Journal of Experimental Psychology: Human Perception and Performance

Visual Object Detection, Categorization, and Identification Tasks Are Associated With Different Time Courses and Sensitivities

Stephan de la Rosa, Rabia N. Choudhery, and Astros Chatziastros

Online First Publication, November 1, 2010. doi: 10.1037/a0020553

CITATION

de la Rosa, S., Choudhery, R. N., & Chatziastros, A. (2010, November 1). Visual Object Detection, Categorization, and Identification Tasks Are Associated With Different Time Courses and Sensitivities. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. doi: 10.1037/a0020553

Journal of Experimental Psychology:
Human Perception and Performance

© 2010 American Psychological Association
0096-1523/10/$12.00 DOI: 10.1037/a0020553

Visual Object Detection, Categorization, and Identification Tasks Are Associated With Different Time Courses and Sensitivities

Stephan de la Rosa, Rabia N. Choudhery, and Astros Chatziastros
Max-Planck-Institute for Biological Cybernetics, Tübingen, Germany

Recent evidence suggests that the recognition of an object's presence and its explicit recognition are temporally closely related. Here we re-examined the time course (using a fine and a coarse temporal resolution) and the sensitivity of three possible component processes of visual object recognition. In particular, participants saw briefly presented (Experiment I to III) or noise masked (Experiment IV) static images of objects and non-object textures. Participants reported the presence of an object, its basic level category, and its subordinate category while we measured recognition performance by means of accuracy and reaction times. All three recognition tasks were clearly separable in terms of their time course and sensitivity. Finally, the use of a coarser temporal sampling of presentation times decreased performance differences between the detection and basic level categorization task suggesting that a fine temporal sampling for the dissociation of recognition performances is important. Overall the three probed recognition processes were associated with different time courses and sensitivities.

Keywords: visual recognition, time course, detection, categorization, identification

According to many influential theories, visual object recognition is not a unitary process but consists of several component processes that are carried out in some temporal order (e.g. Marr & Nishihara, 1978; Biederman, 1987; Nakayama, He, & Shimojo, 1995). Detection, basic-level categorization, and identification are considered to be candidate component processes of object recognition (see e.g. Grill-Spector and Kanwisher, 2005). Here we define detection as the observer's judgments about an object's presence. Furthermore we refer to categorization as the recognition of the object's basic-level category (e.g. dog) and to identification as the recognition of the object's subordinate category (e.g. German Shepherd; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976).

A recent debate concerns the temporal order in which these component processes of object recognition are executed. A common view is that objects or object features are detected in the background before they are recognized in more detail (e.g. Nakayama et al, 1995). Recent evidence, however, suggests that the detection of an object and its categorization are associated with very similar reaction times. These results lead to the suggestion that the visual processes underlying detection and categorization are equally fast (Grill-Spector & Kanwisher, 2005).

According to the first view and several "classic" theories of object recognition, the visual system first detects parts of an object (e.g. object contours). Subsequently these parts are then integrated into an object representation which serves as a basis for a more detailed analysis of the object (e.g. Marr, 1982; Biederman, 1987; Nakayama et al., 1995; see also Treisman & Gelade, 1980). This kind of organization in which the detection precedes categorization or identification allows object recognition to be more efficient because processes underlying object categorization or identification can focus on visual information that pertains to an object rather than processing the entire visual array. According to this view visual processes underlying detection temporally precede visual processes underlying categorization. Hence one would expect that detection is associated with shorter reaction times than categorization or identification.

More recent evidence challenges the view that detection temporally precedes categorization. Grill-Spector and Kanwisher (2005) measured participants' reaction times and accuracy to detect, categorize, and identify images of objects in their natural background for various presentation times. Interestingly, they found that both accuracy and reaction times did not differ significantly between object detection and object categorization at all tested presentation times. Yet, object identification was clearly associated with significantly lower accuracy and significantly higher reaction times for all presentation times. Additionally when Grill-Spector and Kanwisher (2005) asked participants to both detect and categorize an object on a single trial, a trial-by-trial analysis revealed that categorization errors were related to detection errors and vice versa. Grill-Spector and Kanwisher (2005) therefore concluded that object detection and object categorization have the same time course. In contrast identification has a time course that is shifted towards longer reaction times suggesting that its underlying visual processes are slower.

Using the same experimental design, Mack, Gauthier, Sadr, & Palmeri (2008) only partly replicated Grill-Spector and Kanwish-

Stephan de la Rosa, Rabia N. Choudhery, and Astros Chatziastros, Department of Human Perception, Cognition, and Action, Max-Planck-Institute for Biological Cybernetics, Tübingen, Germany.

This research was supported in by the EU Project "Joint Action Science and Technology" (IST-FP6-003747).

Correspondence concerning this article should be addressed to Stephan de la Rosa, Department of Human Perception, Cognition and Action, Max-Planck-Institute for Biological Cybernetics, Spemannstr. 38, 72076 Tübingen, Germany. E-mail: delarosa@kyb.tuebingen.mpg.de

er's (2005) results. They showed that object detection and categorization are tightly temporally coupled when images were presented upright. However, when the same objects images were presented upside-down, participants' detection and categorization performance was significantly different. In particular, the detection of inverted objects was significantly better than their categorization as indicated by larger d′ values for detection. Furthermore, participants exhibited faster response times in the detection task than in the categorization task when object images were inverted. Mack et al. also found the detection of degraded (i.e. phase scrambled) object images to be faster and more accurate than their categorization. These results suggest that under more challenging viewing conditions object detection and categorization performance can be dissociated. However, the finding that detection and categorization of upright object images are associated with the same time course remains unchallenged.

The finding that detection and categorization are tightly temporally linked for upright object images imposes important constraints onto existing theories of object recognition. Here we re-investigated the time course of detection and categorization of upright natural object images (Grill-Spector & Kanwisher, 2005). The close temporal linkage between detection and categorization might be owed to the rapid nature of visual categorization (Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001). Hence, if differences between detection and categorization performance exist, they should occur at very short presentation times. We therefore decided to use a finer temporal sampling at short presentation times than in previous studies to examine the time course of visual recognition.

