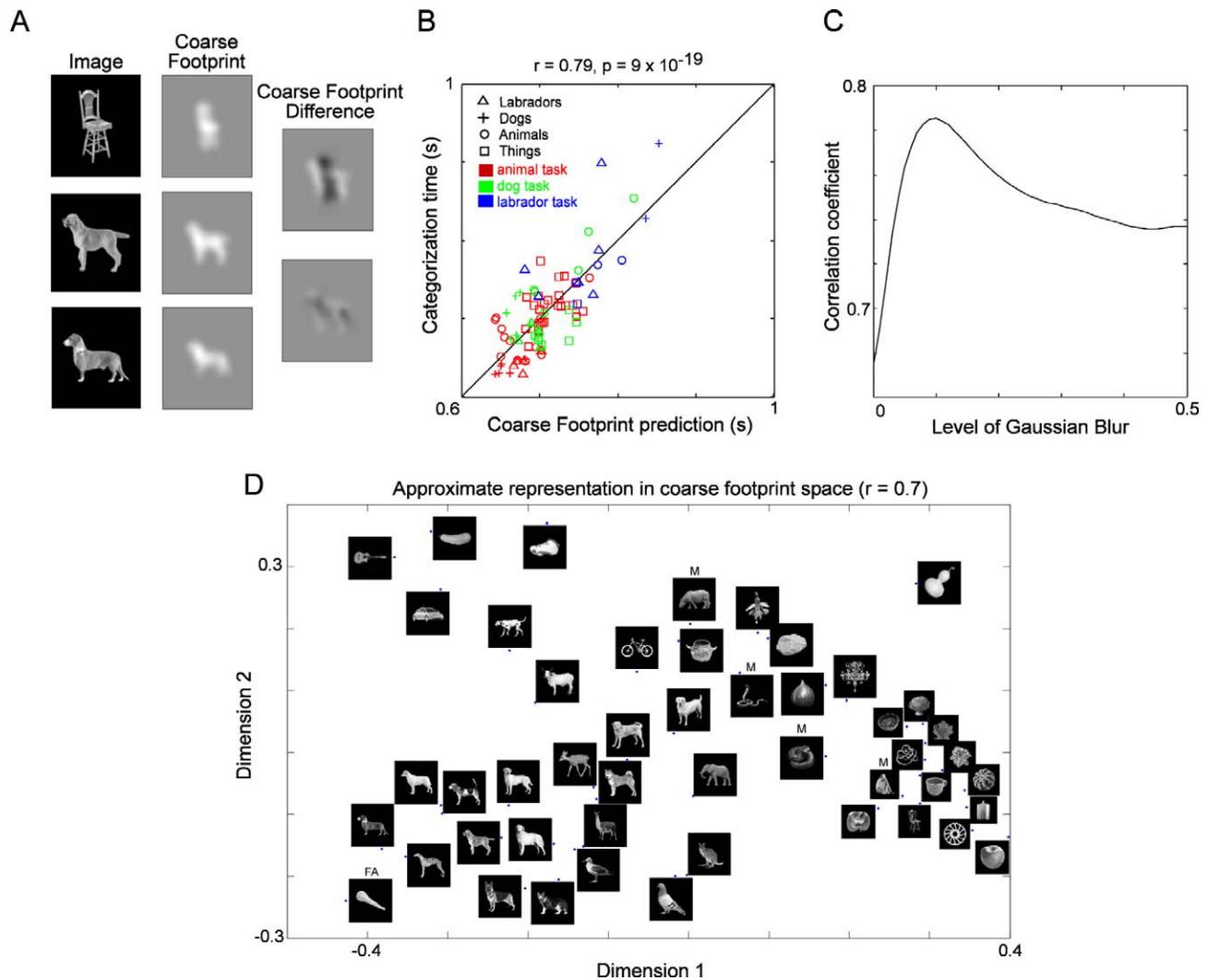


Figure 3. A model of coarse image structure accounts for categorization times. (A) Each image is blurred using a Gaussian blur to create its coarse footprint. The reciprocal of the difference in coarse footprints of each image pair yields a measure of similarity in the coarse structure. These pair-wise similarity measures are then used to calculate within- and between-category similarity, which were then used to predict categorization times. (B) Categorization times plotted against predictions based on coarse footprint for the optimum level of blur. Each point represents a reaction time pair for each stimulus in the animal, dog, and Labrador categorization tasks. Symbol conventions are identical to those in Figure 2D. (C) Correlation between categorization and coarse footprint predictions as a function of standard deviation of the Gaussian blur. (D) Similarity relations between objects in the animal task, as revealed by coarse image structure. Images are represented such that nearby objects have small coarse structure differences and distant objects have large coarse structure differences. The labels above each image represent the category label predicted by a linear classifier trained on this multidimensional representation: *no label* = correct prediction, *FA* = false alarm (i.e., a nonanimal misclassified as an animal), *M* = miss (i.e., animal misclassified as a nonanimal). The classifier misclassified only 5 out of the 48 objects. Three of these misclassifications were atypical animals, which were harder also for humans. Some images have been scaled down to accommodate them in the plot, and other images (but not their data points) have been deleted to avoid clutter.

of the object (Figure 3C). This is close to the value of 0.08 obtained during visual search for objects differing in global arrangement (Sripati & Olson, 2010). We conclude that visual categorization depends on similarity relations driven by coarse object structure.

The results above demonstrate that coarse object structure can predict categorization times, but do not show that it is sufficient to perform category judg-

ments. Specifically, we wondered whether similarity relations, as measured by coarse footprint differences, can be used by a computer-based classifier to predict whether a given object was an animal. To this end, we took the pair-wise distances between all 48 objects in the animal task and performed multidimensional scaling. This yielded the coordinates of each object in a multidimensional space such that the Euclidean

distances best approximate the pair-wise distances. The best-fitting two-dimensional configuration, depicted in Figure 3D, is similar to that obtained using the visual search data (Figure 2F). We then performed a linear discriminant analysis to obtain a linear boundary that predicts the category label of each image. To avoid over-fitting, we performed a leave-one-out cross validation: in other words, we predicted the category label of each object by training the classifier on the coordinates and category labels of the remaining 47 objects. The classifier correctly predicted the category labels of 43 of the 48 objects in the animal task (Figure 3D). Of the four misses, three were atypical animals, suggesting that, like humans, the classifier finds it difficult to categorize atypical animals. Its performance (89% correct) was close to the accuracy of humans on the same task (95% correct; Table 4). We conclude that coarse object structure can predict human category judgments as well as categorization times during animal categorization.

Experiment 2: Animal categorization with varying three-dimensional view

The two essential findings of the previous experiment are that (a) categorization times can be accounted for by visual search and that (b) coarse structure can account for these data. However, these results are based on objects appearing in a canonical leftward facing profile view. In this experiment, we set out to investigate whether these results would generalize to objects varying in their three-dimensional view. Subjects performed an animal categorization task as before, except that the stimuli now consisted of six animals and six nonanimals, each presented in four possible three-dimensional views. Thus an object could be seen in either a profile or oblique view and could be pointing either left or right (Figure 4).

