Brief Reports

Object detection and basic-level categorization: Sometimes you know it is there before you know what it is

Michael L. Mack and Isabel Gauthier
Vanderbilt University, Nashville, Tennessee

Javid Sadr
Harvard University, Cambridge, Massachusetts

and

Thomas J. Palmeri
Vanderbilt University, Nashville, Tennessee

A tight temporal coupling between object detection (*is an object there?*) and object categorization (*what kind of object is it?*) has recently been reported (Grill-Spector & Kanwisher, 2005), suggesting that image segmentation into different objects and categorization of those objects at the basic level may be the very same mechanism. In the present work, we decoupled the time course of detection and categorization through two task manipulations. First, inverted objects were categorized significantly less accurately than upright objects across a range of image presentation durations, but no significant effect on object detection was observed. Second, systematically degrading stimuli affected categorization significantly more than object detection. The time course of object detection and object categorization can be selectively manipulated. They are not intrinsically linked. As soon as you know an object is there, you do not necessarily know what it is.

In a moment's glance, humans are able to recognize objects at various levels of abstraction. Whether detecting the presence of an object, determining what kind of object, or uniquely identifying an object, the human visual system is capable of rapid recognition. An important and long-standing question about object processing is when these different levels of abstraction become available to the perceiver. Given an image of a dog, how much time is required before we know that this image contains any object, or that this object is an animal, a dog, or a golden retriever? Some perceptual decisions are made more quickly than others. But does fastest mean first? Are certain perceptual decisions made prior to others?

The classic study by Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) found that participants were faster at verifying objects at the basic level compared to superordinate (e.g., animal) or subordinate (e.g., robin) levels. Later work showed that this basic-level advantage was malleable; the basic-level advantage can be eliminated in domains of perceptual expertise (Tanaka & Taylor, 1991) and atypical category members are quickly labeled at the subordinate level (Jolicœur, Gluck, & Kosslyn, 1984).

Jolicœur et al. favored the term "entry-level" to describe the level of abstraction verified the fastest. As the entry level at which object knowledge is first accessed, the term suggests that particular object categorizations may be made before others can be started.

The entry level could be more than an entry into object knowledge; it could also reflect an entry into perceptual awareness. By some accounts, objects are segmented from the visual scene before they are categorized (e.g., Nakayama, He, & Shimojo, 1995). By this account, you may know an object is there before you know what it is. But alternatively, image segmentation may be informed by object categorization (e.g., Peterson & Gibson, 1993). Or they may even be the very same mechanism (Grill-Spector & Kanwisher, 2005). By the latter account, as soon as you know an object is there, you know what it is.

A recent study by Grill-Spector and Kanwisher (2005) considered these alternative hypotheses by contrasting the time course of object detection with the time course of basic-level categorization and subordinate-level identification. In their experiments, participants performed verifications of briefly presented images. On object detection

---

T. J. Palmeri, thomas.j.palmeri@vanderbilt.edu

blocks, they verified the presence of any object versus a pattern mask; on basic-level categorization blocks, they verified a particular object category (e.g., dog) versus other categories (e.g., birds or fish); on identification blocks, they verified a particular subordinate-level category (e.g., jeep) versus other objects from the same basic-level category but different subordinate-level categories (e.g., other cars). Stimuli were presented for durations ranging from 17 msec to 167 msec and were followed by a mask to prevent further visual processing (see, e.g., Breitmeyer & Ogmen, 2000).[1] The empirical goal was to compare the time course of verification at detection, categorization, and identification.

First, performance on subordinate identification was significantly worse than object detection and basic-level categorization at all exposure durations. Second, and most intriguing, performance on object detection and basic-level categorization was identical at all exposure durations. Participants were just as accurate at categorizing an object at the basic level as they were at detecting the presence of any object. The conclusion was stated in the article's title: "As soon as you know it is there, you know what it is" (Grill-Spector & Kanwisher, 2005).

Although rapid categorization itself is consistent with a variety of theories of object processing that largely posit feedforward mechanisms (Nosofsky & Kruschke, 1992; Riesenhuber & Poggio, 1999; Thorpe & Fabre-Thorpe, 2001), Grill-Spector and Kanwisher's (2005) results could pose serious challenges to extant theories of object processing and categorization. They argue that because object detection and basic-level categorization have identical time courses, and because their time course is different from that of subordinate-level categorization, object detection and basic-level categorization may be the very same mechanism. In other words, basic-level categorization cannot occur without object detection and, critically, object detection cannot occur without basic-level categorization.

