Research Article

Visual Recognition

As Soon as You Know It Is There, You Know What It Is

Kalanit Grill-Spector[1] and Nancy Kanwisher[2]

[1]*Department of Psychology, Stanford University, and* [2]*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology*

ABSTRACT—*What is the sequence of processing steps involved in visual object recognition? We varied the exposure duration of natural images and measured subjects' performance on three different tasks, each designed to tap a different candidate component process of object recognition. For each exposure duration, accuracy was lower and reaction time longer on a within-category identification task (e.g., distinguishing pigeons from other birds) than on a perceptual categorization task (e.g., birds vs. cars). However, strikingly, at each exposure duration, subjects performed just as quickly and accurately on the categorization task as they did on a task requiring only object detection: By the time subjects knew an image contained an object at all, they already knew its category. These findings place powerful constraints on theories of object recognition.*

Humans recognize objects with astonishing ease and speed (Thorpe, Fize, & Marlot, 1996). In the studies we report here, we used behavioral methods to investigate the sequence of processes involved in visual object recognition in natural scenes. We tested two (non-mutually exclusive) hypotheses: (a) that visual object recognition entails first detecting the presence of the object, before perceptually categorizing it (e.g., as bird, car, or flower), and (b) that objects are perceptually categorized (e.g. bird, car) before they are identified at a finer grain (e.g., pigeon, jeep).

Consistent with the first hypothesis, traditional models of object recognition posit an intermediate stage between low-level visual processing and high-level object recognition at which the object is first segmented from the rest of the image before it is recognized (Bregman, 1981; Driver & Baylis, 1996; Nakayama, He, & Shimojo, 1995; Rubin, 1958). Underlying

this idea is the intuition that an efficient recognition system should not operate indiscriminately on any region of an image, because most regions will not correspond to distinct objects. Instead, researchers have argued that stored object representations should be accessed only for candidate regions selected by a prior image-segmentation process. However, other evidence suggests that object recognition may influence, and perhaps even precede, segmentation (Peterson & Gibson, 1993, 1994; Peterson & Kim, 2001). Thus, the first hypothesis, which suggests that segmentation occurs prior to recognition, is currently subject to vigorous debate (Peterson, 1999; Vecera & Farah, 1997; Vecera & O'Reilly, 1998).

Consistent with the second hypothesis, some behavioral evidence suggests that familiar objects are named faster at the basic level (e.g., car; Rosch, 1978; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) than the superordinate (e.g., vehicle) or subordinate (e.g., Volkswagen Beetle) level. However, this is apparently not true for visually atypical members of a category (Jolicoeur, Gluck, & Kosslyn, 1984). Further, it has been suggested that visual expertise may lead experts to recognize stimuli from their expert category as fast at the subordinate level as the basic level (Rosch et al., 1976; Tanaka, 2001). Thus, the generality of the second hypothesis is also subject to debate.

To test whether object detection precedes perceptual categorization and whether perceptual categorization precedes identification, we measured behavioral performance on three different recognition tasks: object detection, object categorization, and within-category identification. We used displays in which each photograph was presented briefly at one of several exposure durations and then immediately masked (Fig. 1). We reasoned that if one task (Task A) requires additional processing not required by another task (Task B), this extra processing could be detected in two different ways. Insofar as masking truncates visual processing (Breitmeyer & Ogmen, 2000), performance should be lower for a given stimulus duration on Task A than on Task B, because the mask will cut off processing before the longer process is completed. However, because the

Address correspondence to Kalanit Grill-Spector, Department of Psychology, Jordan Hall, Stanford University, Stanford, CA 94305; e-mail: kalanit@psych.stanford.edu.
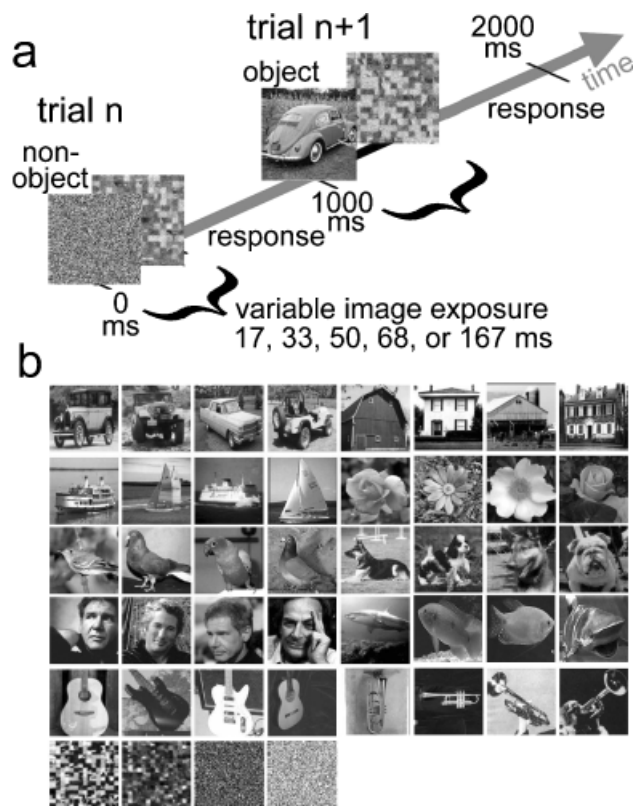
Fig. 1. Example of two trial sequences (a) and the images used in this study (b). Images were presented briefly (for 17, 33, 50, 68, or 167 ms) and were immediately followed by a mask that stayed on for the remainder of the trial, as indicated by the gray bar. During this period, subjects were required to respond according to the task instructions. Trial duration was 1 s in Experiments 2 through 4 (as shown here) and 2 s in Experiment 1.