Methods

Subjects: A total of six subjects were recruited for this experiment and gave informed consent as before. Three of these subjects had participated in Experiment 1. Although we were initially concerned that the observed differences between oblique and profile views arose because these subjects were exposed more to profile views during Experiment 1, a post-hoc analysis revealed no qualitative difference between the results obtained by including or excluding these subjects.

Stimuli: The stimuli comprised 48 gray-scale images (from Hemera Photo Objects), of which 24 were images

of animals (different from those in Experiment 1), and 24 were of nonanimals. There were six unique animals (cow, dog, elephant, horse, stag, and tiger) and six unique nonanimals (motorcycle, shoe, chair, gourd, pepper, and pumpkin). Each of these objects was presented in four distinct three-dimensional views. The four views of each object were profile views and oblique views, pointing either leftwards or rightwards. For each object, we defined the profile view as its sideways view (i.e., the view in which its image was most elongated). The oblique view was chosen to be a view of the same object rotated approximately 45° out of the plane. Objects were presented against a black background and equated for brightness as before. We also equated image size across all objects: to prevent any low-order visual cues from contributing to task performance, we resized the profile views of all objects such that their longer dimension (typically their width) was 4.8°. We then resized each oblique view image such that its height was equal to the corresponding profile view. This was done to achieve the overall effect that the oblique view appeared to be a plausible three-dimensional rotation of the object seen in the profile view.

Categorization and visual search tasks: Subjects performed a categorization task and a visual search task, with task order counterbalanced across subjects. The tasks were exactly the same as described earlier, with the exception of the stimuli.

Results

Subjects' reaction times were highly consistent in both tasks as evidenced by a strong correlation between two independent groups of subjects ($r = 0.51$, $p = 0.001$ across 48 objects in the categorization task, and $r = 0.83$, $p = 0$ across 1,128 object pairs in the search task). Thus, the underlying strategies and/or features used to perform each task did not differ between subjects. Subjects were faster to categorize profile views of animals compared to their oblique views (mean = 654 ms for profile views, 670 ms for oblique views), as revealed by an ANOVA on the categorization reaction time with subject, animal (six levels) and view (four levels) as factors ($p = 0.02$ for main effect of view). We found no such difference for the nonanimals in our set (mean = 691 ms for profile view, 690 ms for oblique views, $p = 0.86$ for main effect of view). We conclude that humans categorize profile views of animals faster than oblique views, but show no such effect for nonanimals.

We then set out to investigate whether the observed categorization times could be predicted using visual search. As before, we calculated for each object the average search time to find this object as target among all other members within its category or vice-versa (i.e.,

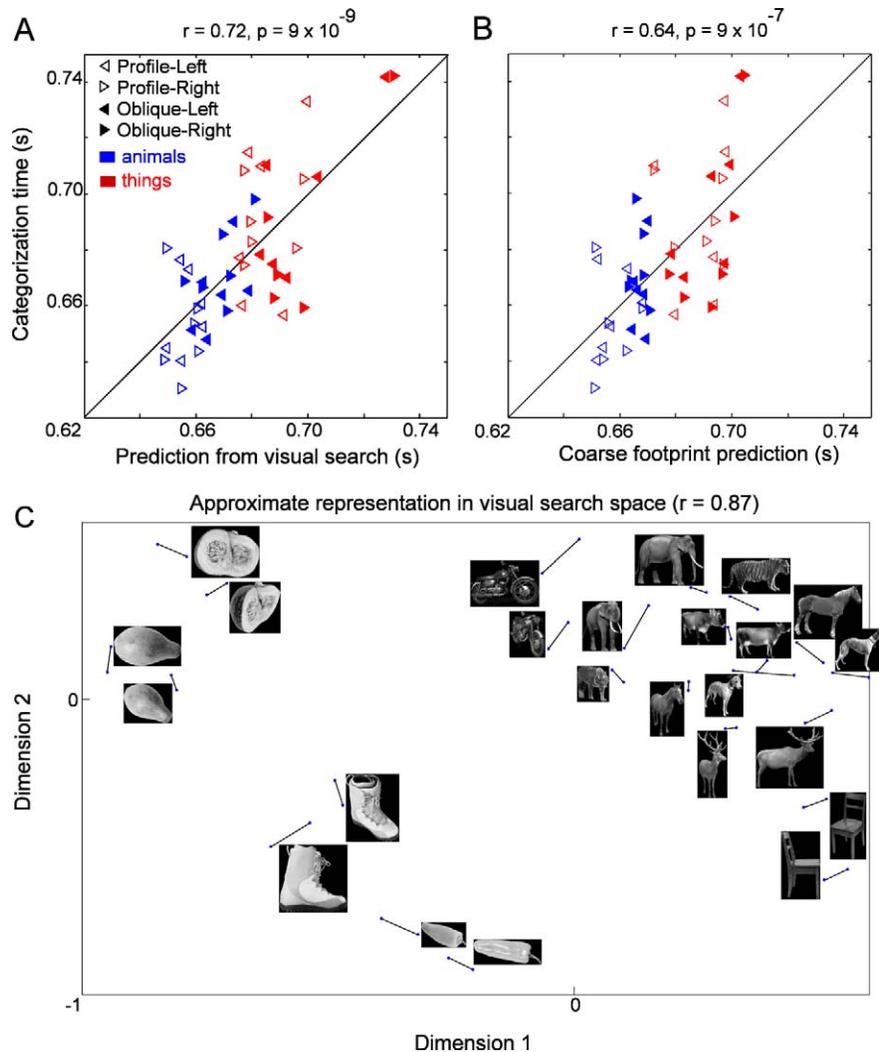


Figure 4. Animal categorization task with objects varying in three-dimensional view (Experiment 2). Subjects were asked to categorize an object as an animal or not as before, except that the objects consisted of animals or nonanimals in four possible three-dimensional views: two profile views (pointing left or right) and two oblique views (pointing left or right). (A) Correlation plot of categorization times for each item (averaged across subjects) versus predictions using search data; different shapes represent the reaction times for animals in the four different views (profile/oblique left/right); *blue* represents reaction times of all animals, and *red* represents reaction times of all things. (B) Correlation plot of observed categorization times versus predictions using the coarse structure model with conventions as in Figure 4A. (C) Approximate representation of animals and nonanimals in visual search space. *Dots* represent all 48 images. *Dots* connected by a line represent pairs of images that are identical short of mirror reflection. All images are to scale except for the cow images, which are scaled down by 80% to accommodate them in the plot.

within-category similarity) and the average time required to search for this object among items outside its category (i.e., between-category similarity). We then fit a linear model based on these within- and between-category similarities for all 48 objects in the task in order to account for their corresponding categorization times. We observed a significant positive correlation across all items between model predictions and observed categorization times ($r = 0.65$, $p = 5 \times 10^{-7}$; data not shown). However, this fit was significant only for nonanimals ($r = 0.49$, $p = 0.02$) and not so for animals ($r = 0.27$, $p = 0.2$). We therefore sought a model

that would account better for the observed categorization times. Upon calculating within- and between-category search times separately for each view, we obtained a much higher degree of fit (Figure 4A; $r = 0.72$, $p = 9 \times 10^{-9}$) with significant correlations within both the set of animals ($r = 0.52$, $p = 0.009$) as well as nonanimals ($r = 0.56$, $p = 0.005$). We conclude that view-dependent similarity relations account for the time taken by humans to categorize objects varying in three-dimensional view.