The notion that detection and basic-level categorization are linked in a stage of processing that precedes subordinate-level categorization runs counter to many extant computational models of perceptual categorization. According to such models, detection, basic-level categorization, and subordinate-level categorization are different perceptual decisions that can be made, not different stages of object processing. The speed of such perceptual decisions is based on factors such as the similarity of an object to stored category exemplars and how quickly exemplars are retrieved from memory (Nosofsky & Palmeri, 1997), and on the availability of perceptual dimensions over time (Lamberts, 2000). Similarly, many extant models of object recognition (e.g., Riesenhuber & Poggio, 1999; Tong, Joyce, & Cottrell, 2005) assume that categorization and identification are decisions that take place after perceptual processing and access to stored knowledge.

A tight temporal coupling between detection and basic-level categorization would have important implications for most extant models of categorization and object recognition. Our aim was to test whether the time courses of these decisions could be dissociated. Following the arguments presented by Grill-Spector and Kanwisher (2005), if object detection and basic-level categorization are the same mechanism, they should show the same time course irrespective of task manipulations. Such a finding would force a major reconceptualization of current models. But, if the time courses of detection and categorization can be dissociated, this would significantly weaken the evidence for an initial stage of detection and categorization in visual processing. Instead, detection, basic-level categorization, and subordinate-level categorization might best be characterized as different kinds of perceptual decisions that can be manipulated by a variety of task factors.

The task manipulations used in the following two experiments were stimulus inversion and stimulus degradation, respectively. Inversion has been shown to impair the recognition of objects and faces (Yin, 1969) and the recognition of animals in briefly presented scenes (Rousselet, Macé, & Fabre-Thorpe, 2003). We predicted that while inversion may disrupt information critical for basic-level categorization, object detection may not be affected because information sufficient for detection (e.g., lower level visual properties and local object parts) will be preserved. Stimulus degradation limits the amount of perceptual information available for visual processing. We predicted that basic-level categorization may require more visual information than object detection leading to significantly worse performance on visually degraded stimuli for categorization than detection.

In both of our experiments we exactly followed the experimental procedures outlined in Grill-Spector and Kanwisher (2005). The only difference was including an inversion (Experiment 1) or stimulus degradation (Experiment 2) manipulation.
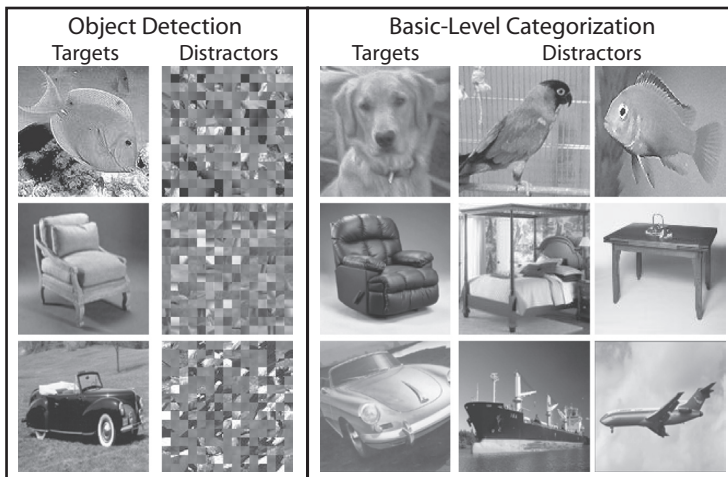
EXPERIMENT 1

Method

Participants. Thirty-six undergraduate students at Vanderbilt University participated for course credit.

Stimuli. Following Grill-Spector and Kanwisher (2005), some stimuli were images of objects from the animal and vehicle domains. Because our inversion manipulation required that the objects have some conventional orientation, our experiment replaced the musical instruments domain used by Grill-Spector and Kanwisher with images of furniture. The images of animals and vehicles were from the same database used by Grill-Spector and Kanwisher and the images of furniture were collected from various Web sources. Each of the three domains consisted of images belonging to three different basic-level categories (from animals, dog, bird, and fish; from furniture, chair, bed, and table; from vehicles, car, boat, and plane). Following Grill-Spector and Kanwisher, nonobject textures were created by randomly scrambling 8×8 pixel squares from natural images, and pattern masks were created by randomly scrambling 1×1 pixel squares from the images used as targets (see Figure 1). Stimuli subtended approximately $5.2º \times 5.2º$ of visual angle and were presented on a 19-in. computer monitor that sat approximately 60 cm from the participants.