masking stimulus is unlikely to cut off processing at all stages, we also compared reaction times across tasks. If Task A requires additional processing not required by Task B for the same stimulus and exposure duration, then reaction times (RTs) should be longer for Task A than for Task B (Sternberg, 1998a, 1998b).

In the *object detection* task, participants were asked to decide whether or not a gray-scale photograph contained an object. Catch trials consisted of scrambled versions of the images (Grill-Spector, Kushnir, Hendler, & Malach, 2000) containing textures or random dot patterns (Fig. 1b). Participants were told that they did not have to recognize the object to report its presence. (This is a liberal test of object detection, as performance could in principle be based on lower-level information such as spatial-frequency composition.) In the *object categorization* task, subjects were asked to categorize the object in the picture at the basic level (e.g., car, house, flower). In the *within-category identification* task, subjects were asked to discriminate exemplars of a particular subordinate-level category (e.g., German shepherd) from other members of the category (e.g., other dogs). In each trial in each of our experiments, subjects viewed an image they had never seen before, so performance could not be affected by prior knowledge of particular images.

Further, objects from each category and subordinate class were depicted in various viewing conditions and in different backgrounds to reduce the probability that subjects would use a small set of low-level features to perform these tasks.

METHOD

Sixty-six subjects (31 male and 35 female, ages 19–41) participated in these experiments. All subjects had normal or corrected-to-normal vision and gave written informed consent to participate in the study.

Experimental Design
Each image was presented for 17, 33, 50, 68, or 167 ms and was immediately followed by a mask that stayed on for the remainder of the trial (Fig. 1a). Images were presented centrally using Psychophysics Toolbox (Brainard, 1997) and subtended a visual angle of 8°. The same subjects participated in all three tasks of a given experiment: detection, categorization, and identification. Stimulus order was counterbalanced for exposure duration and content. Task order was counterbalanced across subjects.

Stimuli
The image database contained more than 4,500 gray-level images from 15 basic categories. Each category included at least 200 images of different exemplars (e.g., different birds) along with at least 100 images from one subordinate-level category (e.g., pigeon). Images from each category and subordinate category appeared in many viewing conditions and backgrounds. Nonobject textures (Fig. 1b) were created by scrambling object pictures into 225 random squares with a size of 8×8 pixels (Experiments 1 and 3) or 14,400 squares with a size of 1×1 pixels (Experiments 2 and 4).

Behavioral Performance
Accuracy scores were corrected for guessing (Green & Swets, 1966): accuracy (corrected for guessing) = 100 ∗ (hits − false alarms)/(1− false alarms).

EXPERIMENT 1: NAMING OBJECTS AT DIFFERENT LEVELS OF SPECIFICITY

In Experiment 1, we measured accuracy on the object detection, categorization, and identification tasks performed on the same natural images (Fig. 1). Fifteen subjects viewed 600 images from 10 object categories and 600 random masks across the three tasks. In each of the tasks, the frequency of each object category was 10%, and for each category, half of the images were from a single subordinate class. In each 2-s trial, an image was presented for one of five different exposure durations and was immediately followed by a masking stimulus for the remainder of the trial. For each task, subjects were presented with

200 images (40 per exposure duration) and 200 random masks. Before each task, subjects were told the level of specificity of the required answers and the response alternatives for that task. For the detection task, subjects pressed one key if the picture contained an object (50% of trials), and another key if it contained a texture with no object (50% of trials). For the categorization task, subjects viewed the same object images and named them at the basic level using the following 10 alternatives: face, bird, dog, fish, flower, house, car, boat, guitar, or trumpet. For the within-category identification task, subjects viewed the same object stimuli and named the following pre-specified targets at the subordinate level: Harrison Ford, pigeon, German shepherd, shark, rose, barn, VW beetle, sailboat, and electric guitar; for other exemplars of the categories (e.g., if a picture contained a dog other than a German shepherd), they were instructed to respond "other."

Results

Figure 2 shows accuracy as a function of stimulus duration for all three tasks. Note that the performance curve for the identification task is shifted to the right of the performance curves for the other two tasks. Accuracy in both the detection and categorization tasks was statistically significantly higher than accuracy in identification for the 33-, 50-, and 68-ms exposure durations, $t(15) > 4.1$, $p < .001$, $d > 2$. Accuracy was lower for identification than for categorization for each of the object categories tested.