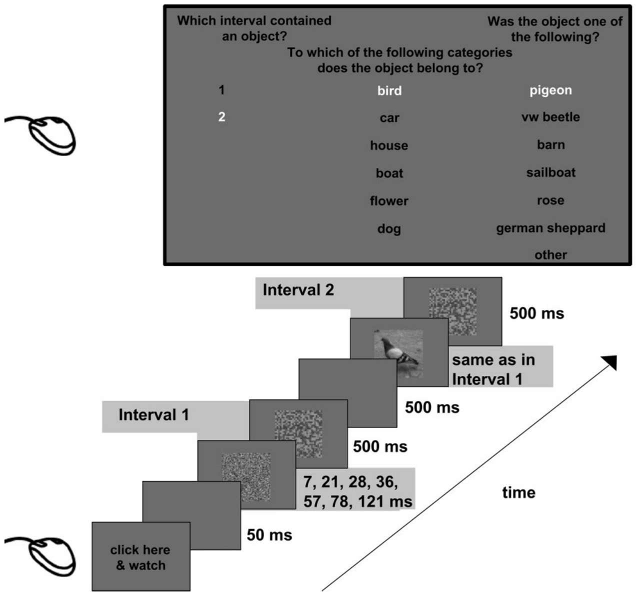

Figure 1. Schematic outline of an experimental trial in Experiment I. The labels to the right of a screen indicate the presentation time (in ms) of the corresponding screen. The mouse symbols indicate the screens for which participants mouse clicks were required to continue to the next screen, i.e. to start a trial and to respond (selected answers were highlighted in white). The answer screen is shown enlarged along with the correct answers for this trial for sake of clarity.

Experiment I: The Time Course of Object Recognition

We measured the time course of object detection, categorization, and identification using a finer temporal resolution for short presentation times than in previous studies (Grill-Spector & Kanwisher, 2005; Mack et al., 2008). Moreover, we were interested in assessing the degree of detail that participants could perceive from a single brief presentation of an image by probing detection, categorizing, and identification on the same trial. If detection is mediated by faster visual processes than categorization, then participants should be able to tell the presence of an object while being unable to categorize it at short presentation times. To this end, participants saw two temporally separated, backward-masked image presentations with one showing an object image (e.g. a dog) and the other a non-object image (visual noise; see Figure 1). Following these two image presentations participants had to indicate on an answer screen the object's presentation interval (detection task), the object's basic-level category (categorization task), and the object's subordinate category (identification task).

Methods

Participants. Ten naïve participants (age range between 18 and 30 years; four females) participated in the experiments. All participants had a normal or corrected-to-normal vision and were naïve to the task and stimuli. Each participant gave informed consent prior to the experiment and was compensated 8€/hour for their participation.

Apparatus and stimuli. Stimuli were presented on a Sony (Tokyo, Japan) Monitor (CPD-G500) by means of the Psychtoolbox (Brainard, 1997; Pelli, 1997). The gamma corrected monitor had a refresh rate of 140 Hz.

The grayscale object images were drawn from six object categories (bird, boat, car, dog, house, and flower). Each category contained 100 images. Within each of the six categories, 50 images were of a particular subordinate category, hereafter referred to exemplar. The six exemplars for the six categories were pigeon, sailboat, VW Beetle, German Shepherd, barn, and rose. The other 50 images in a category were non-exemplar images, that is, images of a different subordinate category than the exemplar category (e.g. all other birds except pigeon). The non-object images were patches of Gaussian visual noise. A new patch of visual noise was used on each trial. The mask was a scrambled version of an object image. To do so each object image was chopped up into 10 pixels by 10 pixels tiles that were then randomly rearranged. All stimuli were presented in the center of the screen with a gray level of 127 pixel (RGB value). All stimuli had the same size (5.89° visual angle), luminance (127 RGB pixel value), and contrast (20 RMS RGB pixel contrast).

Procedure. A trial began with the start screen "click here & watch" presented in the centre of the screen (see Figure 1). Participants had to left click with the mouse on "click here & watch" to start the trial. Following the mouse click, the start screen was replaced with a gray screen for 50 ms to minimize forward masking. The gray screen was followed by two image presentation intervals which were separated by 500 ms inter-stimulus-interval consisting of a gray screen. One interval presented a real-object image (an object from one of the categories) and the other interval presented a non-object image (visual noise). In both intervals the presentation of an image was immediately followed by a mask (scrambled object image) that was visible for 500 ms. Following the image presentation the answer screen was presented. It always presented the same three questions along with the same answer options (see Figure 1). The three questions were designed to measure detection, categorization, and identification, respectively. Participants answered all three questions by selecting one answer option for each question with a mouse click. Participants were instructed that a) they may answer the questions in any order; b) once an answer was selected, it could not be changed; c) all three questions are of equal importance; and d) they should guess an answer if they did not know the answer to a question. Participants had to answer all three questions to move on to the next trail. Once the three questions were answered, the next trial started by presenting “click here & watch” in the centre of the screen.

Forty-two trials constituted a block, and seven blocks an experiment (total of 294 trials). The real-object was pseudo randomly assigned to one of the two intervals on each trial with the restriction that the real-object had to appear in the first and second presentation interval equally likely (50%) within a block. The probability of guessing the correct answer of the detection question was therefore $p = .5$. Seven images from each of the six categories were shown in a block. Hence the probability of guessing the correct answer of the categorization question was $p = 1/6$ (six categories). Out of the seven images that were presented of a given category, six were exemplar images and one was a non-exemplar image. Hence, in total six non-exemplar images were shown within a block across all six categories. The probability of correctly guessing the correct answer (six target exemplars names plus the option “other” [see top right side of Figure 1]) of the identification question was therefore $p = 1/7$. Each object image was presented only once. Both, the object and the non-object images, were presented for the same duration. The presentation time was randomly selected on a given trial from the following presentation times: 7, 21, 28, 36, 57, 78, or 121 ms. Each presentation time was used six times (i.e. the frequency with which each presentation time occurred was counterbalanced) within a block and 42 times during an experiment. That is, the presentation order of presentation times was randomized while the presentation frequency was counterbalanced across presentation times.

