Perception of Objects in Natural Scenes: Is It Really Attention Free?

Karla K. Evans and Anne Treisman
Princeton University

Studies have suggested attention-free semantic processing of natural scenes in which concurrent tasks leave category detection unimpaired (e.g., F. Li, R. VanRullen, C. Koch, & P. Perona, 2002). Could this ability reflect detection of disjunctive feature sets rather than high-level binding? Participants detected an animal target in a rapid serial visual presentation (RSVP) sequence and then reported its identity and location. They frequently failed to identify or to localize targets that they had correctly detected, suggesting that detection was based only on partial processing. Detection of targets was considerably worse in sequences that also contained humans, presumably because of shared features. When 2 targets were presented in RSVP, a prolonged attentional blink appeared that was almost eliminated when both targets were detected without being identified. The results suggest rapid feature analysis mediating detection, followed by attention-demanding binding for identification and localization.

Keywords: natural scene perception, attentional blink, RSVP, visual detection of semantic categories

What information can we get from a visual display before focusing attention on the objects it contains? The perception of rich, multifaceted, and meaningful everyday scenes feels effortless. We assess high-level properties such as the presence of danger or the esthetics of a painting almost instantaneously. Yet studies using much simpler stimuli suggest marked limits to the capacity of visual attention. How can these conflicting observations be reconciled?

Early models proposed two stages of visual processing: an initial parallel stage registering relatively low-level features, such as brightness gradients, motion, and orientation, and a high-level object description stage that operates serially, has limited capacity, and requires focused attention (Neisser, 1967). Demonstrations of change blindness and inattentional blindness support this view, showing that without visual attention, significant events or changes can easily escape our awareness (Mack & Rock, 1998; O'Regan, Rensink, & Clark, 1999; Rensink, O'Regan, & Clark, 1997; Simons & Levin, 1997).

As early as the 1970s, another line of research challenged these views, suggesting a larger capacity and the possibility of attention-free processing of natural scenes. Potter (1975, 1976) showed that high-order representations can be accessed very rapidly from natural scenes presented at rates of up to 10 per second (see also Biederman, 1972). More recently, Thorpe and his colleagues (Thorpe, Fize, & Marlot, 1996) found that both humans and nonhuman primates are capable of rapid and accurate categorization of natural images presented for exposures as brief as 20 ms.

Their participants were asked to detect animal targets in everyday indoor and outdoor scenes. They recorded event-related potentials associated with performance in this go/no-go visual categorization task and found differential cerebral activity reflecting the decision that there was a target present as early as 150 ms after stimulus onset (Fabre-Thorpe, Richard, & Thorpe, 1998; Thorpe et al., 1996; VanRullen & Thorpe, 2001b). The removal of chromatic information from the images had no effect on performance, nor was foveal vision required (Delorme, Richard, & Fabre-Thorpe, 2000; Thorpe, Gegenfurtner, Fabre-Thorpe, & Bülthoff, 2001). Finally, the speed and accuracy of visual categorization seemed to be independent of familiarity and of target category: Performance was as efficient for means of transportation as for the more natural and evolutionarily significant animal category (Thorpe et al., 2001; VanRullen & Thorpe, 2001a).

Fabre-Thorpe, Delorme, Marlot, and Thorpe (2001) proposed a specific mode of visual processing, termed "ultrarapid visual categorization," that relies on a parallel and automatic feedforward mechanism that cannot be sped up, even by extensive training (VanRullen & Thorpe, 2001b). Behavioral and electrophysiological results showed latencies in a go/no-go natural categorization task that were as fast when two images were presented simultaneously as when only a single image was presented (Rousselet, Fabre-Thorpe, & Thorpe, 2002), suggesting that high-level object representations can be formed in parallel without focused attention. Li, VanRullen, Koch, and Perona (2002) found no decrement when participants were asked to detect an animal or a vehicle in a briefly presented natural scene while simultaneously performing a foveal task that is known to be attentionally demanding. In the foveal task the participants had to discriminate the presence of a unique letter in a display of 5 letter *T*s and/or *L*s in random orientations. Surprisingly, if the natural scene was replaced by a large peripheral *T* or *L* or by a bisected two-colored disk to be identified under the same conditions, performance dropped substantially in the dual-task condition. This makes it hard to argue that the perceptual load involved in the *T*–*L* search task was insufficient to prevent additional processing of peripheral stimuli

Correspondence concerning this article should be addressed to Karla K. Evans, Department of Psychology, Princeton University, Princeton, NJ 08544-1010. E-mail: kevans@princeton.edu

(following the load theory of attention proposed by Lavie, 1994). Li and her colleagues concluded that high-level representations can be accessed outside the focus of attention.