Procedure. The experimental procedure was identical to that of Grill-Spector and Kanwisher (2005), with two exceptions. First, we included trials with inverted stimuli. Second, because our study focused on comparing the time course of object detection and basic-

Figure 1. Examples of the stimuli used during the experiment. The left panel shows examples of the stimuli used for object detection trials, with the left column containing examples of target stimuli (images of objects) and the right column containing examples of nonobjects. The right panel shows examples of stimuli used for basic-level categorization trials, with the left column containing examples of target stimuli (dogs, chairs, cars) and the two right columns showing examples of distractors (birds, fish, beds, tables, boats, planes).

level categorization, subordinate categorization was not tested; adding an inversion manipulation for subordinate categorization as well would have made the experimental session too long.

Each block began with an instruction defining the target category for a given block. This instruction was "Object?" in the detection blocks and "Dog?," "Car?," or "Chair?" in the categorization blocks. After these instructions, the block of trials began. On each trial, a stimulus image was presented for 17, 33, 50, 68, or 167 msec followed by a mask; the mask was presented until 2,000 msec had elapsed from the onset of the stimulus image. On categorization trials, target images contained an object from one of three basic-level categories (dog, chair, or car) and distractor images were from other basic-level categories within the same domain (bird and fish, bed and table, boat and plane, respectively). On detection trials, target images contained an object from any of the basic-level categories listed above and distractors consisted of nonobject textures. The poststimulus mask of each trial also served as a prestimulus mask for the next trial (see Figure 2A).

Participants responded by pressing a key labeled "yes" or "no" on each verification trial. A response could be made as soon as the stimulus was presented. If 2,000 msec elapsed before a response was made, the trial was flagged for removal and the next trial began. The experiment consisted of 12 blocks (6 detection and 6 categorization blocks) with each block having 60 trials, for a total of 720 trials in the experiment. The entire experiment lasted roughly 45 min. All participants completed 15 detection and categorization practice trials on stimuli not used during the experiment. The order of detection and categorization blocks was randomized for each participant, and exposure duration and stimulus order were randomized within a block.

Results

Performance was measured using d', with a "hit" defined as correctly verifying a target image from the instructed category and a "false alarm" defined as erroneously verifying a distractor image from that category. We also measured reaction times of hits. The data from nine participants were removed from analysis because these participants failed to respond on more than 15% of the trials.

The key results of Experiment 1 are shown in Figure 2B. The four graphs plot the participants' performance measured by d' and reaction times for upright (top two graphs) and inverted (bottom two graphs) stimuli as a function of task (categorization vs. detection) and as a function of stimulus exposure duration. Replicating Grill-Spector and Kanwisher (2005), the time course of object detection and basic-level categorization were the same for upright stimuli. However, for inverted stimuli, the time course of detection and categorization were dissociated, with poorer performance on categorization than detection.