Results and Discussion

Recognition performance was measured in terms of corrected-for-guessing accuracy scores using the following formula (Macmillan & Creelman, 2005, p. 252):

$$c = [m \times p(c) - 1]/(m - 1) \times 100, \tag{1}$$

where c is the accuracy corrected for guessing in percent, p(c) is the probability of a correct response, m is the number of answer alternatives in a given task; $m = 2$ in the detection task, $m = 6$ in the categorization task, and $m = 7$ in the identification task.

Figure 2 left panel shows the psychometric functions relating accuracy and presentation time for each of the three recognition tasks separately. The psychometric functions for detection, categorization, and identification have clearly different shapes. A repeated-measures analysis of variance (ANOVA) with presentation times and recognition tasks as within-subject factors was conducted to investigate whether the observed differences in Figure 2 left panel bear statistical significance. Both main effects of presentation time, $F(6, 54) = 146.17$, $p < .001$, $\eta^2_{\text{partial}} = 0.942$, and recognition task, $F(2, 18) = 27.30$, $p < .001$, $\eta^2_{\text{partial}} = 0.752$, were significant. The interaction of presentation time and recognition task was also significant suggesting that the presentation time had a different effect on accuracy scores for the three recognition tasks, $F(12, 108) = 23.70$, $p < .001$, $\eta^2_{\text{partial}} = 0.725$. To see which recognition tasks differed from each other, we conducted two separate (detection vs. categorization and categorization vs.

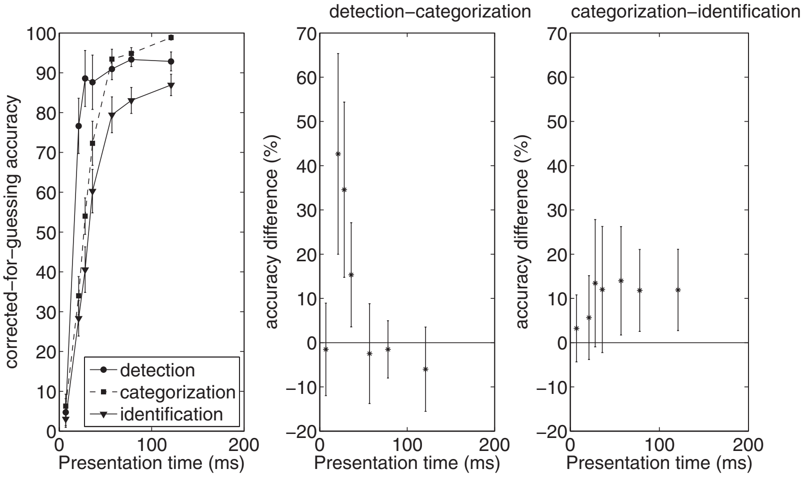

Figure 2. Results of Experiment I. Left: Psychometric functions relating mean accuracy (corrected-for-guessing) to presentation time (in ms) for each of the three recognition tasks in Experiment I. Bars indicate one standard error of the mean. Middle: Mean accuracy difference between detection and categorization performance calculated for each presentation time separately. Right: mean accuracy difference between categorization and identification calculated for each presentation time separately. Bars indicate 95% Bonferroni-corrected confidence intervals in the middle and the right panel.

identification) within-subjects ANOVAs with presentation time and recognition task as within-subject factors.

Detection vs. Categorization

Figure 2 (left panel) suggest that the psychometric function for detection and categorization clearly differ. A within subject ANOVA with presentation time and recognition task as within subject factors was used to investigate the statistical significance of this observation. Both main effects of presentation time, $F(6, 54) = 120.68$, $p < .001$, $\eta^2_{\text{partial}} = 0.931$, and recognition task, $F(1, 9) = 18.67$, $p = .002$, $\eta^2_{\text{partial}} = 0.675$, were significant. The significant interaction of presentation time and recognition task indicates that detection was better than categorization only on some presentation time levels, $F(6, 54) = 34.75$, $p < .001$, $\eta^2_{\text{partial}} = 0.794$. Bonferroni-corrected paired *t*-tests conducted for each presentation time separately revealed that detection performance was better than categorization performance for presentation times at 21, 28, and 36 ms (see Figure 2 middle panel).

Categorization vs. Identification

Figure 2 (left panel) suggests performance differences between the categorization and identification task. A two-way, within-subjects ANOVA with presentation time and recognition task as factors was used to examine these observed differences. We found a significant main effect of presentation time, $F(6, 54) = 155.92$, $p < .001$, $\eta^2_{\text{partial}} = 0.945$, and recognition task, $F(1, 9) = 17.37$, $p = .002$, $\eta^2_{\text{partial}} = 0.659$. The interaction of presentation time and recognition task was also significant, $F(6, 54) = 3.35$, $p = .007$, $\eta^2_{\text{partial}} = 0.271$. Bonferroni-adjusted *t*-tests were used to investigate at which presentation time levels categorization differed significantly from identification performance. Figure 2 right panel shows that categorization was better than identification at presentation times longer than 36 ms.

These results suggest that detection, categorization, and identification are associated with different time courses. The observed differences between the three recognition tasks in Experiment I, however, might be owed to mechanisms other than visual processing speed. In particular the lower performance in the categorization task might be due to fading of categorization response representations in short term memory since participants answered the three questions in the order of detection, categorization, and identification on 99.65% of the trials. Other, at least theoretically possible, confounding factors that might also have biased the results of Experiment I are backward priming (category response facilitation due to the knowledge of the object's identity), and experimental design differences (there were two task relevant presentation intervals for the detection task but only one for categorization and identification task).

We tried to minimize memory, priming, and experimental design effects on recognition performance in Experiment II.