Surprisingly, the curves relating accuracy to stimulus duration were nearly identical and not significantly different for the categorization and detection tasks, despite the greater complexity of the 10-alternative forced-choice categorization task compared with the two-alternative forced-choice object detection task. Performance in the categorization task was similar to categorization performance in previous experiments (Grill-Spector et al., 2000) in which subjects were not told in advance the object categories, so prior knowledge of the possible categories is unlikely to have been critical for producing these results. Hence, object detection accuracy was not higher than object categorization accuracy at any exposure duration.

Discussion

Our data show strikingly similar performance on object detection and object categorization. Two alternatives may account for this surprising result. One is that detection and categorization require the same amount of processing time. Another is that the same amount of stimulus information is necessary for detection and categorization, but categorization requires additional processing. According to the latter hypothesis, RTs should be longer in the categorization task than in the detection task even when accuracy is similar. Our first experiment is not useful in testing this hypothesis because the different tasks had different numbers of response alternatives, a factor that is known to affect RT (Sternberg, 2001). We therefore conducted a second experiment using the same design except that only two response alternatives were used in each task and the proportions of targets and nontargets were equated across tasks.

EXPERIMENT 2: COMPARISON OF DETECTION, CATEGORIZATION, AND IDENTIFICATION PERFORMANCE WITH A TWO-ALTERNATIVE FORCED-CHOICE DESIGN

In Experiment 2, we measured both accuracy and RT for the three tasks. To examine the specificity of categorization that occurs together with detection, we compared subjects' performance when they were asked to categorize objects within the same superordinate category (e.g., cars vs. boats and planes) with their categorization performance when the objects were
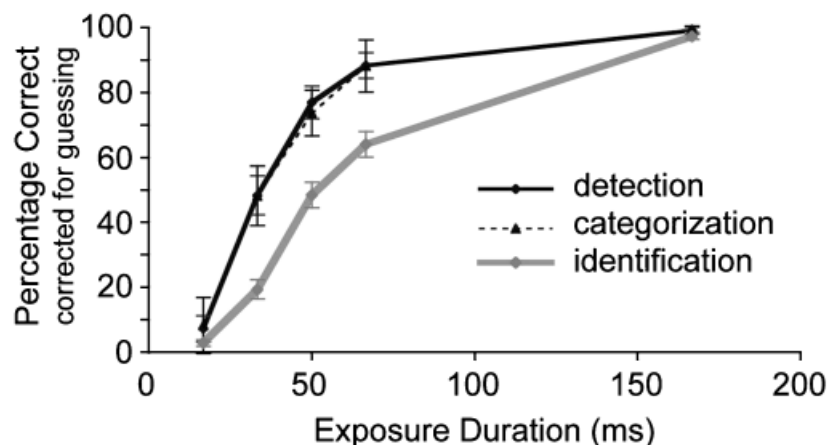

Fig. 2. Naming performance for the three recognition tasks in Experiment 1. The data are averaged across 15 subjects (5 male, 10 female). The *y*-axis denotes accuracy corrected for guessing. Error bars indicate standard errors of the means.

from different superordinate categories (e.g., cars vs. objects excluding vehicles).

Experimental Design

Methods were the same as for Experiment 1 except as follows: (a) We collected both accuracy and RT data, (b) each task was a two-alternative forced-choice task in which 50% of trials contained targets and 50% contained nontargets, (c) three categories were tested (cars, dogs, and guitars) in separate blocks, and (d) trial duration was 1 s instead of 2 s.

In the object detection task, on half of the trials with objects the object belonged to the target category that was used in the categorization and identification tasks, and on the other half the objects were from nine other familiar object categories. Because we wanted to compare performance on the same stimuli across tasks, we report detection performance only for the target category that was tested in the other two tasks in the same block.

In the categorization task, subjects were asked whether each image was from the target category or not (e.g., "car" or "not a car"). For each of the three target categories (cars, guitars, and dogs), subjects participated in two blocks, one in which nontargets were objects from nine other familiar categories, but not from the same superordinate category as the targets, and one in which nontarget objects were from the same superordinate category as the targets. In the latter case, subjects had to distinguish (a) cars versus boats and planes, (b) guitars versus pianos and trumpets, and (c) dogs versus birds and fish.

In the identification task, subjects were asked to determine whether or not each image was an exemplar of the within-category target (e.g., jeep). Distractors were other exemplars from the same basic-level category (e.g., different car models). Half of the images were targets, and half were distractors. Note that subjects had to identify a within-category target, not a particular image. We tested three categories: (a) jeep versus other cars, (b) electric guitar versus other guitars, and (c) German shepherd versus other dogs.