The accuracy data were statistically analyzed by a 2 (task) × 2 (orientation) × 5 (exposure duration) × 3 (stimulus domain) within-subjects ANOVA. Greenhouse–Geisser correction for nonsphericity was applied to all reported ANOVA results, and all analyses were conducted with an α level of .05. Of particular interest was the significant interaction of orientation × task × exposure duration $[F(4,104) = 7.34, MS_e = 1.94]$. Planned comparisons revealed no significant difference between categorization and detection of upright stimuli at most exposure durations [except at 17 msec and 50 msec; $t(27) = 2.84$ and 3.12, respectively]; by contrast, there was a significant difference between categorization and detection of inverted stimuli at all exposure durations $[t(27) > 2.5]$, except at 17 msec. Furthermore, planned comparisons also revealed no significant difference for detecting upright versus inverted stimuli at all exposure durations $[t(27) = 3.16]$, except at 68 msec; but there was a significant difference for categorizing upright versus inverted stimuli at all exposure durations $[t(27) > 2.09]$, except at 17 msec and 33 msec.
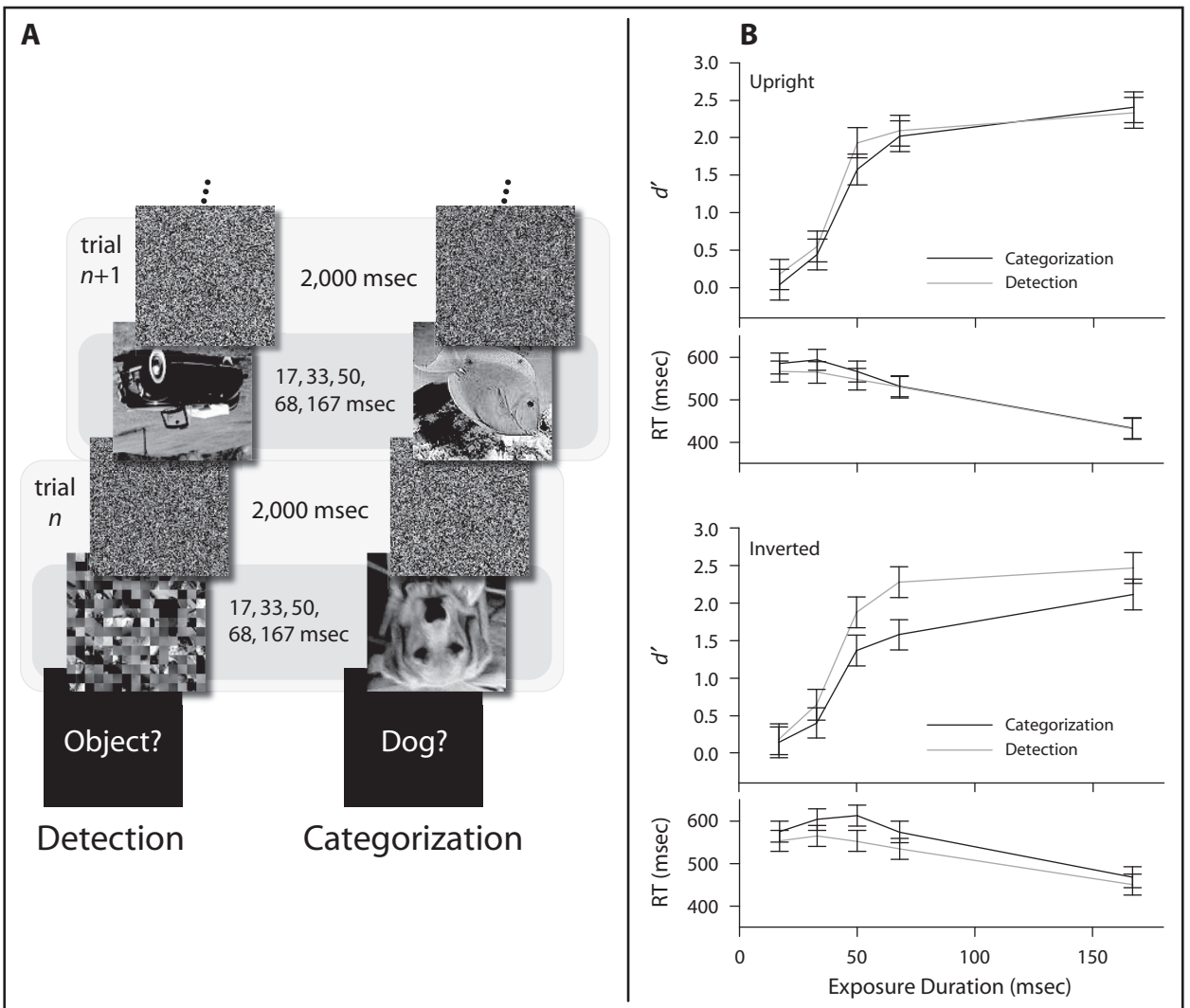
Figure 2. (A) Procedure for Experiment 1. Participants completed both detection and categorization blocks. A block began with presentation of a word describing what participants were to respond "yes" to during the block. A series of trials followed in which a stimulus image was presented for a variable duration of 17, 33, 50, 68, or 167 msec, followed by a mask that was shown until 2 sec had elapsed since the onset of the stimulus image. On half of the trials, the stimulus image was presented inverted. Participants could respond with the onset of the stimulus image and throughout the presentation of the post mask. (B) Accuracy (d') and reaction time (msec) from Experiment 1. The first and third graphs plot accuracy versus stimulus exposure duration, and the second and fourth graphs plot reaction time versus stimulus exposure duration for basic-level categorization (black line) and object detection (gray line). The top two graphs contain data for upright trials, and the bottom two graphs contain data for inverted trials. Error bars represent 95% confidence intervals.