Experiment II: Single Response Control Experiment

Detection, categorization, and identification were measured in separate experiments using a one-interval-forced-choice paradigm (1IFC; Figure 3). In brief, every 2 s, participants saw one backward masked image and had to indicate whether the shown image is of a predefined category. In Experiment IIa ten naïve participants completed one detection and one categorization experiment to compare detection and categorization performance; in Experiment IIb another ten naïve participants completed both one categorization and one identification experiment to compare categorization and identification performance. This experimental design minimized memory and backward priming effects between different recognition tasks by probing only one recognition task at a time. Furthermore we eliminated design differences due to the employment of the same 1AFC task for all three recognition tasks.

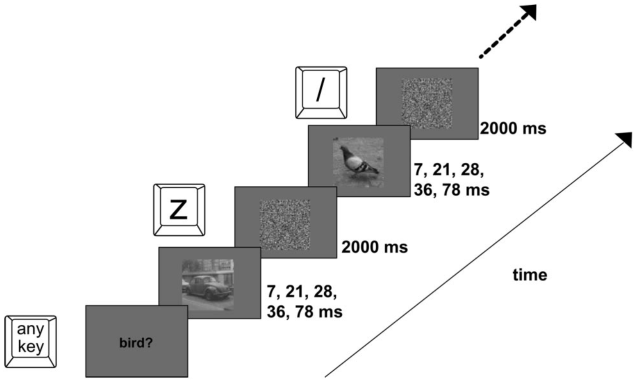

Figure 3. Experimental procedure used in Experiment II and III. The presentation times for each screen are given in ms next to corresponding screen. The keyboard key symbol indicates the screen that required participants' keyboard input, i.e. to start a run (see Methods section for the definition of a run) and to give an answer. The target was defined only at the beginning of a run. For sake of clarity the figure only shows the presentation times for a categorization task in Experiment IIa along with the correct key presses for a right handed participant.

Methods

Participants. Twenty naïve participants (age range between 20 and 30 years; 12 females) participated in Experiment II (ten in Experiment IIa and ten in Experiment IIb). All participants had a normal or corrected-to-normal vision and were naïve to the task and stimuli. Each participant gave informed consent prior to the experiment and was compensated 8€/hour for their participation.

Apparatus and stimuli. The apparatus and stimuli used in Experiment IIa and IIb are identical to Experiment I.

Procedure. The same experimental procedure as outlined in Figure 3 was used for Experiment IIa and IIb. In short participants saw one image every 2 s and indicated for each image presentation whether the shown image matched a predefined target. 50% of the trials were target and 50% of the trials were non-target trials (random assignment). Image presentation times in Experiment IIa and IIb were chosen to highlight differences between the recognition tasks as suggested by Experiment I (see also Figure 2). Consequently the image presentation times were 7, 14, 21, 28, and 78 ms in Experiment IIa and 21, 28, 36, 57, and 121 ms in Experiment IIb. The noise presentation period (a new noise patch was used on every trial) served also as response period. Participants' task was to answer as quickly and as accurately as possible whether the shown image matched the predefined target by pressing the target key ("z" or "/") with their dominant hand and the non-target key with the non-dominant hand. 100 image presentations constituted a run. An experiment, e.g. detection experiment, consisted of three of these runs (Experiment IIa consisted of one detection and one categorization experiment; Experiment IIb consisted of one categorization and one identification experiment). A different target (e.g. bird in the categorization task or pigeon in the identification task) was used for each run except for the detection task, where the target was always an object image and the non-target was always a scrambled object image (see Experiment I for a description of the scrambling method). Participants were verbally informed about the target by the experimenter prior to the run and visually reminded of the target at the beginning of the run by the presentation of the target word in the middle of the screen ("object?" in the detection task, e.g. "bird?" in the categorization task, e.g. "pigeon?" in the identification task). Non-targets were images of other categories and non-exemplar images in the categorization and identification task, respectively. Non-targets in the detection task were scrambled object images. Recognition task testing order was counterbalanced across participants within Experiment IIa and IIb. Note that recognition task was a within-subject-factor and mean RT were calculated from hit trials only.

Results and Discussion

We compared detection, categorization, and identification performance by means of d′ and reaction times (RT). D′ measures are preferable to correction-for-guessing accuracy scores as they more adequately assess the effect of partial knowledge on task performance (Macmillan & Creelman, 2005). As for the d′ calculation we counted a correct target recognition as a hit and the recognition of a non-target as a target as a false alarm.

Experiment IIA: Detection vs. Categorization

D′ values as a function of presentation time are shown for the detection and categorization task in Figure 4A top-left. Detection d′ values seem to be consistently higher than categorization d′ values. A repeated-measures ANOVA with presentation time and recognition task as within-subject factors and d′ as dependent variable was used to investigate significant differences in detection and categorization performance. We found significant main effects of presentation time and recognition task, $F(4, 36) = 73.14$; $p < .001$, $\eta^2_{\text{partial}} = 0.890$, and $F(1, 9) = 44.93$; $p < .001$, $\eta^2_{\text{partial}} = 0.831$ respectively. The interaction of presentation time and recognition task was also significant, $F(4, 36) = 13.31$, $p < .001$, $\eta^2_{\text{partial}} = 0.597$ suggesting that performance differences between the two recognition tasks depended on presentation time. Figure 4A bottom-left panel shows that detection performance was better than categorization performance at all presentation time levels except for 78 ms.

The mean reaction times of the detection and categorization task are shown in the top right panel of Figure 4A. Reaction times are longer for the categorization task than for the detection task at all presentation time levels. We compared detection and categorization reaction times of the hit trials in a repeated-measures ANOVA with presentation time and recognition task as within-subject factors and mean RT as dependent variable. We found a significant main effect of presentation time, $F(4, 36) = 59.54$, $p < .001$,