Results

This experiment replicated Experiment 1 in that accuracy on the detection task (i.e., object vs. texture) and accuracy on the categorization task (e.g., car vs. other vehicles and car vs. objects from other superordinate categories) were similar (Fig. 3), whereas accuracy in the identification task (e.g., jeep vs. other cars) was lower. However, crucially, the new experiment found further that not only accuracy but also RTs were virtually identical for the detection and categorization tasks (Fig. 3), all $ts(15) < 1$, $ps > .07$, $ds < 0.25$. In contrast, RTs were longer for the identification task, even when accuracy in categorization and identification were matched (33–68 ms), $ts(15) > 2.95$, $ps < .01$, $ds > 1.8$.

Our results also demonstrate that categorization performance was virtually identical to detection performance even when nontargets were restricted to the same superordinate category (Fig. 3). For all categories, there was no difference in accuracy for detection and categorization when distractor objects were
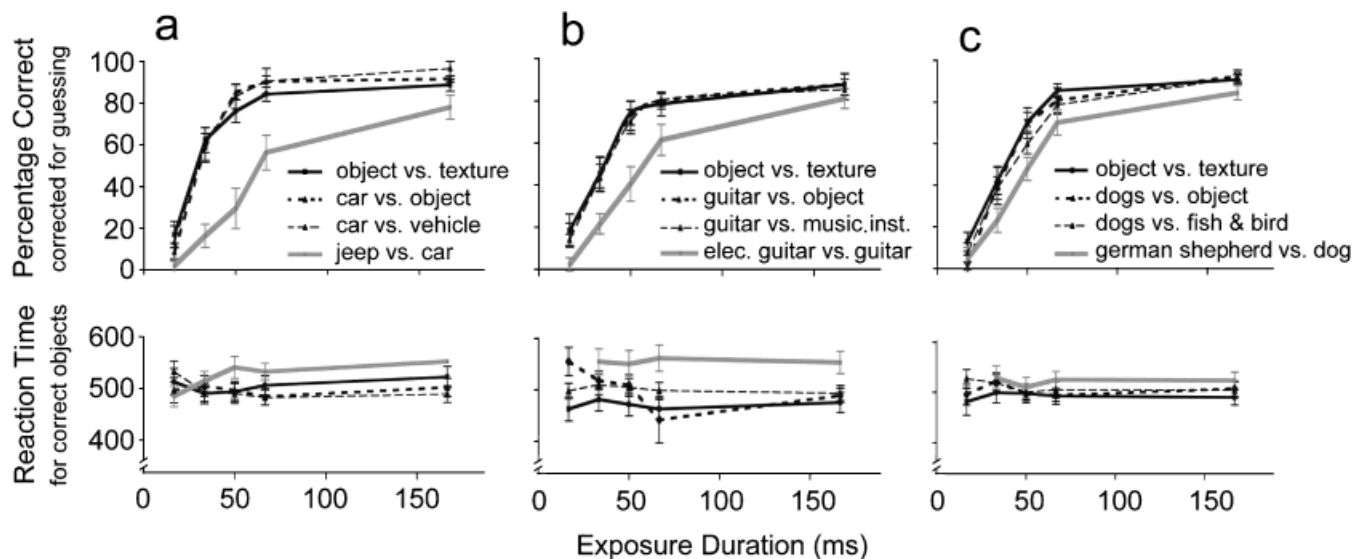

Fig. 3. Recognition performance within and across superordinate categories in Experiment 2. Subjects performed three tasks: detection (e.g., object vs. texture), categorization (e.g., car vs. other vehicle, car vs. other object), and identification (e.g., jeep vs. other car). The graphs present accuracy data (corrected for guessing) and reaction time (RT) data on correct trials for three kinds of categories: (a) vehicles, (b) musical instruments ("music.inst."), and (c) animals. The data are averaged across 15 subjects (9 male, 6 female). Error bars indicate standard errors of the means. RTs are not meaningful when accuracy is at chance. Therefore, RTs are not plotted for the identification task at the 17-ms exposure duration for the guitar and dog blocks. Data for the detection task (object vs. texture) are plotted only for the corresponding basic-level category in each panel; for example, (a) presents the data for cars, which accounted for half the object trials in the detection task.

restricted to the superordinate category. The only exception was higher detection accuracy than accuracy in categorization of dogs versus animals at the 50-ms exposure duration, $t(15) > 2.9$, $p < .02$. There was also no difference in RTs between detection and categorization within the same superordinate category, with two exceptions: guitars versus musical instruments for the 17-, 50-, and 68-ms durations, $ts(15) > 2.9$, $ps < .01$, and dogs versus other animals at the 17- and 33-ms exposure durations, $ts(15) > 2.6$, $ps < .03$. Thus, subjects extracted object categories quite accurately.

Discussion

This experiment demonstrates that object detection and object categorization take the same amount of processing time. The category information extracted during detection is slightly coarser than basic-level information, but considerably finer than superordinate-level information.