Not surprisingly, a significant main effect of exposure duration was also observed [$F(4,104) = 368.96$, $MS_e = 437.32$], with increased performance with increased exposure duration. Significant main effects of orientation [$F(1,26) = 5.31$, $MS_e = 1.09$] and task [$F(1,26) = 22.32$, $MS_e = 24.22$] were also observed, but these can be explained largely through the interactions described above.

The same kind of ANOVA was performed on the reaction time data. A significant interaction of orientation by task was found [$F(1,26) = 18.7$, $MS_e = 2,547$], with longer reaction times for categorization than detection on inverted trials, but not on upright trials. A significant

orientation by exposure duration interaction was also observed [$F(4,104) = 4.25$, $MS_e = 6,427$], with shorter reaction times for upright trials across exposure durations. Finally, a significant task by exposure duration interaction was observed [$F(4,104) = 2.97$, $MS_e = 4,789$], with longer reaction times for categorization at most exposure durations. Significant main effects for orientation, task, and exposure duration where also observed, but these are largely explained by the interactions described above.

Some significant high-order interactions that included the factor of stimulus domain were also revealed by the ANOVAs for both accuracy and reaction times. However,

these interactions were driven by quantitative difference in performance across domains, not qualitative differences in the effects of other factors. Because we made no explicit attempt to equate the stimulus domains on any perceptual characteristics, it was not surprising to find modest quantitative differences across domains.

Discussion

The results of Experiment 1 replicate the finding of Grill-Spector and Kanwisher (2005) that for canonically oriented stimuli, the time course of object detection and basic-level categorization are the same. However when stimuli are inverted, the time course of basic-level categorization, but not object detection, is significantly impacted in both accuracy and reaction times. This dissociation of the time course of categorization from that of detection suggests that these two kinds of perceptual decisions need not be intrinsically linked.
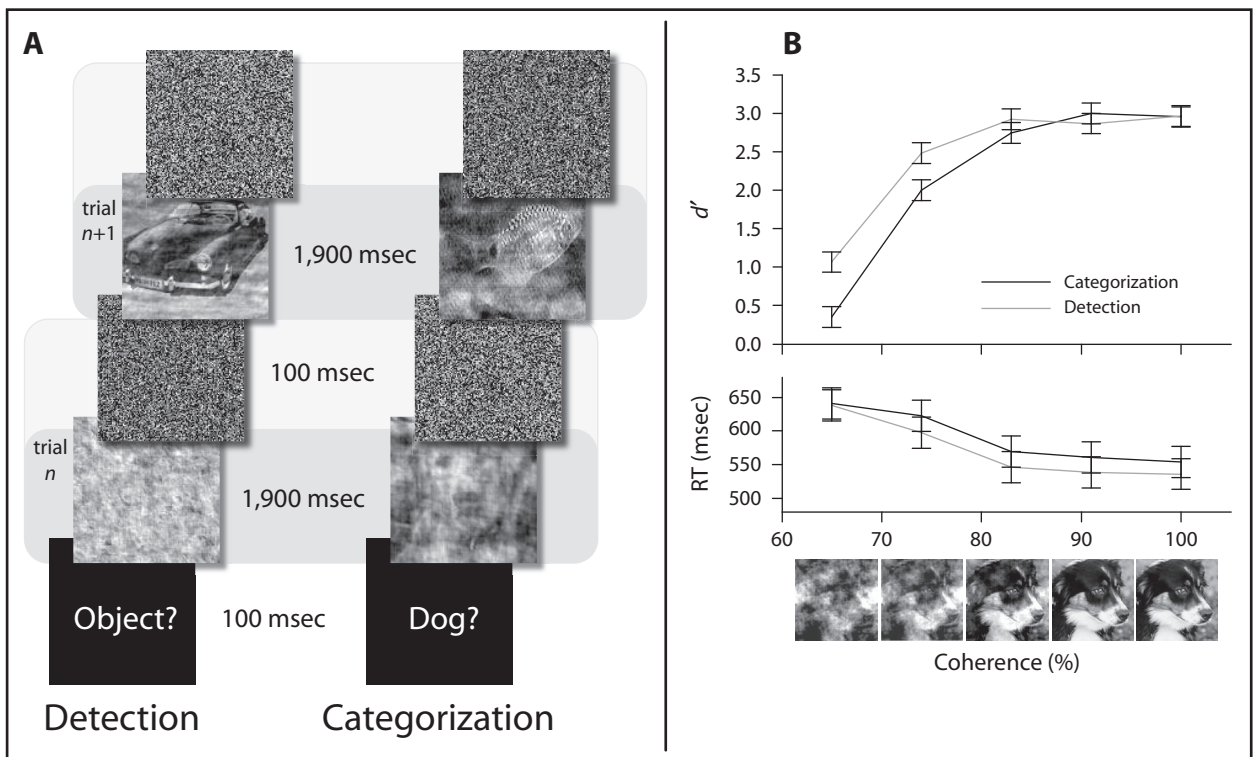
EXPERIMENT 2