To investigate whether coarse object similarity could account for categorization of animals varying in view,

we started as before with the reciprocal of the difference in coarse footprint between two images as their pair-wise similarity. Just as for the visual search data, we then calculated the within- and between-category similarity separately for each view and used these to fit the categorization times. We obtained a positive correlation between the coarse footprint prediction and observed categorization times ($r = 0.64$, $p = 9 \times 10^{-7}$; Figure 4B, see also Table 3). We were also able to predict the category of the item using the visual search data (accuracy = 94%) or using the coarse footprint data (accuracy = 83%). These accuracy levels approached the performance of humans on the same task (98%). We concluded that coarse object similarity can account for categorization even when objects vary in three-dimensional view.

To visualize the similarity relations that underlie categorization, we performed multidimensional scaling on the visual search data as before. The best-fitting two-dimensional configuration is shown in Figure 4C. Several interesting patterns can be seen in these plots: the first and most striking trend is that in both animals and nonanimals, object views related by mirror reflection lie close together. This is consistent with the mirror-image confusion observed in humans (Gross & Bornstein, 1978) and in neuronal activity in high-level visual cortex (Rollenhagen & Olson, 2000).

Second, different views of inanimate objects tended to form distinct clusters (Figure 4C), whereas this clustering is present but not as apparent among animals. The weaker clustering among animals could be due to the greater similarity within animals compared to within nonanimals. Alternatively, the presence of such clustering among animals might have been obscured by the multidimensional scaling and projection of the data into two dimensions. We therefore performed an analysis on the original visual search data. For each view of each object, we compared the reaction times to search for it among its three other views with the reaction times to search for it among all views of all other objects. For every object in our set (both animals and things), within-object search times were longer on average compared to between-object search times (mean = 2068 ms for within-object search, mean = 1200 ms for between-object search, $p = 4 \times 10^{-6}$, t -test). We conclude that the different views of each object (in both animals and nonanimals) form distinct clusters in visual search space.

Third, profile and oblique views appear to form distinct clusters in the case of animals (i.e., in Figure 4C, oblique views appear on the left and profile views on the right). We confirmed that this trend is present in the visual search data by comparing search times for oblique views among profile views across objects with search times for objects in the same view (i.e., profile among profile or oblique among oblique views). For

animals, the average search times for oblique views among profile views (1,376 ms) was significantly smaller than search times for profile among profile (1680 ms, $p = 4 \times 10^{-13}$, ANOVA) or oblique among oblique views (1790 ms, $p = 6 \times 10^{-22}$, ANOVA). In contrast, for nonanimals, average search times for oblique views among profile views (891 ms) did not differ significantly from search for profile among profile views (897 ms, $p = 0.87$, ANOVA) but were significantly different from oblique among oblique views (1007 ms, $p = 0.0003$, ANOVA). We conclude that similarity relations based on visual search reflect both view-dependent and view-invariant representations.

Based on the results of Experiments 1 and 2, we conclude that similarity relations from visual search can account for categorization even when objects vary in view and that these similarity relations are driven by coarse object similarity.

Experiment 3: Vehicle categorization

We performed additional experiments to investigate whether the above results would generalize to other categories. In this experiment, subjects performed a vehicle categorization task on 48 objects (24 vehicles and 24 nonvehicles) and an oddball visual search task on all 1,128 pairs of stimuli (48 choose 2) used in the categorization task.

Methods

Subjects: A total of six new subjects were recruited for this experiment. Because a few of the nonvehicles in the set were animals, we were concerned that prior exposure to the animal categorization task might induce response conflict. Therefore, we only chose subjects who had not previously performed Experiments 1 and 2.

Stimuli: The image set for the vehicles categorization experiment consisted of 48 gray-scale images (from Hemera Photo Objects), comprising 24 vehicles and 24 nonvehicles. The 24 nonvehicles consisted of 12 natural objects (animals such as caribou or cat; fruits such as melon or gourd) and 12 man-made objects (furniture such as chair or table; household items such as light bulb or teapot). All images in the vehicle experiment were chosen to have an oblique view pointing to the left.

Categorization and visual search tasks: Subjects performed a vehicles categorization task and a visual search task, with task parameters as before.

Results

Subjects were highly consistent in both tasks as evidenced by a strong correlation in reaction times between two independent groups of subjects ($r=0.61$, $p=4.6 \times 10^{-5}$ across 24 objects in the categorization task and $r=0.80$, $p=0$ across 1,128 object pairs in the search task). Thus, the underlying strategies and/or features used to perform each task did not differ between subjects. Subjects were faster to categorize an object as a vehicle than rejecting it (mean = 714 ms for vehicles, 770 ms for nonvehicles, $p=6 \times 10^{-17}$, ANOVA). We then used the search times to predict categorization times exactly as before and observed a strong positive correlation between model predictions and categorization times ($r=0.62$, $p=2 \times 10^{-6}$). We found a marked improvement in the performance of the model when it was fit separately on the category and noncategory data—we accordingly show the results using this model ($r=0.71$, $p=2 \times 10^{-8}$; Table 1, Figure 5A). This correlation remained significant when computed separately for the set of vehicles ($r=0.43$, $p=0.04$) and across the set of nonvehicles ($r=0.48$, $p=0.02$). These categorization times were also predicted well by coarse object similarity ($r=0.67$, $p=2 \times 10^{-7}$; Table 3, Figure 5B). In both cases, it can be seen that the models not only account for individual categorization times of vehicles and nonvehicles, but also for general trends such as the longer categorization times for nonvehicles compared to vehicles.

To visualize the similarity relations that underlie vehicle categorization, we performed multidimensional scaling on the visual search data as before (Figure 5C). The resulting plot reveals distinct clusters for vehicles and nonvehicles, which form the basis for the ability of the visual search data to predict rapid visual categorization in humans. Indeed, when we trained a linear classifier on the visual search data, we were able to predict the category (vehicle or not) with an accuracy of 88% correct, approaching the accuracy of humans on this task (93% correct; Table 4). We were also able to predict the category of the item using the coarse footprint data (accuracy = 81%; Table 4). We conclude that vehicle categorization can be accounted for by similarity relations based on visual search, which are in turn driven by coarse object similarity.

Experiment 4: Tool categorization

The objects investigated in the above experiments (animals and vehicles) consisted of categories defined primarily by their visual form. Would our results generalize to categories defined by their motor func-

tion? We reasoned that the category of tools, which are defined primarily by their function and by motor affordances, would be a natural choice to investigate this question. Specifically, we hypothesized that categorization of an object as a tool would be predicted poorly by visual search or coarse object structure, both of which are based solely on visual appearance. Alternatively, it is possible that, under conditions of rapid visual presentation used here, tool categorization is based primarily on visual appearance rather than on motor representations.