EXPERIMENT 3: WAS DETECTION PERFORMANCE BASED ON OBJECT CATEGORY INFORMATION?

Our results consistently show that categorization and detection performance are similar. A straightforward interpretation of these results is that these two processes are linked. However, an alternative account is that detection and categorization are distinct and the observed linkage arose because subjects used object category information in the detection task. One possibility is that the masking stimulus obliterated low-level visual representations, forcing subjects to rely on high-level representations to perform the detection task. If this account is correct, then detection performance should be superior to categorization performance for unmasked stimuli.

We tested this prediction in Experiment 3, in which stimuli were followed by an equiluminant blank screen instead of a masking pattern. Methods were otherwise identical to those of Experiment 2.

Results

Because stimuli were not masked, accuracy for detection and categorization was at ceiling and did not vary significantly with exposure duration (Fig. 4). There were no statistically significant differences between detection and categorization in RT or accuracy for any of the image exposures at all durations, $ts(24) < 1.4$, $ps > .1$, $ds < 0.3$. In contrast, RTs were significantly slower, by approximately 100 ms, for identification than for both detection and categorization at all durations, $ts(24) > 4.5$, $ps < .001$, $ds > 1.2$. Accuracy in both the detection and the categorization tasks was also higher than accuracy in identification at all durations, $ts(24) > 2.8$, $ps < .01$ $ds > 1$. Therefore, detection performance and categorization performance were similar in both accuracy and RTs even when stimuli were not masked, indicating that the apparent similarity of process-
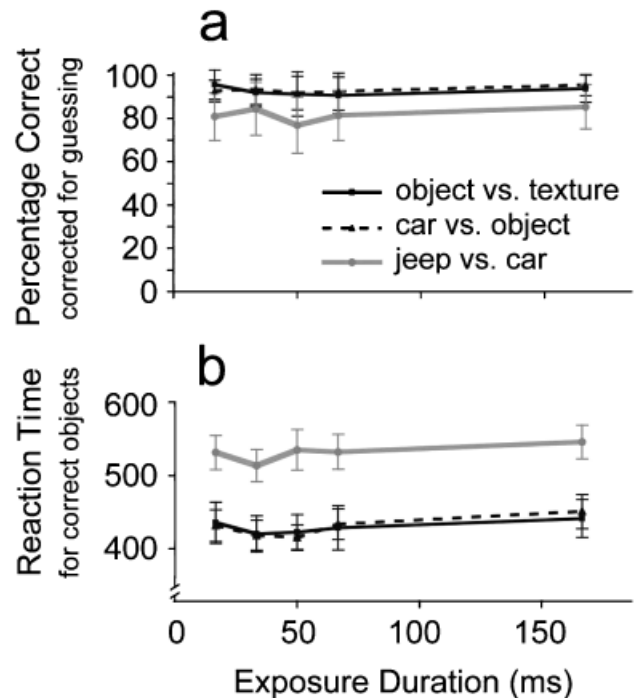

Fig. 4. Detection (object vs. texture), categorization (car vs. object), and identification (jeep vs. car) performance on unmasked stimuli in Experiment 3. Data were measured for 24 subjects (10 male, 14 female). Error bars indicate standard errors of the means. The data plotted here are from the experimental block in which we examined performance on car stimuli. Performance was measured by both accuracy corrected for guessing (a) and reaction time on correct trials (b).

ing time required for detection and categorization in the previous experiments was not an artifact of masking.

Discussion

Experiments 1 through 3 provide evidence that detection and categorization performance require the same amount of information and processing time. Two possible mechanisms might account for this result: (a) Detection and categorization may be mediated by the same mechanism, or (b) detection and categorization may be computed by distinct mechanisms but require similar total amounts of processing. We tested these hypotheses in the next experiment by investigating whether detection and categorization are correlated on a trial-by-trial basis, or whether either task can be successfully performed without the other on a given trial.

EXPERIMENT 4: COMPARING PERFORMANCE IN TWO TASKS ON A TRIAL-BY-TRIAL BASIS

If detection and categorization are directly linked, then success (or failure) at detection will predict success (or failure) at categorization on a trial-by-trial basis, and vice versa. However, if detection and categorization are computed independently, then detection and categorization performance might not show trial-by-trial correlations. To test these predictions, we modified the

experimental paradigm so that subjects made two independent responses on each trial.

Experimental Design

The trial sequence consisted of an image that appeared for 17 (or 33) ms, a masking stimulus (which consisted of a texture pattern created by dividing object images into 225 squares and then scrambling the squares) that was shown for 500 ms, a second image that appeared for 17 (or 33) ms, and another masking stimulus that was shown for 2,966 (or 2,934) ms. In each trial, only one of the pictures contained an object, and the other was a random dot pattern. The 17- and 33-ms exposures were run in separate blocks. Each block contained 128 trials.

In the *detection and categorization* version of the experiment, subjects were asked in which interval (first or second) the object appeared (detection task) and whether the object was a car or a face (categorization task). The objects were cars in half of the trials and faces in the other half. Objects occurred with equal probability in the first and second intervals.