As a complement to an inversion manipulation, we sought converging evidence for a temporal dissociation

of detection and categorization using a stimulus degradation manipulation. Experiment 1 varied both exposure duration and inversion. But for both practical and statistical reasons, Experiment 2 used a single exposure duration (100 msec) but varied stimulus degradation across a relatively wide range of values. To the extent that basic-level categorization, at this particular exposure duration, requires more visual information than object detection, we predicted that degrading stimuli will lead to worse performance for categorization than detection. As was the case in Experiment 1, this experiment exactly followed the experimental procedures outlined in Grill-Spector and Kanwisher (2005), except for the inclusion of stimulus degradation at just one stimulus exposure duration for upright stimuli.

Method

Participants. Twenty-three undergraduate students at Vanderbilt University participated in this experiment to satisfy credit requirements for introductory psychology courses.

Stimuli. The stimulus set was identical to the first experiment. Each image was degraded using the random image structure evolution (RISE) procedure (Sadr & Sinha, 2004). The RISE method degrades images by systematically introducing randomization to the

Figure 3. (A) Procedure for Experiment 2. Participants completed both detection and categorization blocks. A block began with a word describing the target category during the block. This was followed by a series of trials in which a stimulus image was presented for 100 msec, followed by a mask that was presented for 1,900 msec. The stimuli were degraded to a variable level. Participants could respond with the onset of the stimulus image and throughout the presentation of the post mask. (B) Accuracy (d') and reaction time (msec) from Experiment 2. Accuracy versus phase coherence percentage is plotted in the top graph, and reaction time versus phase coherence percentage is plotted in the bottom graph for basic-level categorization (black line) and object detection (gray line). An example of the RISE degradation is shown at the bottom along the horizontal axis. The source image, shown on the right at 100% phase coherence, can be systematically degraded to specific levels with RISE. The images below the graph show the range of phase coherence levels used in Experiment 2. From left to right, the images are examples of 65%, 73.75%, 82.5%, 91.25%, and 100% phase coherence. Error bars represent 95% confidence intervals.

phase component of the image while preserving the power spectrum. This results in an image that is degraded to a particular level of phase coherence, but also retains all the low level attributes from the power spectrum of the original image. In this experiment, the degradation of the images ranged from 65% to 100% phase coherence with 8.75% steps (see Figure 3B).

Procedure. This experiment was conducted in the same manner as Experiment 1, with the following exceptions: Stimulus exposure duration was fixed at 100 msec, an exposure duration that yielded identical performance for categorization and detection in Experiment 1. The integrity of the images was manipulated by presenting the stimuli with a phase coherence of 65%, 73.75%, 82.5%, 91.25%, or 100% using the RISE procedure (see Figure 3A for example trials). The experiment had six blocks (three categorization blocks and three detection blocks) with each block containing 120 trials (24 trials at each degradation level) for a total of 720 trials. All participants completed a 15 trial detection and categorization practice block on stimuli from categories that were not used during the experiment. Detection and categorization block order was randomized for each participant and degradation level and stimulus order within a block were randomized.