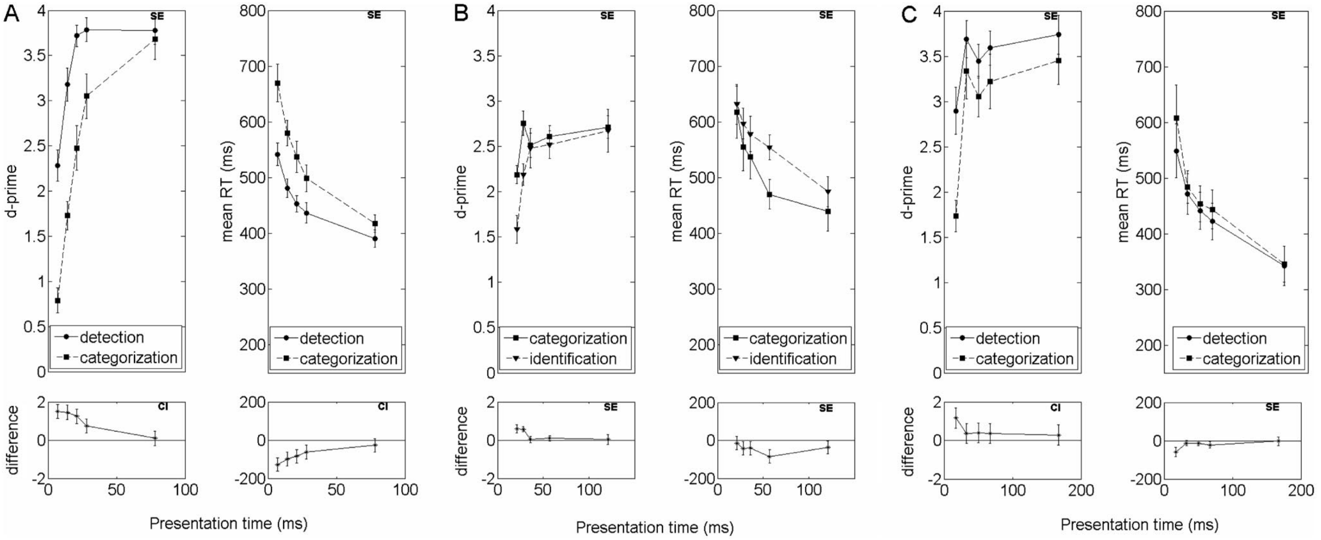

Figure 4. Results of Experiment II and III. Panel A, B, and C show the results of Experiment IIa (detection-categorization) and Experiment IIb (categorization-identification), and Experiment III (detection-categorization), respectively. All three panels are organized in the same way. The top-left panel shows d′ as a function of presentation time for each recognition task separately. The bottom left panel shows mean d′ differences between the two recognition tasks for each presentation time separately. The top-right panel plots mean RT as a function of presentation time for each recognition task separately. The bottom right panel shows the mean RT differences between the two recognition tasks for each presentation time separately. The letters *SE* or CI in top-right corner of each graph indicate whether the bars indicate one standard error from the mean (SE) or the 95% Bonferroni corrected confidence interval (CI).

$\eta^2_{\text{partial}} = 0.869$, and recognition task, $F(1, 9) = 29.08$, $p < .001$, $\eta^2_{\text{partial}} = 0.763$. The significant interaction between presentation time and recognition task implies that RT differences between the two tasks depend on presentation time, $F(4, 36) = 6.02$, $p = .001$, $\eta^2_{\text{partial}} = 0.401$. Figure 4A bottom-right panel shows that detection is significantly faster than categorization at all presentation time levels except for 78 ms.

Experiment IIB: Categorization vs. Identification

D′ values for the categorization and identification task are shown in the top left panel of Figure 4B. Categorization d′ seem to slightly but consistently higher than identification d′. A repeated measures ANOVA with recognition task and presentation time as within-subject factors and d′ as dependent variable was used to investigate d′ difference between the categorization and identification task. We found a significant main effect of recognition task, $F(1, 9) = 8.77$; $p = .016$, $\eta^2_{\text{partial}} = 0.493$, and presentation time, $F(4, 36) = 8.11$; $p < .001$, $\eta^2_{\text{partial}} = 0.473$. The interaction between recognition task and presentation time was not significant, $F(4, 36) = 2.00$, $p = .113$, $\eta^2_{\text{partial}} = 0.183$. Categorization performance was better than identification performance (Figure 4B bottom-left).

The top right panel of Figure 4B shows that identification reaction times are always longer than categorization reaction times for all presentation times. We conducted a repeated-measures ANOVA with recognition task and presentation time as within-subject factors and mean RT of the hit trials as dependent variable to compare the reaction times of the categorization and identification task. We found significant main effects of recognition task, $F(1, 9) = 15.21$; $p = .004$, $\eta^2_{\text{partial}} = 0.628$, and presentation time, $F(4, 36) = 44.05$; $p < .001$, $\eta^2_{\text{partial}} = 0.830$. The interaction of recognition task and presentation time was not significant, $F(4, 36) = 2.32$, $p = .076$, $\eta^2_{\text{partial}} = 0.205$. Figure 4B bottom-right panel shows that categorization responses were generally faster than identification responses.

Experiment II was designed to minimize the effect of memory, priming, and differences in the experimental design on recognition performance. We found detection, categorization, and identification performance to be significantly different as we did in Experiment I. It therefore seems unlikely that the results of Experiment I were solely caused by memory, priming, or design differences. Our reaction time analysis further shows that detection is the fastest and identification the slowest of the three recognition processes. Overall, the results of Experiment II further support the idea that detection, categorization, and identification are associated with different time courses.

Why do we, but not others, observe significant differences between detection and categorization performance? Note, that the present experimental paradigm is identical to Mack et al. (2008) and very similar to Grill-Spector and Kanwisher's (2008) paradigm. The only difference between the present experiment and Grill-Spector and Kanwisher's (2005) Experiment 2 is the type of mask that was employed for forward and backward masking. One might argue that Grill-Spector and Kanwisher (2005) scrambled object image mask might provide a stronger masking effect than the noise mask used in the present experiment. The difference in the strength of the masking effect might explain why Grill-Spector and Kanwisher (2005) fail to find any differences between detection and categorization performance while we did. Although this is a possible explanation, other evidence suggests that this explanation is unlikely. Mack et al. (2008) also used a noise mask but did not find performance differences between detection and categorization for upright images. Hence a noise mask seems to be sufficiently strong to allow a replication of Grill-Spector and Kanwisher's (2005) results. We therefore suggest an alternative explanation to explain the discrepancy between the present and previous findings. We observe significant performance differences between detection and categorization only in a time range from 7 to 36 ms (see Figure 2 and Figure 4A). This time range has not been intensively probed in previous experiments. The discrepancy between present and previous reports could be therefore simply owed to the choice of presentation times. We investigated this possibility in Experiment III.

