Visual Recognition: as soon as you know it is there, you know what it is.

Kalanit Grill-Spector¹ & Nancy Kanwisher²

¹Department of Psychology, Stanford

²Department of Brain and Cognitive Sciences, MIT

December 16th, 2003

Communication Author:

Kalanit Grill-Spector

Dept. of Psychology

Jordan Hall

Stanford University

Stanford, CA 94305

Tel: 1-650-725-2457

Fax: 1-650-725-5699

Email: kalanit@psych.stanford.edu

ABSTRACT

What is the sequence of processing steps involved in visual object recognition? We varied the exposure duration of natural image stimuli 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 versus cars). However, strikingly, subjects performed just as quickly and accurately at each exposure duration 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.

INTRODUCTION

Humans recognize objects with astonishing ease and speed (Thorpe, Fize, & Marlot, 1996). 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: (i) that visual object recognition entails first detecting the presence of the object, before perceptually categorizing it (e.g., bird, car, flower) and (ii) 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, 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 such 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 segmentation, and may 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 (Rosch, 1978; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) (e.g., car) than the super-ordinate (e.g., vehicle) or subordinate level (e.g., Volkswagen Beetle). 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 at the subordinate level as fast 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 (Hypothesis 1) and whether perceptual categorization precedes identification (Hypothesis 2) we measured behavioral performance on three different recognition tasks: object detection, object categorization, and within-category identification. We used displays in which photographs were presented briefly at one of several stimulus 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 Task B, because the mask will cut off processing before the longer process is completed. However, because the masking stimulus is unlikely to cut off processing at all stages we also compared reaction times across tasks. Here the logic is that if Task A requires additional processing not required by anot required by Task B for the same stimulus and exposure duration, then reaction times should be longer for Task A than Task B (Sternberg, 1998a, 1998b).

In the *object detection* task, participants were asked to decide whether a gray scale photograph contained an object or not. Catch trials consisted of scrambled versions of these images (Grill-Spector, Kushnir, Hendler, & Malach, 2000) containing textures or random dot patterns (Fig. 1b-bottom left). Participants were told that they did not have to recognize the object to report its presence. (Note that this is a liberal test of object detection, as performance could in principle be based on lower-level information such as the spatial frequency composition of the images). 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). Importantly, 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

General Methods

A total of 67 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: Images were presented for 17, 33, 50, 68 or 167ms and were immediately followed by a mask that stayed on for the remainder of the trial (Fig. 1). Images were presented centrally on a power Mac using Psychophysics Toolbox (<u>http://color.psych.ucsb.edu/psychtoolbox</u>) and subtended a visual angle of 8°. Subjects' responses were collected from keyboard presses (except for Experiment 1 in which subjects named image content according to task instructions). The same subjects participated in all three tasks of a given experiment: detection, categorization and identification. Stimuli order was counter-balanced for exposure duration and content. Task order was counterbalanced across subjects.

Stimuli: The image database contained 4500 gray level images from 15 basic categories. Each category included at least 200 images of different exemplars of that category (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. Non-object textures (Fig.1b-bottom) were created by scrambling object pictures into 225 random squares of size 8x8 pixels (Experiments 1 and 3) or 14400 squares of 1x1 pixels (Experiments 2 and 4).

Behavioral Performance: Accuracy scores were corrected for guessing: $accuracy_{corrected} = 100 * \frac{hits - false_alarms}{1 - false_alarms};$ hits: "x" present, subject responded "x";

false_alarms: "x not present", subject responded "x.

Experiment 1: Naming Objects from Ten Categories at Different Levels of Specificity

In Experiment 1 we measured accuracy on object detection, categorization, and identification performed on the same natural images (Fig. 1). 15 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 category was 10% (of all object stimuli) and for each category half of the images were from a single subordinate class. In each trial an image was presented for one of five different exposures and was immediately followed by a masking stimulus for the remainder of a 2s trial (Fig. 1). For each task, subjects were presented with 200 images (40 per exposure) and 200 random masks. Subjects were informed before each task the level of specificity of the required answers, and the response alternatives for that task. They were told that if they were unsure of the answer they must guess.