In the *detection and identification* version of the experiment, subjects decided on each trial in which interval (first or second) a face appeared (detection task) and whether the face was Harrison Ford or a different man (identification task). Half the trials contained different pictures of the target individual, and the other half contained pictures of other male faces (some were the faces of famous actors). Male faces appeared with equal probability in the first and second intervals.

The order of the responses within a trial, the order of the two versions of the experiment, and the order of 33-ms and 17-ms blocks were counterbalanced across subjects.

Results

Categorization performance was significantly better for objects that were detection hits than for those that were detection misses (see Figs. 5a and 5b), $t(12) = 4.5$, $p < .001$, $d = 1.2$, for the 17-ms exposure duration and $t(12) = 3.6$, $p < .003$, $d = 1.3$, for the 33-ms exposure duration. Categorization performance on detection misses was not different from chance, $t(12) = 0.5$, $p > .1$, for the 17-ms exposure duration. Crucially, the converse was also true: Detection performance was significantly better for categorization hits (on objects) than for categorization misses (on objects), $t(12) = 4.6$, $p < .001$, $d = 1.7$, for the 17-ms exposure duration and $t(12) = 4.45$, $p < .001$, $d = 1.6$, for the 33-ms exposure duration; also, detection performance was at chance for categorization misses. A two-way analysis of variance of performance as a function of task (detection or categorization) and success (hit or miss in the second task) showed a main effect of success, $F(1, 1) > 12$, $p < .003$, for the 17-ms exposure duration and $F(1, 1) > 17$, $p < .001$, for the 33-ms exposure duration, but there was no significant difference between tasks or interaction between task and
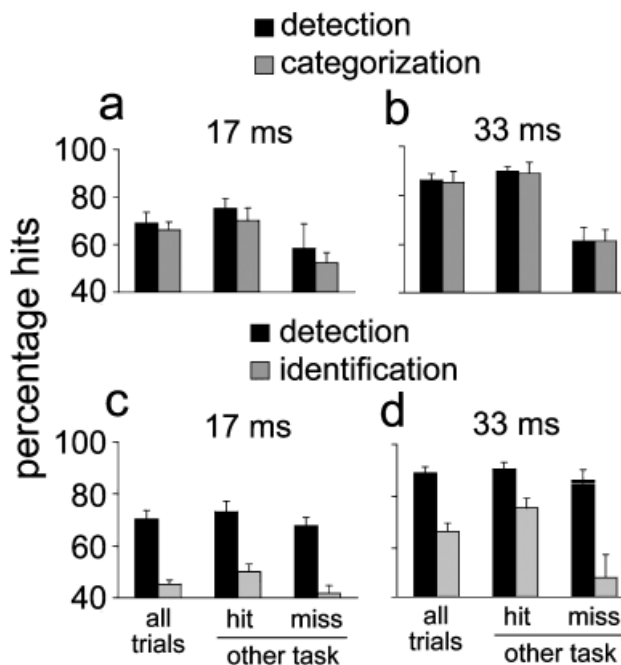

Fig. 5. Experiment 4 results: raw hit rate in one task both overall and as a function of success (hit) or failure (miss) at the second task. The data are averaged across 12 subjects (7 male, 5 female). Chance level is 50%. Results for detection and categorization (a, b) and for detection and identification (c, d) are shown separately for 17-ms and 33-ms exposure durations.

success at the other task. Thus, success on each task predicted success on the other task.

Comparison between face detection and identification within the same trial revealed completely different results (Figs. 5c and 5d). First, detection performance was significantly higher than identification performance, $t(12) = 3.9$, $p < .01$, $d = 1.5$, for the 17-ms exposure duration and $t(12) = 4.5$, $p < .001$, $d = 2$, for the 33-ms exposure duration. Second, identification performance depended on detection performance, but detection did not depend on identification. Thus, identification performance was better for detection hits than for detection misses, $t(12) = 2.9$, $p < .02$, $d = 1.3$, for the 33-ms exposure duration (at 17 ms, identification performance was at chance), but detection performance was not different for identification hit or miss trials, both $ts(12) < 1.5$, $ps > .1$. A two-way analysis of variance of performance on one task as a function of hit or miss at the other task revealed an interaction at the exposure of 33 ms, $F(1, 1) > 5.6$, $p < .03$.

Overall, these findings indicate that detection and categorization are linked, whereas detection occurs prior to identification.

GENERAL DISCUSSION

The same two phenomena occurred with striking consistency in these experiments: (a) Subjects did not require more processing time for object categorization than for object detection, whereas (b) comparable performance on the identification task required

substantially more processing time than was required for either detection or categorization.