Methods

Subjects: Six new subjects were recruited for this experiment—none of the subjects had previously performed Experiments 1, 2, or 3.

Stimuli: The image set for the tools categorization experiment consisted of 48 gray-scale images (from Hemera Photo Objects), comprising 24 tools and 24 nontools. The nontools consisted of 12 natural objects (animals such as tiger, dog, or goat; fruits such as pineapple or banana) and 12 man-made objects (furniture such as bench or sofa; musical instruments such as bugle or guitar). All objects were presented in a profile view pointing towards the left, in a canonical horizontal pose.

Categorization and visual search tasks: Subjects performed a tools categorization task and a visual search task, with task parameters as before.

Results

Subjects were highly consistent in their responses ($r=0.66$, $p=1.9 \times 10^{-6}$ for reaction times between two independent groups of subjects in the categorization task; $r=0.68$, $p=0$ in the visual search task). Subjects tended to categorize tools faster than nontools but this trend did not attain statistical significance (mean = 819 ms for tools, 833 ms for nontools, $p=0.07$, ANOVA). Upon using visual search times to predict categorization times as before (by fitting the model to category and noncategory items separately, we obtained a positive and significant correlation ($r=0.62$, $p=3 \times 10^{-6}$; Figure 6A). This correlation persisted even upon calculating it separately across tools ($r=0.7$, $p=0.0001$) and across nontools ($r=0.45$, $p=0.03$). Thus, rapid visual categorization of tools appears to depend on their perceived visual similarity relations. The categorization data was also predicted using the coarse footprint model ($r=0.63$, $p=2 \times 10^{-6}$; Table 3, Figure 6B). Notably, the best correlation was obtained at zero blur (i.e., no blurring). This optimal blur (blur = 0) differs from the values (blur = 0.1) obtained for animal and vehicle categorization. Thus, higher spatial detail

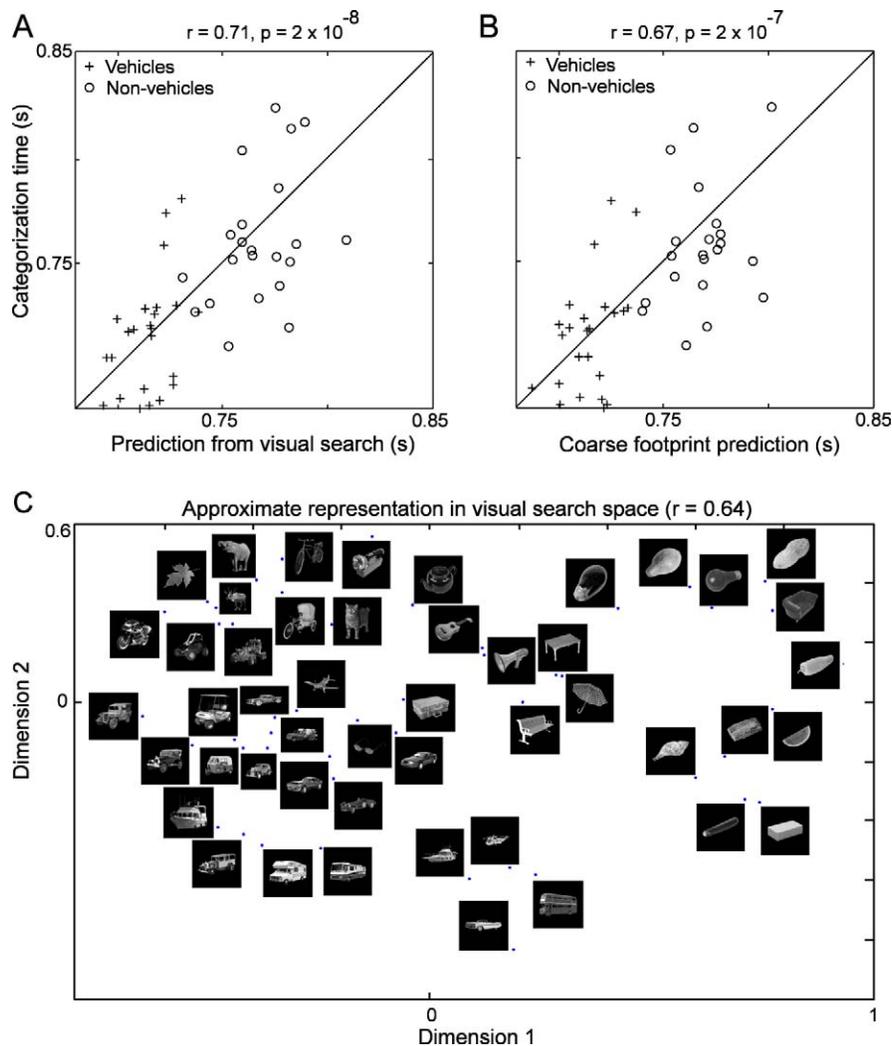


Figure 5. Vehicles categorization task (Experiment 3). Subjects were asked to categorize an object as a vehicle or not. (A) Correlation plot of categorization times for each item in the vehicles task (averaged across subjects) against predictions using visual search data. Plus symbols represent reaction times of all vehicles, and circles represent reaction times of all nonvehicles. (B) Correlation between the observed categorization times against predictions using coarse footprint with the same conventions as in Figure 5A. (C) Approximate representation of vehicles and nonvehicles in visual search space, showing separable clusters according to category. Some images are scaled down to accommodate them in the plot.

may be required for tool categorization compared to animal or vehicle categorization.

To visualize the similarity relations that underlie tool categorization, we performed multidimensional scaling on the visual search data as before (Figure 6C). The resulting plot reveals distinct clusters for tools and nontools, which form the basis for the ability of the visual search data to predict categorization data. Indeed, when we trained a linear classifier on the visual search data, we were able to predict the category (tool or not) with an accuracy of 79% correct, approaching, but not as good as, the accuracy of humans on this task (93% correct; Table 4). We were also able to predict the category of the item using the coarse footprint data (accuracy = 88%; Table 4). We conclude that tool

categorization can be accounted for by similarity relations based on visual search, which are in turn driven by pixel-level similarity between objects.

Can aspect ratio account for categorization?