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 from the following ten alternatives: face, bird, dog, fish, flower,

house, car, boat, guitar, or trumpet. Synonyms were treated as correct answers. 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, versus other exemplars from the same category.

Results

Figure 2 shows that accuracy on the identification task as a function of stimulus duration is shifted to the right of the performance curves for the other two tasks. Accuracy in both detection and categorization tasks was statistically significantly higher than identification for 33, 50 and 68ms exposure durations (all ps<0.001, t-test). The lower accuracy for identification compared to categorization at short exposures occurred 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 ten-alternative forced choice categorization task compared to the two-alternative forced choice object detection task (all ps>0.01, t-test). Performance in the categorization task is similar to 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 be critical for obtaining these results. Hence, object detection accuracy is not higher at each exposure duration than object categorization accuracy.

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 may be necessary for detection and categorization, but categorization may require additional processing. The latter hypothesis predicts that reaction times (RT) in the categorization task will be longer than 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 reaction time (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 proportion of targets and nontargets was equated across tasks.

Experiment 2: Comparison between Detection, Categorization and Identification Performance using a Two-Alternative Forced Choice Design.

Here we measured both accuracy and RT for object detection, categorization and identification. 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 super-ordinate category (e.g., cars vs. boats and planes) to their categorization performance when the objects were from different super-ordinate categories (e.g., cars vs. objects excluding vehicles).

Methods were the same as for Experiment 1 except as follows: (i) We collected both accuracy and reaction time data. (ii) Each task was a two-alternative forced choice task in which 50% of trials contained targets and 50% non-targets. (iii) Three categories were tested (cars, dogs, and guitars) and (iv) Trial duration was 1s instead of 2s.

In the *object detection task*, half of the object trials were of the target category that was used in the corresponding categorization and identification tasks and half were objects from nine familiar object categories. Because we wanted to compare performance across tasks on the same stimuli, we report detection performance only on the target category that was tested in the other two tasks. In the *categorization task*, subjects were asked whether the image was from the target category or not (e.g., "car" or "not a car"). We tested three target categories: cars, guitars and dogs. For each target category subjects participated in 2 blocks of this experiment with different non-target objects: (a) Non-targets were objects from nine familiar categories, but not from the same super-ordinate category as the target category or (b) Non-target objects were from the same super-ordinate category as the target category. In the latter case subjects had to distinguish (i) cars versus boats and planes, (ii) guitars versus pianos and trumpets and (iii) dogs versus birds and fish.

In the *identification task*, subjects were asked to determine for each image whether it was the within-category target or not. Distractors were other exemplars from the same basic level category. Half of the images were different pictures of the within-category target (e.g., jeep) and half of the images were other types of objects from the same basic level category (e.g., different car models). Critically, subjects had to identify a within-category target and not a particular image. We tested three categories: (1) jeep versus car, (2) electric guitar versus guitar and (3) German shepherd versus dog. *Results*

This experiment replicated the finding from Experiment 1 that accuracy on the detection and categorization tasks was similar (Fig. 3), whereas accuracy in the identification task was lower (i.e., the curve for the identification task was shifted to the right). However, crucially, the new experiment further found that not only accuracy but also reaction time were virtually identical for the detection and categorization tasks (Fig. 3, all ps>0.07, t-test, n=15). In contrast, reaction times were longer for the identification task, even when accuracy in categorization and identification were matched (all ps<0.01, t-test).

Our results also demonstrate that categorization performance was virtually identical to detection performance even when non-targets were restricted to the same

super-ordinate category (Fig. 3). Thus, subjects extracted object categories quite accurately. The only exception was for categorization of dogs: performance discriminating dogs vs. inanimate objects was similar to subjects' ability to detect dogs (vs. textures), but performance discriminating dogs versus other animals was lower. Accuracy was lower for 50 ms exposures and RT were longer for 17 and 33 ms exposures (Fig. 3c).

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

Experiment 3: Was Detection Delayed to Later Object Recognition Stages?

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 linkage arises because subjects used object category information in the detection task. One possibility is that the masking stimulus may have obliterated lowlevel visual representations, forcing subjects to rely on high-level representations to perform the detection task. If this is the correct account then detection performance should be superior to categorization performance for unmasked stimuli.