Results

The results of Experiment 2 are shown in Figure 3B. d' and reaction times are plotted for object detection (gray lines) and basic-level categorization (black lines) as a function of level of phase coherence. Data were analyzed with a 2 (task) $\times$ 5 (degradation) $\times$ 3 (domain) within-subjects ANOVA. Participants were more accurate at object detection than basic-level categorization at lower phase coherence levels, as evidenced by a significant interaction of task and phase coherence [$F(4,88) = 19.52$, $MS_e = 6.01$]. A significant main effect of task was also observed [$F(1,22) = 19.73$, $MS_e = 10.96$], with better performance for object detection than basic-level categorization. Not surprisingly, there was a main effect of stimulus degradation [$F(4,88) = 285.96$, $MS_e = 125.76$], with better performance with higher levels of phase coherence. There were also modest quantitative differences found in the domain and phase coherence interaction [$F(8,176) = 3.58$, $MS_e = 0.97$], although the key qualitative difference between detection and categorization was present across domains. For reaction times, the ANOVA revealed significant main effects of task [$F(1,22) = 5.9$, $MS_e = 9,719$] and phase coherence [$F(4,88) = 52.1$, $MS_e = 9,091$]. Reaction times were slower for categorization and were slower at increasing levels of stimulus degradation.

Discussion

Experiment 2 focused on one temporal snapshot from Experiment 1, where conventionally oriented (upright) stimuli produced identical performance for object detection and basic-level categorization. Within this snapshot, limiting the amount of perceptual information through stimulus degradation resulted in significantly worse performance (lower accuracy and slower reaction times) on basic-level categorization than object detection. Manipulating the phase coherence of stimuli affected both detection and categorization, but affected categorization significantly more. Manipulating the integrity of stimuli through degradation uncovers another example of a temporal dissociation between object detection and basic-level categorization.

GENERAL DISCUSSION

Grill-Spector and Kanwisher (2005) reported a tight temporal coupling between object detection and basic-level categorization. We replicated this finding for upright images. But we were able to dissociate the time course of object detection and categorization by image inversion (Experiment 1) and image degradation (Experiment 2).

Grill-Spector and Kanwisher's (2005) original findings were so provocative because they suggested that a single stage of processing that both segments an image into objects and categorizes those objects at a basic level—and does so at the very same time—might precede processing that identifies objects at more specific levels (and perhaps at more abstract levels as well). Whereas Peterson and colleagues (e.g., Peterson & Gibson, 1993, 1994) have suggested that categorization may influence image segmentation, this is the first suggestion we know of that categorization and segmentation may be the very same mechanism or that segmentation cannot occur without categorization, and vice versa. In addition, any suggestion that basic-level categorization is a perceptual stage of processing intimately tied with figure–ground segmentation that precedes subordinate-level identification poses a significant theoretical challenge. Many models of perceptual categorization (see, e.g., Nosofsky & Kruschke, 1992) and object recognition (e.g., Riesenhuber & Poggio, 1999) assume that categorization and identification represent different kinds of perceptual decisions, not different stages of perceptual processing. And most models assume that some process of image segmentation, if not an explicit decision about the presence of an object per se, precedes categorization and identification.

Subordinate-level identification may take more time than either basic-level categorization or object detection because of a slower accumulation of perceptual evidence favoring one perceptual decision over another (see, e.g., Nosofsky & Palmeri, 1997). For example, deciding that an image contains a dog can be made quickly because most of the perceptual information unequivocally favors the decision that a dog is present. By contrast, deciding that an image contains a Labrador retriever rather than a German shepherd takes more time because a lot of the perceptual information is consistent with either decision. So why might deciding that an image contains any object take the same amount of time as deciding that an image contains a dog? It is possible that for upright images, the most salient information for deciding that any object is present in the image may indeed be the very same coarse shape information that is used for a rapid basic-level categorization.

However, we observed that by inverting or by degrading the image, categorization was significantly slowed down relative to object detection. So although detection and categorization may use the same perceptual information when it is possible and advantageous to do so—for upright intact images—detection and categorization do not always show the same time course. The speed of per-

ceptual decisions can be influenced by a variety of task factors. We conjecture that while rapid detection decisions can be made using coarse shape information salient in upright intact images, rapid detection decisions could also be made on the basis of a whole host of local image features whose recognition would far be less impacted by either inversion or degradation. These ideas need to be formally evaluated using computational models.