Experiment III: The Effect of Temporal Resolution on the Dissociation of Object Recognition Processes

Experiment III is identical to Experiment IIa with the only difference that Experiment III uses the previously reported presentation times, namely 17, 33, 50, 68, and 167 ms. We reasoned that if the choice of presentation times is critical for dissociating detection and categorization, we should not replicate our results with the previously employed presentation times.

Methods

Participants. 20 naïve participants participated in Experiment III (age range 20–27 years, 11 females). All participants had corrected-to-normal vision and gave their informed consent prior to the experiment. They received 8€/hour as compensation for their participation in the experiment.

All other methods were identical to Experiment IIa with the following exceptions. The monitor refresh rate was set to 60 Hz and the presentation times were 17, 33, 50, 68, and 167 ms.

Results and Discussion

We calculated the d′ values for each participant, presentation time, and recognition task separately. Figure 4C top-left panel shows the d′ values for the detection and categorization task for each presentation time separately. Although detection d′ had a tendency to be higher than categorization d′, the error bars suggest that detection and categorization d′ might not sufficiently far apart to reach significance. A repeated-measures ANOVA with presentation time and recognition task as within-subject factors was used to investigate differences between detection and categorization performance. We found a significant effect of presentation time, $F(4, 36) = 24.28$; $p < .001$, $\eta^2_{\text{partial}} = 0.730$, and recognition task, $F(1, 9) = 6.21$, $p = .034$, $\eta^2_{\text{partial}} = 0.408$. However the significant main effect of recognition task requires a more refined analysis since the interaction of presentation time and recognition task was significant, $F(4, 36) = 2.65$, $p = .049$, $\eta^2_{\text{partial}} = 0.227$, suggesting that recognition tasks differ only at some presentation time levels. We examined the significant interaction by calculating the mean d′ differences for each participant along with the corresponding 95% Bonferroni corrected confidence intervals (see Figure 4C, bottom-left panel). We found that detection and categorization only differed significantly at 17 ms.

Previous studies (Grill-Spector & Kanwhisher, 2005; Mack et al. 2008) used regular *t*-tests instead of Bonferroni-corrected *t*-test to compare detection and categorization performance. Since the latter ones have less statistical power, the result that detection and categorization do not differ over a large time range might be owed to the Bonferroni correction. We therefore compared detection and categorization performance for each presentation separately by means of non-corrected *t*-tests. We found detection and categorization performance was not significantly different at all presentation times except at 17 ms, $t(9) = 3.62$, $p = .006$, and 50 ms, $t(9) = 2.45$, $p = .037$. Interestingly Mack et al. (2008) obtained the exact same overall result in their comparison of detection and categorization performance of upright images. Based on this overall result they concluded that the time course of detection and categorization of upright images are the same. Using the same type of analysis as in previous studies, we replicate previous findings indicating that our results are not owed to the type of statistical analysis.

The reaction times of the categorization and identification task look very similar (Figure 4C top right panel). We analyzed the reaction times of the hit trials in a repeated-measures ANOVA with recognition task and presentation times as within-subject factors. We failed to find a significant main effect for recognition task, $F(1, 9) = 2.57$, $p = .143$, $\eta^2_{\text{partial}} = 0.222$, but we found a significant main effect of presentation time, $F(4, 36) = 47.23$, $p < .001$, $\eta^2_{\text{partial}} = 0.840$. The interaction between recognition task and presentation time was not significant, $F(4, 36) = 2.35$, $p = .072$, $\eta^2_{\text{partial}} = 0.207$.

Experiment III demonstrates that when using presentation times employed in previous studies, detection and categorization tasks are associated with similar performance over a broad range of presentation times. If we were to base our conclusions on these results only, we would arrive at very similar conclusion as Grill-Spector & Kanwhisher (2005) and Mack et al. (2008). That is, detection and categorization performance of upright images cannot easily dissociated and therefore seem to be associated with the same time course. Overall, we suggest that previous studies were unable to dissociate detection and categorization performance due to a coarse temporal resolution at short presentation times.

Experiment IV: The Sensitivity of Object Recognition

Visual performance can be dissociated in terms of its processing speed and its sensitivity. We were interested in whether we could find differences between the three recognition task performances also in terms of their sensitivity. We measured the accuracy of detecting, categorizing, and identifying simultaneously masked object images as a function of signal-to-noise (S/N) ratio. We used the same experimental procedure as in Experiment I with the only exception that images in both presentation intervals were presented for 500 ms and were simultaneously masked. No backward mask was applied in Experiment IV. After the two image presentations, participants made a detection, categorization, and identification judgment as in Experiment I.

Methods

The methods were similar to those of Experiment I. We will describe only the differences between the methods of Experiment I and IV.

Participants. Ten naïve participants participated in Experiment IV (age range: 20–31 years; 4 females). All participants had normal or corrected-to-normal vision and gave their informed consent prior to their participation. Participants received 8€/hour as compensation for their participation.

Stimuli and apparatus. The object image and the non-object image (visual noise patch) were both transparently overlapped by a patch of visual noise (simultaneously masked) and always presented for 500 ms. No backward mask was employed. The used image RMS contrasts were 0.19, 0.58, 1.62, 2.14, 2.67, 3.97, and 5.27 cd/m^2. Each image contrast was used 6 times within a block and 42 times during the experiment.