The fact that visual search—which is known to be sensitive to low-level image differences such as brightness, color, size, aspect ratio, etc. (Wolfe & Horowitz, 2004)—can explain categorization raises the possibility that subjects may have performed these tasks using simple features that may have differed between category and noncategory items. Because images in all tasks were equated for size, contrast, and brightness,

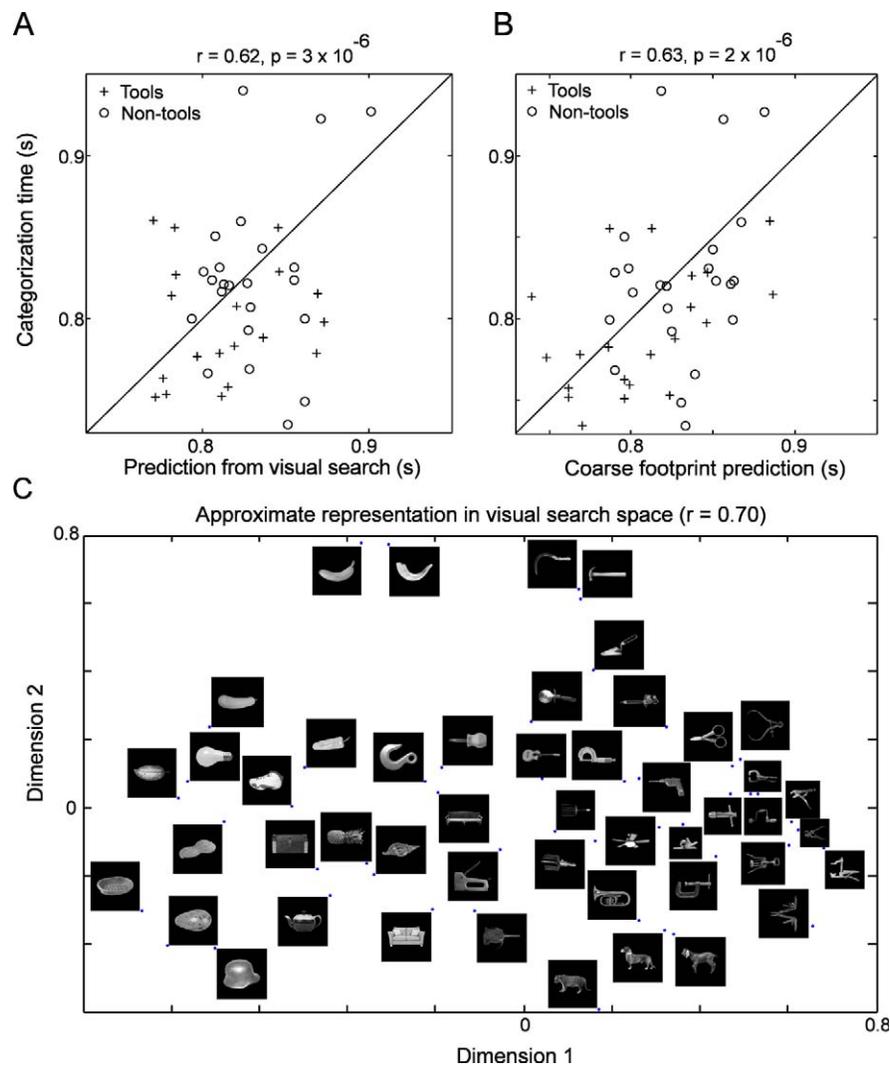


Figure 6. Tools categorization task (Experiment 4). Subjects were asked to categorize an object as a tool or not. (A) Correlation plot of categorization times for each item in the tools task against predictions using visual search data. Plus symbols represent reaction times of all tools, and circles represent reaction times of all nontools. (B) Correlation between observed categorization times in the tools task against predictions using coarse footprint with the same conventions as in Figure 6A. (C) Approximate representation of tools and nontools in visual search space, showing separable clusters according to category. Some images are scaled down to accommodate them in the plot.

we chose for further analysis a low-level feature that was most likely to vary between categories, namely aspect ratio.

We set out to test whether the aspect ratio of an object could explain categorization performance in each task. For each category, we trained a linear classifier to predict the category of each item. Classifiers trained on aspect ratio performed poorly compared to those based on visual search or coarse footprint (classifier accuracy using aspect ratio: 62% for animals, 50% for animals varying in view, 63% for vehicles, and 63% for tools; Table 4). We then used pair-wise similarity measures (calculated as the reciprocal of the absolute difference in aspect ratio) as before to predict

categorization times. Models based on aspect ratio similarity yielded predictions that were positively and significantly correlated with categorization times on all tasks (correlations: $r = 0.35$, $p = 0.001$ for animals; $r = 0.55$, $p = 5 \times 10^{-5}$ for animals varying in view; and $r = 0.69$, $p = 4 \times 10^{-8}$ for vehicles) with the sole exception of tools ($r = 0.18$, $p = 0.2$). These correlations were generally lower than those obtained using visual search or coarse footprint (for animals and tools) except in the case of vehicles where they compare favorably. Although we cannot rule out aspect ratio altogether, the higher correlations and accuracy obtained using coarse structure suggests that coarse structure is the most likely candidate.

Discussion

Here we measured the time taken by humans to categorize objects across three diverse categories (animals, vehicles, and tools), as well as across three hierarchical levels in the animal category: superordinate (animal), basic (dog), and subordinate (Labrador). We hypothesized that categorization of an object could be explained using its similarity to items within and outside its category. To this end, we used visual search to measure the perceived similarity between the objects used in each task and used this data to predict categorization performance. The main result of this study is that categorization can be predicted using similarity relations as measured by visual search. This result held true across various categories as well as across hierarchical levels. Our model quantitatively accounts for several well-known categorization phenomena using similarity relations alone without invoking verbal, semantic, or hierarchy-related factors. Furthermore, we were able to predict these categorization times using coarse image content. Taken together, our results suggest that visual search and categorization are based on a common underlying object representation that depends on coarse image content. We discuss these findings below in relation to other studies of categorization.

Relation between categorization and visual search

The main result of this study is that visual search can predict rapid visual categorization. Our result is concordant with the view that, although neuronal activity in the visual cortex is modulated by attention and task demands (Maunsell & Treue, 2006), neuronal selectivity remains unaffected (Martinez-Trujillo & Treue, 2004; McAdams & Maunsell, 1999; Suzuki, Matsumoto, & Tanaka, 2006). The finding that search predicts categorization is surprising because they are disparate tasks: one involves scrutiny of individual items (categorization) whereas the other involves scrutiny of several items in an array (visual search). But they may be more similar than they appear because both tasks involve feature matching, in one case between the object and other category/noncategory items (categorization) and in the other, between target and distractors (search). Our results imply that rapid visual categorization and visual search depend on a shared object representation.

Our finding that search times for a category member among noncategory members increase from superordinate to basic to subordinate levels is consistent with a previous report (Large et al., 2004). However we have,

to our knowledge, demonstrated for the first time that visual search similarity accounts for visual categorization. Like most categorization studies, we have only dealt with discriminating one category from all other noncategory members (dogs vs. nondogs). We propose that visual search similarity might predict categorization even on tasks that involve discriminating between two categories (e.g., dogs vs. cars). We note that either outcome may be interesting: discriminating between two categories might be more complex because it involves two category representations; alternatively it may be simpler because both alternatives in the task have low visual variability (Bowers & Jones, 2008). A final point of interest concerns how the results would change with visual expertise: would search and categorization performance be different for dog (or other category) experts, leaving their correlation unchanged? Or would categorization alone change leaving search unaffected? The answers to these intriguing questions will clarify the link between categorization and visual search reported here.