It is clear that some perceptual decisions can be made more quickly than others. But of course, just because one kind of decision is made more quickly than another kind of decision, this does not mean that one decision has to be made before the other decision can be made (see, e.g., Mack, Wong, Gauthier, Tanaka, & Palmeri, 2007; Palmeri, Wong, & Gauthier, 2004). It is also just as clear that some kinds of perceptual information may be available before other kinds of perceptual information (e.g., Lamberts, 2000; Oliva & Schyns, 1997). But the availability of perceptual information does not mean that decisions about early perceptual information must be made before decisions about later perceptual information. Suggesting that coarse perceptual information is available relatively early, and that this information may be sufficient for knowing that an object is there or knowing what it is, is quite different from suggesting that early categorization decisions must be made before later categorizations at more subordinate or superordinate levels can be made.

One possible criticism of our experiments is that both stimulus inversion and degradation do not represent situations of "normal" object categorization. In other words, the significant differences in the time course of detection and categorization that we report may reflect differences between a normal recognition process for visually intact, upright objects and a specialized recognition process for degraded or inverted objects. Indeed, Grill-Spector and Kanwisher (2005) noted that there may exist "extreme" situations in which a dissociation between object detection and basic-level categorization would be expected. For example, a distant bird soaring against a blue sky may easily be mistaken for an airplane without any doubt of its being an object. The manipulations used in our experiments hardly represent extreme situations. Any model of object recognition will need to account for intact object recognition as well as the effects of continuously increasing image degradation. It is unlikely that a model would need to posit different systems for intact versus degraded objects in order to account for our results. Furthermore, our visual system successfully recognizes noncanonically oriented objects every day. Although different recognition mechanisms for upright versus inverted faces have been suggested, we know of no such suggestion for familiar everyday object categories.

To summarize, the evidence we report suggests that object detection and basic-level categorization are different types of perceptual decisions that can be made easier or harder, slower or faster depending on a variety of task factors. They do not necessarily represent a common mechanism of early perceptual processing. As soon as you know that an object is there, you do not necessarily know what it is.

REFERENCES

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

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.

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

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

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review*, **107**, 227-260.

Mack, M. L., Wong, A. C.-N., Gauthier, I., Tanaka, J. W., & Palmeri, T. J. (2007). Unraveling the time course of perceptual categorization: Does fastest mean first? In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 1253-1258). Mahwah, NJ: Erlbaum.

Nakayama, K., He, Z. J., & Shimojo, S. (Eds.) (1995). *Visual surface representation: A critical link between lower-level and higher-level vision*. Cambridge, MA: MIT Press.

Nosofsky, R. M., & Kruschke, J. K. (1992). Investigations of an exemplar-based connectionist model of category learning. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 28, pp. 207-250). San Diego: Academic Press.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, **104**, 266-300.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, **34**, 72-107.

Palmeri, T. J., Wong, A. C.-N., & Gauthier, I. (2004). Computational approaches to the development of perceptual expertise. *Trends in Cognitive Sciences*, **8**, 378-386.

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

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

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, **2**, 1019-1025.

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.

Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision*, **3**, 440-455.

Sadr, J., & Sinha, P. (2004). Object recognition and random image structure evolution. *Cognitive Science*, **28**, 259-287.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, **23**, 457-482.

Thorpe, S. J., & Fabre-Thorpe, M. (2001). Perspectives: Neuroscience—Seeking categories in the brain. *Science*, **291**, 260-263.

Tong, M. H., Joyce, C. A., & Cottrell, G. W. (2005). Are Greebles special? Or, why the fusiform fish area would be recruited for sword expertise (if we had one). In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 2196-2201). Mahwah, NJ: Erlbaum.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, **81**, 141-145.

NOTE

1. Of course, this assumes that masking directly truncates visual processing. If it did, this technique would provide a direct window into the temporal dynamics of object processing. But masking does not disrupt all processing following mask onset (see, e.g., Keysers, Xiao, Foldiak, & Perrett, 2001), so this technique influences the amount of time available for direct visual processing, limiting the amount of visual information that can be extracted from a stimulus. So, like Grill-Spector and Kanwisher (2005), we must also examine reactions times in order to fully understand the temporal dynamics of processing. There are other complementary techniques for examining temporal dynamics, such as varying the speed–accuracy instructions to participants and providing a response signal (e.g., Mack, Wong, Gauthier, Tanaka, & Palmeri, 2007). We elected to work within Grill-Spector and Kanwisher's masking paradigm in order to address their recent findings most directly.