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

Results

Because stimuli were not masked accuracy was at ceiling for detection and categorization, and did not vary significantly with exposure duration (Fig.4-left). Importantly, there were no statistically significant differences in RT or accuracy between detection and categorization for any of the image exposures (all durations ps>0.1, t-test, n=24). In contrast, RT for identification were significantly slower than both detection and categorization by approximately 100ms (all ps<0.01, t-test). Accuracy in both detection and categorization tasks was also higher than identification at all durations (all ps<0.01,t-test). Therefore, detection and categorization performance was similar in both accuracy and reaction times even when stimuli were not masked, suggesting that similarity of processing time required for detection and categorization is not an artifact of masking.

Discussion

Experiments 1-3 provide evidence that detection and categorization performance require the same amount of information and processing time. Two possible mechanisms might account for this result: *(i)* detection and categorization are mediated by the same mechanism or *(ii)* detection and categorization are computed by distinct mechanisms, but the total amount of processing is similar in the two tasks. These hypotheses were tested in the next experiment by asking whether detection and categorization are correlated on a trial-by-trial basis, or whether either task can be successful 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 will not show trial-by-trial correlations. To test these predictions, we modified the experimental paradigm such that subjects made two independent responses on each trial. On each trial an image appeared for 17ms, followed by a masking stimulus that was shown for 500ms, a second image then appeared for 17ms, followed by a masking stimulus that was shown for 2966ms. In each trial only one of the pictures contained the object and the other was a random dot pattern. We performed the same experiment with 33ms exposures.

In the *detection and categorization* version, 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*). In half of the trials objects were cars and half were faces. Objects occurred with equal probability in both intervals.

In the *detection and identification* version 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*). All trials contained a male face that appeared with equal probability in both intervals; half of the stimuli were different pictures of the target individual.

Each experiment contained 128 trials. The order of the two responses within a trial, the order of these experiments and exposure duration was counterbalanced across subjects.

Results

Figure 5 shows that categorization performance was significantly better on detection hits than misses (17ms:p<0.001, 33ms:p<0.003, t-test, n=12). Importantly, categorization performance on detection misses was not different from chance level (17ms exposures, p>0.01, t-test). Crucially, the converse was also true: detection performance was significantly better on categorization hits than misses (17ms: p<0.001, 33ms: p<0.001, t-test) and detection performance was at chance for categorization misses. A 2-way ANOVA of performance as a function of task (detection/categorization)

and success (hit/miss in the second task) showed a main effect of success (17ms:F>12, p<0.003; 33ms: F>17, p<0.001), but there was no significant difference between tasks, or interaction between task and success at the other task. Thus, success on each task predicted success on the other task.

In contrast, comparison between face detection and identification within the same trial revealed completely different results (Fig.5c,d). First, detection performance was significantly higher than identification (17ms: p<0.01 33ms: p<0.001, t-test). Second, identification performance depended on detection performance, but detection did not depend on identification. Thus, identification performance on detection hits was better than on detection misses (33ms: p<0.01, t-test), but detection performance was not different for identification hit or miss trials (both ps>0.1, t-test). A 2-way ANOVA of performance at one task as a function or hit or miss at the other task revealed an interaction at exposures of 33ms (F>5.6, p<0.03).

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

GENERAL DISCUSSION

The same two phenomena occur with striking consistency: i) subjects do not require more processing time for object categorization than for object detection, whereas ii) comparable performance on the identification task requires substantially more processing time than required for either detection or categorization.

Our data provide evidence against Hypothesis 1, according to which 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 "super-ordinate level" (Rosch et al., 1976). Second, because Hypothesis 1 holds that object detection is prior to categorization, it predicts that on some trials objects will be correctly detected but not categorized, whereas the opposite will 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 independent 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, 1995; Rubin, 1958) and suggest instead that that categorization and segmentation are closely linked. This conclusion is consistent with the findings of Peterson et al. (Peterson & Gibson, 1993, 1994; Peterson & Kim, 2001; Peterson 2003; Peterson & Lampignano, 2003), although our conclusions differ slightly from theirs: whereas Peterson and colleagues conclude 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 template-like 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

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

information about both the object category and about 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, versus an account in terms of the sequence of processing in object recognition. Preliminary evidence from MEG and ERPs favors the idea that object segmentation and categorization occur at the same time (Halgren, Mendola, Chong, & Dale, 2003; Liu, Harris, & Kanwisher, 2002).