Our data provide evidence against the hypothesis that objects are detected before they are recognized. First, in none of our experiments did object categorization require either longer stimulus durations or longer processing time than object detection. Instead, as soon as subjects could detect an object at all, they already knew its category. The level of categorization that occurred with object detection was slightly more crude than the traditional basic level (Rosch et al., 1976), but considerably finer than the superordinate level (Rosch et al., 1976). Second, if object detection is prior to categorization, on some trials objects should be correctly detected but not categorized, whereas the opposite should not occur. This prediction was not upheld: On trials when categorization performance failed, detection performance was no better than chance (the opposite was also true). These data suggest that detection does not occur prior to and independently of categorization. Instead, detection and categorization are apparently linked: When either process fails on a given trial, so does the other.[1]

Because figure-ground segregation should be sufficient for accurate performance on our object detection task, our findings challenge the traditional view that figure-ground segregation precedes object recognition (Bregman, 1981; Driver & Baylis, 1996; Nakayama et al., 1995; Rubin, 1958) and suggest instead that categorization and segmentation are closely linked. This conclusion is consistent with the findings of Peterson and her colleagues (Peterson, 2003; Peterson & Gibson, 1993, 1994; Peterson & Kim, 2001; Peterson & Lampignano, 2003), although our conclusions differ slightly from theirs: Whereas Peterson and her colleagues concluded that categorization influences segmentation, we suggest that conscious object segmentation and categorization are based on the same mechanism. A recent computational model (Borenstein & Ullman, 2002) suggests one way such a linkage between segmentation and categorization may arise. If incoming images are matched to templatelike image fragments (learned from real-world experience with objects) in which each subregion of each fragment is labeled as either figure or ground, the resulting fragment-based representation of an object would contain information about both the object category and the figure-ground segmentation of the image.

An alternative account of our finding of similar performance for object detection and categorization invokes constraints on perceptual awareness (Hochstein & Ahissar, 2002). According to this account, object detection may occur prior to categorization, but the conscious decision stage may have access only to the output of the categorization stage. Neural measurements may ultimately provide the best test between an account of our

data in terms of constraints on awareness and an account in terms of the sequence of processing in object recognition. Preliminary evidence from magnetoencephalography (MEG) and event-related potentials favors the idea that object segmentation and categorization occur at the same time (Halgren, Mendola, Chong, & Dale, 2003; Liu, Harris, & Kanwisher, 2002).

Performance in the detection task and performance in the categorization task were similar, but comparable performance in the identification task always required longer exposures and more processing time. On average, 65 more milliseconds were necessary for identification than for categorization even when accuracy in the categorization and identification tasks was matched. Further, success at identification depended on success at detection, but success at detection did not depend on success at identification. These results indicate that identification occurs after the category has been determined. This finding was obtained not only for objects but also for faces, and is consistent with prior findings from MEG (Liu et al., 2002), but not with claims that expertise leads to a change in the initial level of perceptual categorization of stimuli, such as faces, on which subjects have gained expertise (Rosch et al., 1976; Tanaka, 2001).

From these behavioral data, we cannot determine whether the extra time needed for identification compared with categorization reflects the engagement of a different mechanism or simply a longer engagement of the same mechanism. Some evidence for the latter view comes from neural measures. First, functional magnetic resonance imaging (fMRI) studies in humans have shown that the same cortical regions are engaged in the detection and the identification of stimuli of a given category (Grill-Spector, 2003, Grill-Spector, Knouf, & Kanwisher, 2004). Second, electrophysiological studies in monkeys have shown that stimulus selectivity of neurons in higher-order visual areas increases as exposure duration increases (Keysers, Xiao, Foldiak, & Perrett, 2001; Kovacs, Vogels, & Orban, 1995; Sugase, Yamane, Ueno, & Kawano, 1999; Tamura & Tanaka, 2001). It is possible that the initial neuronal responses are sufficient for detection and categorization, and later neural responses are necessary for identification.

From a computational point of view, capturing an object's category rapidly may expedite identification by restricting processes that match the input with an internal representation to the relevant category (instead of requiring a search across all internal object representations). Traditional psychophysical analyses, usually applied to simpler stimuli, offer a useful perspective here. Graham (1989) has shown that if detection and discrimination (categorization) performance are based on the outputs of the same perceptual analyzers, then categorization performance can be equivalent to or even better than detection performance whenever the two discriminanda engage independent analyzers. This analysis suggests that the present data can be explained in terms of a system in which (a) object

---

[1]There are probably some extreme conditions in which detection can occur without categorization, but these may be a special case of data-limited conditions (e.g., blurry images), rather than resource-limited conditions (Norman & Bobrow, 1975).

detection and object categorization performance are based on the same perceptual analyzers, which would be consistent with evidence from fMRI (Grill-Spector, 2003; Grill-Spector et al., 2004), and (b) categorization of different basic-level categories engages largely independent and nonoverlapping perceptual analyzers (in contrast to recent claims by Haxby et al., 2001), but (c) identification of different stimuli within a category engages overlapping perceptual analyzers.