Results and Discussion

Our statistical analysis was done on signal-to-noise ratio (S/N) expressed in decibels (dB):

$$\mathrm{S/N(db)} = 20 \times \log(\mathrm{RMS}_{\mathrm{image}}/\mathrm{RMS}_{\mathrm{noise}}), \qquad (2)$$

The accuracy scores were corrected for guessing according to equation 1. The psychometric functions relating corrected-for-guessing accuracy to S/N ratio are shown for each recognition task separately in the left panel of Figure 5. Detection and categorization psychometric functions seem to differ more than categorization and identification psychometric functions. To see whether any of the observed differences bear statistical significance, we calculated a repeated-measures ANOVA with S/N ratio and recognition task as within-subject factor. Both the main effect of S/N ratio, $F(6, 54) = 341.39$, $p < .001$, $\eta^2_{\text{partial}} = 0.974$, and recognition task, $F(2, 18) = 113.48$, $p < .001$, $\eta^2_{\text{partial}} = 0.927$, was significant. We also found the interaction of S/N ratio and recognition task to be significant, $F(12, 108) = 13.83$, $p < .001$, $\eta^2_{\text{partial}} = 0.606$, suggesting that the stimulus contrast had a different effect on the three recognition tasks. We therefore compared detection vs. categorization performance and categorization vs. identification performance in two separate ANOVAs with recognition task and S/N ratio as within-subject factors.

Detection vs. Categorization

Figure 5 left panel suggest large performance differences between the detection and categorization task. A two-way within subject ANOVA with presentation times and recognition tasks as factors was used to investigate this difference. The main effect of recognition task, $F(1, 9) = 79.58$, $p < .001$, $\eta^2_{\text{partial}} = 0.898$, and S/N ratio, $F(6, 54) = 294.40$, $p < .001$, $\eta^2_{\text{partial}} = 0.970$, was significant. The interaction of S/N ratio and recognition was also significant, $F(6, 54) = 11.14$, $p < .001$, $\eta^2_{\text{partial}} = 0.553$, suggesting that stimulus contrast had a different effect on detection and categorization performance. A post-hoc comparison of detection and categorization performance at each level of S/N ratio with Bonferroni-corrected paired *t*-test (Figure 5 middle panel) shows that the detection and categorization performance differed significantly from S/N ratio at intermediate S/N ratios (between −16.15 and −11.84 dB inclusive).

Categorization vs. Identification

Categorization and identification performance seem to differ at higher signal-to-noise ratios as suggested by Figure 5 left panel. A

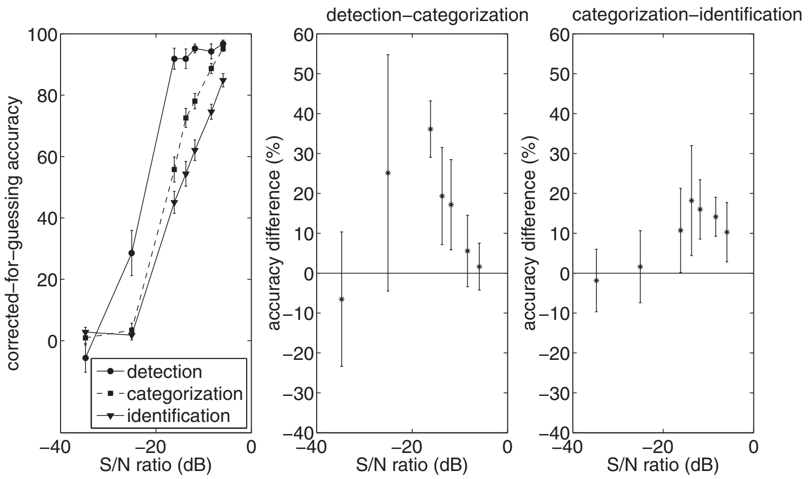

Figure 5. Results of Experiment IV. Left: Mean accuracy scores as a function of S/N ratio (dB) shown for each recognition task separately. Bars indicate one standard error of the mean. Middle: Mean accuracy differences between the detection and categorization task shown for each S/N ratio separately. Left: Mean accuracy difference between the categorization and identification task shown for each S/N ratio separately. Bars indicate the 95% Bonferroni-corrected confidence intervals in the middle and the right panel.

two-way within subjects ANOVA with presentation time and recognition task was used to investigate these differences. Both main effects of recognition task, $F(1, 9) = 62.28$, $p < .001$, $\eta^2_{\text{partial}} = 0.874$; and S/N ratio, $F(6, 54) = 300.83$, $p < .001$, $\eta^2_{\text{partial}} = 0.971$, were significant. The interaction of S/N ratio and recognition task was also significant, $F(6, 54) = 8.87$, $p < .001$, $\eta^2_{\text{partial}} = 0.496$. It seems that the performance differences between recognition tasks depend on the S/N ratio. A post hoc comparison of categorization and identification performance with Bonferroni-corrected paired *t*-test (Figure 5 right panel) shows that categorization and identification performance differed significantly at higher S/N ratios, namely between -13.73 and -5.92 dB (inclusive).

Taken together, we find the three recognition processes to differ in terms of their sensitivity with decreasing sensitivity in the order of detection, categorization, and identification. We therefore can dissociate the recognition of upright images across the three recognition tasks in terms of their sensitivity.

General Discussion

We examined the time course and the sensitivity of detection, categorization, and identification to examine whether they are dissociable in terms their recognition performance. We find that all three recognition processes are associated with different recognition performance in Experiment I, II, and IV suggesting that the three recognition tasks have different time courses and sensitivities. We conclude that the processing speed and the sensitivity of recognition processes decreases in the order of detection, categorization, and identification.

We suggest the following explanation for the discrepant findings between current and previous studies about detection and categorization performance differences (Grill-Spector & Kanwisher, 2005; Mack et al., 2008). Using presentation times of previous studies, we had difficulties to dissociate detection and categorization performance for upright images and therefore would come to very similar conclusions as previous studies (Experiment III). On the other hand, the use of a finer temporal sampling at short presentation times allowed us to clearly distinguish between detection and categorization performance (Experiment II). We suggest that the fine temporal sampling at short presentation times is major driving factor behind the discrepancies of the current and previous results. Further support for the idea that detection and categorization performance difference might occur at very short presentation times, come from studies demonstrating that object categorization is very rapid (Thorpe et al. 1996; VanRullen & Thorpe, 2001).