Comparisons of performance across categories

One concern regarding comparing performance across categories is that the tasks may involve different degrees of visual variability. However, visual variability may be fundamental to the definition of a category (e.g., Labradors are intrinsically less variable than animals), and elucidating its contribution to categorization will require comparing performance on categories with different degrees of variability. Our model implicitly incorporates this factor into the measures of within- and between-category similarity: if category items are more variable, they will be more easily distinguished among each other, leading to small within-category similarity.

In general, differences in categorization performance may simply arise from differences between the noncategory items rather than the category items. Our model incorporates this factor because it is sensitive to both within- and between-category similarity. However, ruling out this explanation is hard: for instance, keeping identical distractors in the animal and dog tasks would result in the dog task becoming equivalent to the animal task for all practical purposes—indeed, recent studies have found identical performance in such equated tasks (Macé et al., 2009). At the other extreme, selecting all distractors to be from a single category (e.g., using dogs as distractors in the Labrador task) may invoke two category representations in the brain, as opposed to only one when the distractors are heterogeneous and belong to no particular category. Although categorization might always be based on

similarity relations, the extent to which it is context, set, or distractor dependent remains to be fully elucidated.

Relation to other studies of categorization

Our finding that superordinate (i.e., animal) categorization is easiest is inconsistent with the classical basic-level advantage where basic level categorization (i.e., dog) is easier than both superordinate and subordinate level classification (Rosch et al., 1976). However, the classic Rosch studies required subjects to make verbal responses, whereas recent studies have required manual (key press) responses indicating the category (Large et al., 2004; Macé et al., 2009; Mack et al., 2009). In all these studies including ours, the advantage of the basic level over the superordinate level is reversed. Thus the classical basic-level advantage over the superordinate level may have arisen due to verbal influences. In contrast, the advantage of the basic level over the subordinate level is a highly robust finding reported in every study including ours (de la Rosa, Choudhery, & Chatziastros, 2011; Grill-Spector & Kanwisher, 2005; Mack et al., 2009; Mack & Palmeri, 2010; Rosch et al., 1976).

Our finding that atypical animals take longer to be categorized as animals is concordant with classic findings regarding atypicality (McCloskey & Glucksberg, 1978; Rosch & Mervis, 1975; Rosch et al., 1976). It is entirely possible, however, that atypicality depends entirely on context; a pigeon would hardly be atypical among birds. Our model accounts for context automatically because both within- and between-category similarity are average measures sensitive to the frequency of occurrence of stimuli within a given set. Our finding of longer categorization times for snakes is discordant with the early detection of fear-relevant stimuli in visual search (Lobue & DeLoache, 2008; Ohman, Flykt, & Esteves, 2001). This may be due to a difference in the two tasks (animal categorization vs. visual search), although whether the early detection of snakes can be truly attributed to their emotional valence rather than their shape or to search asymmetries is not clear (Lipp, Derakshan, Waters, & Logies, 2004). Indeed, our visual search data show no systematic difference between snakes and other atypical animals (Figure 2E).

We have found that for animals, vehicles, and tools, the categorization time for an object depends primarily on its similarity to members outside its category and to a smaller degree on its similarity to members of its own category. This finding is consistent with a previous study demonstrating that dissimilarity (to outside-category members) rather than similarity (to within-category members) determines categorization (Stewart & Morin, 2007). However, their relative contributions towards categorization may differ across tasks—and they do in the categories we tested (Table 1). The

general form of our model is consistent with previously proposed models that predict categorization times based on similarity relations and an explicit category boundary (Hampton, 1998; McKinley & Nosofsky, 1996; Sigala, Gabbiana, & Logothetis, 2002). Our approach differs from these studies in two important respects: we have used naturalistic categorization tasks where both the features and the category boundary are unknown. To resolve the features, we made explicit measurements of similarity using visual search and modeled them using coarse object structure. Rather than explicitly modeling the unknown category boundary, we simply measured the similarity of each object to members in and outside its category. In our formulation, the similarity of an item to both items within or across its category is inversely related to its distance from the category boundary, wherever it may be located. Thus, our data may be accounted for equally well by models that posit increased reaction times near the category boundary (Grinband, Hirsch, & Ferrera, 2006; Maddox et al., 1998; McKinley & Nosofsky, 1996). Distinguishing between these possibilities will require measuring categorization times for stimuli equidistant from the category boundary but differing in within- or between-category distances.

Influence of three-dimensional object view

Our finding that animal categorization is influenced by object view is concordant with reports of view-dependence in the recognition of three-dimensional objects (Buelthoff & Edelman, 1992; Buelthoff, Edelman, & Tarr, 1995; Palmer, 1999; Riesenhuber & Poggio, 2000) and in neuronal object representations in higher visual areas (Freiwald & Tsao, 2010; Logothetis, Pauls, Bülthoff, & Poggio, 1994; Vogels, Biederman, Bar, & Lorincz, 2001). This view dependence might be specific to some categories: objects like animals have elongated, bilaterally symmetric bodies whose distinctive features are maximally visible in a sideways (profile) view, making them easier to categorize in a profile view. In contrast, other objects such as fruits contain distinctive features at all views, suggesting that their categorization is equally easy at all views. Indeed, subjects' performance on animals was view-dependent, whereas their performance on nonanimals was view-independent (Experiment 2). We propose that whether categorization is view-invariant or not will depend on the availability of diagnostic features at different three-dimensional views.

We have also found that in visual search, profile and oblique views of animals form separate clusters, whereas across the set of nonanimals, different views of an object cluster together (Figure 4C). Although at first glance, this appears to be a fundamental difference between animate and inanimate objects, it could arise

simply because animals are more similar to each other than are nonanimals. These possibilities can be distinguished by testing object representations of inanimate and animate objects equated for perceived similarity. We propose that object representations are fundamentally view-dependent but that object identity can be extracted in a view-invariant manner, especially for perceptually distinct objects. Our proposal is consistent with evidence that the neuronal representation of faces (being perceptually similar) is view-dependent in the posterior face patches in monkeys and view-invariant in subsequent stages of processing (Freiwald & Tsao, 2010).

Influence of nonvisual factors on visual categorization

We have found that visual search explains categorization of animals and vehicles (which are defined primarily by visual form) as well as that of tools (which are primarily defined by their motor function). The last result is surprising because we expected motor representations to be activated for tool categorization but purely visual representations to be activated for visual search. However, it is possible that categories such as tools might still have a distinctive visual appearance. For instance, most tools are elongated objects and contain handles for grasping, which could be diagnostic for categorization. In general, the finding that coarse object structure can predict categorization times implies that the underlying representation is primarily visual and does not involve verbal or semantic factors. Although we cannot rule out semantic or verbal contributions to categorization, our results limit their influence as follows. The discrepancy in the quality-of-fit between categorization and similarity based on visual search ($r = 0.85$; Figure 2D) compared to similarity based on coarse structure ($r = 0.68$; Figure 3B) might arise from two sources: an inability of coarse structure to account for visual similarity, or alternatively, from verbal or semantic influences on both visual search and categorization that are absent in the coarse structure model. That coarse structure cannot completely account for visual similarity is evident from a simple example: two out-of-phase checkerboards differ maximally in coarse structure but are perceptually indistinguishable. Nonetheless, coarse structure, as instantiated here, is a first step towards elucidating the object representations that underlie categorization.