These findings seem to be at odds with the conventional view that focused attention is required for high-level visual processing. Focused attention has been shown to be needed for many visual tasks, including feature-binding to form objects and change detection in complex scenes (Mack & Rock, 1998; O'Regan et al., 1999; Rensink et al., 1997; Treisman, 1998; Treisman & Gelade, 1980; Wolfe & Bennett, 1997). How does one reconcile these apparently paradoxical findings? Are natural scenes "intrinsically superior stimuli to 'simple' geometric shapes," making their processing more efficient, as suggested by Braun (2003, p. 8)? Does scene categorization reflect a particular parallel mode of processing, as suggested by Thorpe and Fabre-Thorpe (2001)? Perhaps the meaningful context generates some form of synergistic priming facilitating the detection of target objects that naturally belong in the scene.

We propose another possible account that should be tested before reaching these conclusions. The hypothesis is that in scene categorization tasks, participants may rapidly and in parallel detect disjunctive sets of unbound features of the target category and then use these to discriminate between scenes that do or do not contain the target without necessarily fully identifying it. As we discuss later, the features we are referring to need not be simple hardwired physical features but may be learned features of intermediate complexity (cf. Ullman, Vidal-Naquet, & Sali, 2002) that characterize a target category (e.g., beaks or open wings for birds, smooth shapes and metallic textures for vehicles). Rather than a fixed sequence of processing stages, perception would comprise a first rapid pass through the hierarchy of visual processing up to the nodes in a visual identification network, followed by optional reentry to earlier levels for more detailed analysis and binding (Di Lollo, Enns, & Rensink, 2000; Hochstein & Ahissar, 2002; Treisman, 1996). In the animal categorization task, the initial pass would register a number of animal features in parallel without binding them. These could prime the target category in the identification network sufficiently for a detection response. Automatic feature priming of semantic nodes could be highly adaptive in everyday vision, helping observers to overcome the impoverished perception shown with arbitrary stimuli in laboratory tasks. If the scene is still available, attention would then be brought to bear on the location of a target feature, allowing the set of features to be bound together and an object file containing a more specific representation of the animal to be formed (Kahneman, Treisman, & Gibbs, 1992).

The idea that feature detection could mediate high-level discriminations, such as the presence of any animal, may seem initially implausible. However, we should note that feature search is not limited to a single target feature. Instances of a disjunctive set of at least four features (any of which could appear as target) can be detected through parallel processing (Treisman, 1988). Moreover, efficient search is not restricted to simple elementary features. There is evidence for parallel detection of fairly complex features, including shape from shading (Ramachandran, 1988) and three-dimensional orientation (Enns & Rensink, 1991). Levin, Takarae, Miner, and Keil (2001) explored the discrimination of animals and artifacts in search through displays of line drawings or colored pictures. They found that target–distractor differences in

curvilinearity or rectilinearity, in global contour shape, and in the visual typicality of parts and form played an important part.

Our first set of experiments tested this feature priming idea by exploring what information participants have about the targets they detect, in particular whether they can identify the type of animal better than would be expected from unbound feature information. We also tested to what degree the targets are bound to their locations in the scene. Finally, we compared the accuracy of detecting an animal target in a rapid sequence of natural scenes when the distractor scenes contain images of humans and when they do not. The feature sets that characterize animals are likely to overlap in part with those that characterize humans. Thus, if animal detection depends on the detection of disjunctive sets of animal features, the presence of humans should make the task substantially more difficult.

The second set of experiments used an attentional blink paradigm to further test the attention capacity available for visual categorization in natural scenes. Note that attention limits and parallel processing are already probed in detection of a single category target in a rapid serial visual presentation (RSVP) of single scenes because (a) a single scene often contains many objects, and (b) several scenes are presented in rapid succession. However, the load can be substantially increased by presenting more than one target in the RSVP sequence (or by presenting two scenes at once). Our goal in Experiments 4–7 was to see whether target detection induces a temporary refractoriness of attention as it does with other supposedly simpler targets like letters and shapes (Broadbent & Broadbent, 1987; Chun & Potter, 1995; Raymond, Shapiro, & Arnell, 1992). Finally, we compared search for two successive targets from the same category (thus mediated by the same set of features) and from two different categories (requiring a larger combined set of target features).

We should emphasize that without knowing the set of intermediate level features that are used, our hypothesis is difficult to prove. All we claim for the present experiments is that they are consistent with the hypothesis that animal detection could be based on parallel detection of disjunctive features without attentional binding and that this alternative account should not be dismissed without further consideration, because it could reconcile two apparently conflicting sets of findings. Of course, research will be needed to explore in more detail the nature of the features mediating detection, if such in fact exist.