In contrast to the similarity of performance in the detection and categorization tasks, comparable performance in the identification task always required longer exposures and more processing time than categorization. On average 65 more milliseconds were necessary for identification compared to 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). Thus, argues against other claims that expertise leads to a change in the initial level of perceptual categorization of expert stimuli such as faces (Rosch et al., 1976; Tanaka, 2001).

From these behavioral data we cannot determine whether the extra time needed for identification compared to categorization reflects the engagement of a different mechanism, or simply a longer engagement of the same mechanism involved in categorization. Some evidence for the latter view comes from neural measures. First, fMRI studies in humans have shown that the same cortical regions are engaged in both the detection and identification of stimuli of a given category (Grill-Spector, 2003). Second, electrophysiology in monkeys has shown that stimulus selectivity of neurons in higher order visual areas changes over time (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 later neural responses are necessary for identification.

From a computational point of view, capturing the category rapidly may expedite identification by restricting the matching of the input to the internal representation only to the relevant category (instead of searching 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 an explanation of the present data in which (i) object detection and object categorization performance is based on the same perceptual analyzers, consistent with evidence from fMRI (Grill-Spector, 2003) and (ii) categorization of different basic level categories engage largely independent and non-overlapping perceptual analyzers (in contrast to recent claims by Haxby et al., (2001)), but (ii) identification of different stimuli within a category engage overlapping perceptual analyzers.

In sum, we show that although substantially more processing is required in order 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 model of object recognition

ACKNOWLEDGEMENTS

We would like to thank Bart Anderson, Galia Avidan, Jon Driver, Uri Hasson, David Heeger, Elinor McKone, Peter Neri and Mary Potter for fruitful discussions and comments on the manuscript. We would like to thank Mary Peterson and Simon Thorpe for their important comments on the manuscript. We would like to thank AJ Margolis and Jenna Boller for their help in running some of the experiments. This research was funded by Human Science Frontiers fellowship LT0670 to KGS and grant EY13455 to NK.

Correspondence should be addressed to Kalanit Grill-Spector, kalanit@psych.stanford.edu.

REFERENCES

- Borenstein, E., & Ullman, S. (2002). *Class-specific top-down segmentation*. Paper presented at the European Conference on Computer Vision (ECCV) 2002.
- Bregman, A. L. (1981). Asking the "what for" question in auditory perception. In *Perceptual Organization* (pp. 99-118). Hillsdale, NJ: Lawerence Erlbaum Associates.
- Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: a comparison, review, and update. *Perceptual Psychophysics, 62*(8), 1572-1595.
- Driver, J., & Baylis, G. C. (1996). Edge-assignment and figure-ground segmentation in short-term visual matching. *Cognitive Psychology*, *31*(3), 248-306.

Graham, N. (1989). Visual Pattern Analyzers. New York: Oxford University Press.

- Green, D., & Swets, J. (1966). Signal detection theory and psychophysics. John Wiley & Sons Ltd.
- 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., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of objectselective activation correlate with recognition performance in humans. *Nature Neuroscience*, *3*(8), 837-843.
- Halgren, E., Mendola, J., Chong, C. D., & Dale, A. M. (2003). Cortical activation to illusory shapes as measured with magnetoencephalography. *Neuroimage*, 18(4), 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(5539), 2425-2430.
- Hochstein, S., & Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791-804.
- Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: making the connection. *Cognitive Psychology, 16*(2), 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 Science U S A*, 92(12), 5587-5591.
- Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nature Neuroscience*, *5*(9), 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, Osherson, D.N., (Ed.), *An invitation to cognitive science: visual cognition*. Cambridge: MIT.
- Peterson, M. A. (1999). What's in a stage name? Comment on Vecera and O'Reilly (1998). *Journal of Experimental Psychology, 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 figureground 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*(4), 808-822.
- Rosch, E. (1978). Principles of categorization. In B. E. Rosch & Lloyd (Ed.), *Cognition* and categorization. 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. W. Beardslee, M. (Ed.), *Readings in perception*. 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: 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: MIT Press.
- Sternberg, S. (2001). Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. *Acta Psychologica, 106*(1-2), 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*(6747), 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*(5), 384-399.
- Tanaka, J. (2001). The entry point of face recognition: evidence for face expertise. *Journal of Experimental Psychology General, 130*(3), 534-543.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520-522.
- Vecera, S. P., & Farah, M. J. (1997). Is visual image segmentation a bottom-up or an interactive process? *Perceptual Psychophysics*, 59(8), 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(2), 441-462.