Evidence for coarse object representations in vision

Our finding that coarse object structure accounts for categorization is consistent with four lines of evidence.

First, coarse structure is sufficient for categorization: removing other details such as color (Delorme et al., 2000), removing high spatial frequencies (Nandakumar & Malik, 2009), and even reducing objects to silhouettes (Quinn et al., 2001) all have modest effects on categorization. However, the level of coarseness required may be task-dependent (Collin & McMullen, 2005; Harel & Bentin, 2009; Morrison & Schyns, 2001). Indeed, our results suggest that at least for animal and vehicle categorization, the relevant spatial frequencies are similar but that a finer spatial scale is required for tool categorization. Second, visual processing proceeds in a coarse-to-fine manner (Bar, 2003; Bar et al., 2006; Bullier, 2001; Frazor, Albrecht, Geisler, & Crane, 2004; Kveraga, Boshyan, & Bar, 2007; Macé, Delorme, Richard, & Fabre-Thorpe, 2010; Macé, Thorpe, & Fabre-Thorpe, 2005; Morrison & Schyns, 2001; Navon, 1977; Sripathi & Olson, 2009). This early availability of coarse information is consistent with evidence that animal categorization is extremely fast (Rousselet et al., 2002; Thorpe, Fize, & Marlot, 1996) and that it involves feedforward processing (Serre, Oliva, & Poggio, 2007; Thorpe et al., 1996). Third, our results are consistent with evidence that coarse structure influences object representations in visual cortex (Bermudez, Vicente, Romero, Perez, & Gonzalez, 2009; Frazor et al., 2004; Sripathi & Olson, 2009, 2010) and with evidence that neuronal representations in monkey inferotemporal cortex (Kiani, Esteky, Mirpour, & Tanaka, 2007) as well as voxel-based representations in human object-selective cortex (Kriegeskorte et al., 2008) contain category information. Finally, our results accord with suggestions that categorization may be based on features with intermediate complexity since these features are likely to occur at coarse spatial scales (Delorme, Richard, & Fabre-Thorpe, 2010; Fabre-Thorpe, 2011; Ullman, Vidal-Naquet, & Sali, 2002).

Conclusions

Coarse object structure may be important for categorization because it is insensitive to the idiosyncratic finer details (e.g., antlers of a deer) that may be irrelevant for categorization but relevant for identification. At the same time, it is sensitive to large-scale structural differences (e.g., presence of head, body, and legs) that distinguish one category from another. Coarse structure may allow for object category to be decoded sooner than object identity (Hung, Kreiman, Poggio, & DiCarlo, 2005; Matsumoto, Sugase-Miyamoto, & Okada, 2005; Young & Yamane, 1992), but this may be due to differences in discriminability rather than a fundamental temporal separation (Bowers &

Jones, 2008; Grill-Spector & Kanwisher, 2005; Mack, Gauthier, Sadr, & Palmeri, 2008; Mack & Palmeri, 2010, 2011). Separate representations are also unlikely given that different categorization and identification tasks may require different spatial frequencies (Collin & McMullen, 2005; Harel & Bentin, 2009; Morrison & Schyns, 2001). We propose that both coarse and fine structure are integrated into a unified object representation that subserves a wide variety of visual and cognitive tasks.

Acknowledgments

KM was supported by a Kishore Vaigyanik Protsahan Yojana fellowship from the Government of India. This research was funded by a startup grant from the Indian Institute of Science and an Intermediate Fellowship (SPA) from the Wellcome Trust–DBT India Alliance. We thank T. Vighneshvel and Zhivago Kalathuriyan for assistance with collecting data.

Commercial relationships: none.

Corresponding author: S. P. Arun.

Email: sparun@cns.iisc.ernet.in.

Address: Centre for Neuroscience, Indian Institute of Science, Bangalore, India.

References

- Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, 15(4), 600–609.
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Schmidt, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences USA*, 103(2), 449–454.
- Bermudez, M. A., Vicente, A. F., Romero, M. C., Perez, R., & Gonzalez, F. (2009). Spatial frequency components influence cell activity in the inferotemporal cortex. *Visual Neuroscience*, 26(4), 421–428.
- Bowers, J. S., & Jones, K. W. (2008). Detecting objects is easier than categorizing them. *Quarterly Journal of Experimental Psychology (Hove)*, 61(4), 552–557.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.
- Buelthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences USA*, 89(1), 60–64.
- Buelthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex*, 5(3), 247–260.
- Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, 36(2–3), 96–107.
- Collin, C. A., & McMullen, P. A. (2005). Subordinate-level categorization relies on high spatial frequencies to a greater degree than basic-level categorization. *Perception & Psychophysics*, 67(2), 354–364.
- de la Rosa, S., Choudhery, R. N., & Chatziastros, A. (2011). Visual object detection, categorization, and identification tasks are associated with different time courses and sensitivities. *Journal of Experimental Psychology: Human Perception & Performance*, 37(1), 38–47.
- Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. *Vision Research*, 40, 2187–2200.
- Delorme, A., Richard, G., & Fabre-Thorpe, M. (2010). Key visual features for rapid categorization of animals in natural scenes. *Frontiers in Psychology*, 1, 21.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3), 433–458.
- Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology*, 2, 243.
- Frazor, R. A., Albrecht, D. G., Geisler, W. S., & Crane, A. M. (2004). Visual cortex neurons of monkeys and cats: Temporal dynamics of the spatial frequency response function. *Journal of Neurophysiology*, 91(6), 2607–2627.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, 291(5502), 312–316.
- Freiwald, W. A., & Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, 330(6005), 845–851.
- Girard, P., & Koenig-Robert, R. (2011). Ultra-rapid categorization of fourier-spectrum equalized natural images: Macaques and humans perform similarly. *PLoS One*, 6(2):e16453.
- Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*, 52(2), 125–157.
- Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you