It is important to note that time courses of the three recognition tasks are not indicative of the absolute speed of the visual processes underlying detection, categorization, and identification. The reaction times are influenced by many factors that are not associated with the speed of the underlying recognition processes (e.g. spatial frequency, decision speed, planning and execution of the motor response). Hence the reaction times in the current study only provide information about the relative speed of the three recognition processes when the parameters (e.g. masking levels) of the present study are used.

We employed fewer object categories in our experiments than Grill-Spector and Kanwisher (2005) in theirs. Could some of the difference between the findings of their and our study be attributed to the different number of categories? We think that the lower number of categories in the present study is not the major driving force behind the observation that categorization has a different time course than detection for the following reason. Although Mack and colleagues (2008) employed nine stimulus categories while we employed only six, our results of Experiment III very closely replicate their results of detection and categorization performance being very similar when we use their presentation times. This demonstrates that we are able to obtain results that resemble previous reports of the close relatedness of detection and categorization performance with only six categories. However, when we use the same six categories and apply a finer temporal sampling at shorter presentation times as in Experiment II, detection and categorization performance clearly differ. We therefore think that the major driving force behind the observed differences between detection and categorization performance is the finer temporal sampling but not the fewer number of categories.

Finally to gain more insight into how visual processes underlying detection, categorization, and identification are related, we investigated the relatedness of errors in Experiments I and IV. In Experiment I and IV participants saw one stimulus (in a two alternative forced choice task) and made a detection, categorization, and identification judgment based on this single presentation while we recorded the accuracy for each recognition task on every trial. Because recognition judgments were based on the same stimulus presentation, the results of Experiment I and IV allow an examination of how closely an error in one recognition task is related to an error in another recognition task. The phi correlations between the correctness of detection and categorization indicate that an error in the categorization task was significantly related to an error in the detection task in both Experiments (see Table 1). The size of the correlation coefficient however suggests only a small to moderate relationship between detection and categorization errors. Furthermore, participants were more likely to conduct a categorization error when they did a detection error than to conduct a detection error when they did a categorization error in both experiments (see Table 1). These asymmetric dependencies between detection and categorization suggest that categorization errors are more dependent on detection errors than the other way around. As for categorization and identification errors, we found them to be significantly correlated in Experiment I and IV (see Table 1). The correlation of categorization and identification errors has a higher value than the correlation for detection and categorization errors. Moreover the conditional probabilities indicate that participants were more likely to conduct an identification error after a categorization error than the other way around. The difference between these two conditional probabilities was smaller than for detection and categorization. The high correlation between categorization and identification errors and the lesser asymmetric dependencies between these two recognition tasks suggest a stronger relationship between categorization and identification responses. Finally the correlations between detection and identification were moderate to small and about the same magnitude as the correlation between detection and categorization in Experiment I and IV. Moreover, the probability to conduct an identification error given a detection error was higher than the probability of conducting a detection error given an identification error in both experiments. Overall the error analysis shows a moderate to small relationship between errors of each pair of the three recognition tasks suggesting that errors are neither completely independent nor completely dependent. Furthermore the asymmetrical conditional error probabilities are indicative of errors being passed on from one recognition task to another one in mainly one direction.

Table 1
Correlation and Conditional Probabilities Between the Correctness of Responses of the Three Recognition Tasks in Experiment I (Time Course) and IV (Sensitivity)

Statistic	Experiment I	Experiment VI
φ(detection-categorization)	0.283**	0.324**
P(detection|categorization)	0.26	0.31
P(categorization|detection)	0.65	0.76
φ(categorization-identification)	0.665**	0.598**
P(categorization|identification)	0.67	0.68
P(identification|categorization)	0.89	0.85
φ(detection-identification)	0.239**	0.265**
P(detection|identification)	0.21	0.25
P(identification|detection)	0.71	0.78

** Indicates a significant phi correlation on the 0.001 level.

The results that the three recognition tasks are associated with different performances and are partly related are easily reconciled with of a feed-forward organization of visual recognition in which different recognition processes solve different recognition tasks and lower order processes provide partial input to higher order processes (e.g. Biederman, 1987). According to this view, detection, categorization, and identification would be mediated by recognition processes that are located at increasingly higher levels within the visual processing hierarchy. Dependencies between task performances arise from the output of preceding recognition processes providing (partial) input to subsequent recognition processes. The presence of task performance dependencies are also in line with a coarse-to-fine view of visual recognition (e.g. Schyns, 1998; Bar et al., 2006). According to this view, the first coarse low-spatial frequency object representation only provides sufficient information for object detection but not for object categorization or identification. Because the initial representation is subsequently filled with higher spatial frequencies components more detailed information about the object is available to the observer allowing object categorization and identification when longer presentation times are used.

Our results are however more difficult to explain in terms of object recognition models that assume a higher order scene (e.g. categorical) representation to be the initial conscious percept (e.g. Hochstein & Ahissar, 2002) as our results suggest that detection has the shortest processing times.

Overall our results support models of object recognition that postulate feed-forward and feedback mechanisms between visual processes. Our results provide further support for the idea that objects are not instantly recognized as soon we have the notion of "something being there" (e.g. Holm, Eriksson, and Andersson, 2008).

References

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, M., Dale, A. M., & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences, 103,* 446–454.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

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

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science, 16,* 152–160.

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

Holm, L., Eriksson, J., & Andersson, L. (2008). Looking as if you know: Systematic object inspection precedes object recognition. *Journal of*

Vision, 8, 1–7. Retrieved from http://journalofvision.org/8/4/14/, doi: 10.1167/8.4.14

Mack, M., 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,* 28–35.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory. A user's guide.* Mahwah, NJ: Lawrence Erlbaum.

Marr, D. (1982). *Vision.* San Francisco, CA: W. H. Freeman.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three dimensional structure. *Proceedings of the Royal Society of London B, 200,* 269–294.

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,* 1–70, Cambridge, M: MIT Press.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies, *Spatial Vision, 10,* 437–442.

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

Schyns, P. G. (1998). Diagnostic recognition: Task constraints, object information, and their interactions. *Cognition, 67,* 147–179.

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

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception, 30,* 655–668.

Received July 9, 2009
Revision received April 29, 2010
Accepted May 5, 2010 ■