Figure 1: Experimental design

(a) Images were presented briefly for (17, 33, 50, 68 or 167ms) 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. (b) Examples of the images used in this study.



Figure 2: Naming performance for three recognition tasks.

Data are averaged across 15 subjects (5 male and 10 female subjects). Y-axis denotes accuracy of subjects corrected for guessing on object pictures; Error bars indicate SEM (standard error of the mean).



Figure 3: Experiment 2: Recognition performance within and across super-ordinate categories. Left: Accuracy data (corrected for guessing). Right: Reaction time (RT) data on correct trials. Data are averaged across 15 subjects (9 male 6 female). Error bars indicate SEM, RTs are not meaningful for the shortest exposure duration for the identification task, as accuracy was at chance; this value is not plotted in these graphs. Detection task (object vs. texture) data are plotted only for the corresponding basic level category in each panel.(a) Vehicles: differences between detection and categorization performance were not statistically significant in either RT or accuracy (all ps>0.07, t-test). Accuracy in identification task was lower than both categorization and detection for 33, 50, 68 and 167ms duration (p<0.0001,t-test) and RT were longer for 50, 68 and 167ms (p<0.001,t-test).(b) Musical Instruments: differences between detection and categorization performance were not statistically significant in either RT or accuracy (ps>0.1, t-test). Detection RT were faster than categorization within the super-ordinate category task (i.e., guitars vs. musical instruments) for 17, 50 and 68ms stimulus durations (p<0.01, t-test). Accuracy in the identification task was lower than both categorization and detection (p<0.01,t-test) and RT were longer (p<0.01,t-test) for all durations. (c) Animals: differences between detection accuracy and categorization of dogs (vs. inanimate objects) were not statistically significant (ps>0.08, t-test). Accuracy in detection was statistically significantly higher than categorization within super-ordinate category (i.e., dogs vs. birds and fish) only for 50ms (p<0.02, t-test, other ps>0.07,t-test). RT differences between the detection and categorization within super-ordinate class were statistically significant only for 17 and 33ms (p<0.03,t-test). Accuracy in detection was statistically significantly higher than identification for 33, 50, 68, 167ms (p<0.01,t-test). Accuracy in categorization was statistically significantly higher than identification for 33, 50ms). RTs in identification task were significantly slower from categorization and detection for all durations (p<0.01,t-test).



Figure 4: Detection, categorization and identification performance on unmasked stimuli.

Data were measured for 24 subjects (10 male and 14 female subjects). Error bars indicate SEM. Detection task data (object vs. texture) are plotted for cars. (a) Accuracy (corrected for guessing). (b) RT on correct trials



Detection / Categorization

Figure 5: Experiment 4 Results

Data are averaged across 12 subjects (7 male and 5 female). Raw hit rate in one task as a function of success (hit) or failure (miss) at the second task. Chance level is 50%. **a-b detection and categorization:** *Black:* performance in the detection task as a function of hit and miss in the categorization task. *Gray:* performance in the categorization task as a function of hits=1.57, but on categorization misses d'=0.50; Categorization d' on ca tegorization hits=1.37, but on detection misses d'=0.24. **c-d detection and identification:** *Black:* performance in the detection task as a function task. *Gray:* performance in the identification task. *Gray:* performance in the identification task as a function of hit or miss in the identification task as a function of hit and miss in the identification task. *Gray:* performance in the identification task as a function of hit or miss in the identification task as a function of hit or miss in the identification task as a function of hit or miss in the identification task as a function of hit or miss in the identification task. *Gray:* performance in the identification task as a function of hit or miss in the detection task. At 33ms detection d' on identification hits=3.1 and on misses d'=2.9; Identification d' on detection hits=1.25, but on misses d'=0.36.