- know what it is. *Psychological Science*, 16(2), 152–160.
- Grinband, J., Hirsch, J., & Ferrera, V. P. (2006). A neural representation of categorization uncertainty in the human brain. *Neuron*, 49(5), 757–763.
- Gross, C. G., & Bornstein, M. (1978). Left and right in science and art. *Leonardo*, 11, 29–38.
- Hampton, J. A. (1998). Similarity-based categorization and fuzziness of natural categories. *Cognition*, 65(2–3), 137–165.
- Harel, A., & Bentin, S. (2009). Stimulus type, level of categorization, and spatial-frequencies utilization: Implications for perceptual categorization hierarchies. *Journal of Experimental Psychology: Human Perception & Performance*, 35(4), 1264–1273.
- Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, 310(5749), 863–866.
- Joubert, O. R., Rousselet, G. A., Fabre-Thorpe, M., & Fize, D. (2009). Rapid visual categorization of natural scene contexts with equalized amplitude spectrum and increasing phase noise. *Journal of Vision*, 9(1):2, 1–16, <http://www.journalofvision.org/content/9/1/2>, doi:10.1167/9.1.2. [[PubMed](#)] [[Article](#)]
- Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, 97(6), 4296–4309.
- Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., et al. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6), 1126–1141.
- Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience*, 27(48), 13232–13240.
- Large, M.-E., Kiss, I., & McMullen, P. A. (2004). Electrophysiological correlates of object categorization: Back to basics. *Cognitive Brain Research*, 20(3), 415–426.
- Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences USA*, 99(14), 9596–9601.
- Lipp, O. V., Derakshan, N., Waters, A. M., & Logies, S. (2004). Snakes and cats in the flower bed: Fast detection is not specific to pictures of fear-relevant animals. *Emotion*, 4(3), 233–250.
- Lobue, V., & DeLoache, J. S. (2008). Detecting the snake in the grass: Attention to fear-relevant stimuli by adults and young children. *Psychological Science*, 19(3), 284–289.
- Logothetis, N. K., Pauls, J., Bülthoff, H. H., & Poggio, T. (1994). View-dependent object recognition by monkeys. *Current Biology*, 4(5), 401–414.
- Macé, M. J.-M., Delorme, A., Richard, G., & Fabre-Thorpe, M. (2010). Spotting animals in natural scenes: Efficiency of humans and monkeys at very low contrasts. *Animal Cognition*, 13(3), 405–418.
- Macé, M. J.-M., Joubert, O. R., Nespoulous, J.-L., & Fabre-Thorpe, M. (2009). The time-course of visual categorizations: You spot the animal faster than the bird. *PLoS One*, 4(6), e5927.
- Macé, M. J.-M., Thorpe, S. J., & Fabre-Thorpe, M. (2005). Rapid categorization of achromatic natural scenes: How robust at very low contrasts? *European Journal of Neuroscience*, 21(7), 2007–2018.
- Mack, M. L., Gauthier, I., Sadr, J., & Palmeri, T. J. (2008). Object detection and basic-level categorization: Sometimes you know it is there before you know what it is. *Psychonomic Bulletin & Review*, 15(1), 28–35.
- Mack, M. L., & Palmeri, T. J. (2010). Decoupling object detection and categorization. *Journal of Experimental Psychology: Human Perception & Performance*, 36(5), 1067–1079.
- Mack, M. L., & Palmeri, T. J. (2011). The timing of visual object categorization. *Frontiers in Psychology*, 2, 165.
- Mack, M. L., Wong, A. C.-N., Gauthier, I., Tanaka, J. W., & Palmeri, T. J. (2009). Time course of visual object categorization: Fastest does not necessarily mean first. *Vision Research*, 49(15), 1961–1968.
- Maddox, W. T., Ashby, F. G., & Gottlob, L. R. (1998). Response time distributions in multidimensional perceptual categorization. *Perception & Psychophysics*, 60(4), 620–637.
- Margolis, E., & Laurence, S. (Eds.). (1999). *Concepts: Core Readings*. Cambridge, MA: Bradford Books.
- Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current Biology*, 14(9), 744–751.
- Matsumoto, N., Sugase-Miyamoto, Y., & Okada, M. (2005). Categorical signals in a single-trial neuron activity of the inferotemporal cortex. *NeuroReport*, 16(15), 1707–1710.

- Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neuroscience*, 29(6), 317–322.
- McAdams, C. J., & Maunsell, J. H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *Journal of Neuroscience*, 19(1), 431–441.
- McCloskey, M., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition*, 6, 462–472.
- McKinley, S. C., & Nosofsky, R. M. (1996). Selective attention and the formation of linear decision boundaries. *Journal of Experimental Psychology: Human Perception & Performance*, 22(2), 294–317.
- Minda, J. P., & Smith, J. D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 28(2), 275–292.
- Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin & Review*, 8(3), 454–469.
- Nandakumar, C., & Malik, J. (2009). Understanding rapid category detection via multiply degraded images. *Journal of Vision*, 9(6):19, 1–8, <http://www.journalofvision.org/content/9/6/19>, doi:10.1167/9.6.19. [PubMed] [Article]
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9, 353–383.
- Ohman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General*, 130(3), 466–478.
- Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.
- Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, 460, 94–97.
- Quinn, P. C., Eimas, P. D., & Tarr, M. J. (2001). Perceptual categorization of cat and dog silhouettes by 3- to 4-month-old infants. *Journal of Experimental Child Psychology*, 79(1), 78–94.
- Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience*, 3(Suppl.), 1199–1204.
- Rollenhagen, J. E., & Olson, C. R. (2000). Mirror-image confusion in single neurons of the macaque inferotemporal cortex. *Science*, 287(5457), 1506–1508.
- Rosch, E., & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573–605.
- Rosch, E., Mervis, C., Gray, W., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382–439.
- Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience*, 5(7), 629–630.
- Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences USA*, 104(15), 6424–6429.
- Sigala, N., Gabbiani, F., & Logothetis, N. K. (2002). Visual categorization and object representation in monkeys and humans. *Journal of Cognitive Neuroscience*, 14(2), 187–198.
- Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in primate temporal cortex. *Nature*, 415, 318–320.
- Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition*, 65(2–3), 167–196.
- Sripati, A. P., & Olson, C. R. (2009). Representing the forest before the trees: A global advantage effect in monkey inferotemporal cortex. *Journal of Neuroscience*, 29(24), 7788–7796.
- Sripati, A. P., & Olson, C. R. (2010). Global image dissimilarity in macaque inferotemporal cortex predicts human visual search efficiency. *Journal of Neuroscience*, 30(4), 1258–1269.
- Stewart, N., & Morin, C. (2007). Dissimilarity is used as evidence of category membership in multidimensional perceptual categorization: A test of the similarity-dissimilarity generalized context model. *Quarterly Journal of Experimental Psychology (Colchester)*, 60(10), 1337–1346.
- Suzuki, W., Matsumoto, K., & Tanaka, K. (2006). Neuronal responses to object images in the macaque inferotemporal cortex at different stimulus discrimination levels. *Journal of Neuroscience*, 26(41), 10524–10535.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582), 520–522.
- Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, 5(7), 682–687.
- Vogels, R., Biederman, I., Bar, M., & Lorincz, A. (2001). Inferior temporal neurons show greater sensitivity to nonaccidental than to metric shape

- differences. *Journal of Cognitive Neuroscience*, 13(4), 444–453.
- Wichmann, F. A., Braun, D. I., & Gegenfurtner, K. R. (2006). Phase noise and the classification of natural images. *Vision Research*, 46(8–9), 1520–1529.
- Wichmann, F. A., Drewes, J., Rosas, P., & Gegenfurtner, K. R. (2010). Animal detection in natural scenes: critical features revisited. *Journal of Vision*, 10(4):6, 1–27, <http://www.journalofvision.org/content/10/4/6>, doi:10.1167/10.4.6. [[PubMed](#)] [[Article](#)]
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews: Neuroscience*, 5(6), 495–501.
- Young, M. P., & Yamane, S. (1992). Sparse population coding of faces in the inferotemporal cortex. *Science*, 256(5061), 1327–1331.