In sum, we have shown that although substantially more processing is required to precisely identify an object than to determine its general category, it takes no longer to determine an object's category than to simply detect its presence. Overall, these findings provide important constraints for any future theory of object recognition.

REFERENCES

Borenstein, E., & Ullman, S. (2002, August). *Class-specific top-down segmentation*. Paper presented at the European Conference on Computer Vision, Glasgow, Scotland.

Brainard, D.H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Bregman, A.L. (1981). Asking the "what for" question in auditory perception. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization* (pp. 99–118). Hillsdale, NJ: Erlbaum.

Breitmeyer, B.G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perceptual Psychophysics*, *62*, 1572–1595.

Driver, J., & Baylis, G.C. (1996). Edge-assignment and figure-ground segmentation in short-term visual matching. *Cognitive Psychology*, *31*, 248–306.

Graham, N. (1989). *Visual pattern analyzers*. New York: Oxford University Press.

Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics*. New York: John Wiley & Sons.

Grill-Spector, K. (2003). The functional organization of the ventral visual pathway and its relationship to object recognition. In N. Kanwisher & J. Duncan (Eds.), *Attention and performance XX: Functional brain imaging of visual cognition* (pp. 169–193). London: Oxford University Press.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, *7*, 555–562.

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience*, *3*, 837–843.

Halgren, E., Mendola, J., Chong, C.D., & Dale, A.M. (2003). Cortical activation to illusory shapes as measured with magnetoencephalography. *NeuroImage*, *18*, 1001–1009.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–2430.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*, 791–804.

Jolicoeur, P., Gluck, M.A., & Kosslyn, S.M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, *16*, 243–275.

Keysers, C., Xiao, D.K., Foldiak, P., & Perrett, D.I. (2001). The speed of sight. *Journal of Cognitive Neuroscience*, *13*(1), 90–101.

Kovacs, G., Vogels, R., & Orban, G.A. (1995). Cortical correlate of pattern backward masking. *Proceedings of the National Academy of Sciences, USA*, *92*, 5587–5591.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, *5*, 910–916.

Nakayama, K., He, Z.J., & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In S.M. Kosslyn & D.N. Osherson (Eds.), *An invitation to cognitive science: Visual cognition* (pp. 1–70). Cambridge, MA: MIT Press.

Norman, D.A., & Bobrow, D.G. (1975). On data-limited and resource limited processes. *Cognitive Psychology*, *7*, 44–64.

Peterson, M.A. (1999). What's in a stage name? Comment on Vecera and O'Reilly (1998). *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 276–286.

Peterson, M.A. (2003). Overlapping partial configurations in object memory: An alternative solution to classic problems in perception and recognition. In M.A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects and scenes: Analytic and holistic processes* (pp. 269–294). New York: Oxford University Press.

Peterson, M.A., & Gibson, B.S. (1993). Shape recognition contributions to figure-ground organization in three-dimensional display. *Cognitive Psychology*, *25*, 383–429.

Peterson, M.A., & Gibson, B.S. (1994). Must shape recognition follow figure-ground organization? An assumption in peril. *Psychological Science*, *5*, 253–259.

Peterson, M.A., & Kim, J.H. (2001). On what is bound in figures and grounds. *Visual Cognition*, *8*, 329–348.

Peterson, M.A., & Lampignano, D.W. (2003). Implicit memory for novel figure-ground displays includes a history of cross-border competition. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 808–822.

Rosch, E. (1978). Principles of categorization. In B.E. Rosch & B.B. Lloyd (Eds.), *Cognition and categorization* (pp. 28–49). Hillsdale, NJ: Erlbaum.

Rosch, E., Mervis, C.B., Gray, M.D., Johnson, D.M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Rubin, E. (1958). Figure and ground. In D.C. Beardslee & M. Wertheimer (Eds.), *Readings in perception* (pp. 194–203). New York: Van Nostrand.

Sternberg, S. (1998a). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Eds.), *An invitation to cognitive science: Methods, models and conceptual results* (Vol. 4, pp. 703–863). Cambridge, MA: MIT Press.

Sternberg, S. (1998b). Inferring mental operations from reaction-time data: How we compare objects. In D. Scarborough & S. Sternberg (Eds.), *An invitation to cognitive science: Methods, models and conceptual results* (Vol. 4, pp. 365–454). Cambridge, MA: MIT Press.

Sternberg, S. (2001). Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychologica*, *106*, 147–246.

Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, *400*, 869–873.

Tamura, H., & Tanaka, K. (2001). Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cerebral Cortex*, *11*, 384–399.

Tanaka, J. (2001). The entry point of face recognition: Evidence for face expertise. *Journal of Experimental Psychology: General*, *130*, 534–543.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.

Vecera, S.P., & Farah, M.J. (1997). Is visual image segmentation a bottom-up or an interactive process? *Perceptual Psychophysics*, *59*, 1280–1296.

Vecera, S.P., & O'Reilly, R.C. (1998). Figure-ground organization and object recognition processes: An interactive account. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 441–462.