Experiment 1

If animal detection is based on priming from disjunctive sets of typical animal features, as we suggest, some testable predictions should follow. Although detection might be quite efficient, identification of which animal is present should be much less so. In many cases, the features might point to a superordinate category, like bird (wings, beaks, feathers), mammal (fur, four legs), or fish (body shape, gills, etc.), without identifying the particular species within that category. We used RSVP of a string of pictures to reduce or eliminate the time available for feature binding following category detection. We set the presentation rate very high, to approximate the durations and detection rates reported by Li et al. (2002), because their findings were the basis for some of the most dramatic claims of attention-free detection of categories.

Participants were asked to press a key as soon as they detected a target. To minimize the memory load and eliminate any interference from subsequent pictures, we designed the experiment so that the detection response interrupted the RSVP sequence, and we instructed participants immediately both to identify and to localize the target they had just seen (to the left, right, or center of the display). If detection is mediated only by unbound features, identification should frequently fail, as should localization if the features are not bound to their spatial locations.

Detection should also be impeded by the presence of distractors that share similar features. Thus, if an animal target is embedded in a sequence of scenes that also contains humans, performance should be worse than if it is presented in a sequence of scenes without humans. To rule out the possibility that, rather than interfering through feature overlap, human distractors might be intrinsically distracting and disrupt the search by attracting attention away from the animal targets, we tested a separate group of participants with the same sets of distractor slides, with and without humans, but replaced animal targets with vehicle targets. Because these share few or no features with humans, detection should be equally high with or without people.

Method

Participants. Sixteen students from Princeton University (9 men, 7 women), between the ages of 18 and 30 years, with normal or corrected-to-normal vision, participated in this study for course credit. Informed consent was obtained before the beginning of the experiment. Half the participants were tested with animal targets and half with vehicle targets. The type of distractors (human or not) was varied within participants.

Apparatus and stimuli. The stimuli used in this and subsequent experiments were displayed on a G3 Mac computer (screen resolution: 1024 × 768; vertical refresh rate: 75 Hz) in a dimly lit room at a viewing distance of 57 cm. Stimuli were color photographs of natural scenes taken from a large commercial CD-ROM library (Corel, 1996). We selected from this bank 2,456 distractors, 288 animal target images, and 288 vehicle images, chosen to be as varied as possible. Animal targets included birds, fish, insects, mammals, and reptiles. Vehicle targets included cars, trucks, military and civil airplanes, boats, helicopters, trains, and hot air balloons. The animal and vehicle targets in the pictures were mostly of single animals or vehicles and included both full body images and face only images of animals and whole vehicles as well as just the front ends of vehicles (see Figure 1). There were no restrictions on the size or number of animals or vehicles in any particular photograph, but the locations were constrained so that an equal number of each category were to the left, to the right, or in the central area of the slide. The distractor images were also drawn from a wide range, including outdoor and indoor scenes, natural landscapes (oceans, mountains, forests, deserts, and beaches) and cityscapes, pictures of people, food, plants, buildings, and other man-made objects (see Figure 1). The images of people in the scenes included faces of people as well as full body images, and there were scenes with a single person and scenes with groups of people. The images comprised 800 × 570 pixels, subtending roughly 20° visual angle in width and 13° visual angle in height. They were presented at the center of the screen for 75 ms each with no ISI.

Procedure. Participants initiated each trial by pressing a button on the keyboard. A small fixation cross then appeared in the center of the screen where the images were to appear. After a 1,000-ms delay, an RSVP of six temporal frames began with no blank intervals between them. Each temporal frame consisted of one image lasting 75 ms. A target scene (containing one or more animals or vehicles) appeared in half the trials, at either the second, third, fourth, or fifth frame, chosen randomly on each trial with the constraint that there were equal numbers in each of these positions (see Figure 2). Therefore, a target was followed equally often by four, three, two, or one image frames. The target appeared with equal probability to the left, center, or right side of the image.

Participants were asked to respond as quickly and accurately as possible by pressing an assigned key each time a target was present (an animal for one group and a vehicle for the other). When the RSVP sequence did not contain a target, participants were asked to press another assigned key. As soon as the participants pressed the key indicating that they detected a target, they were prompted with a question asking them what animal or

Figure 1. Examples of photographs used in Experiments 1–7. First row: Animal target slides. Second row: Vehicle target slides. Third row: People distractor slides. Fourth row: No people distractor slides. All images are from the Corel Stock Photo Library (Corel, 1996). Copyright 1996 by Corel. Reprinted with permission.
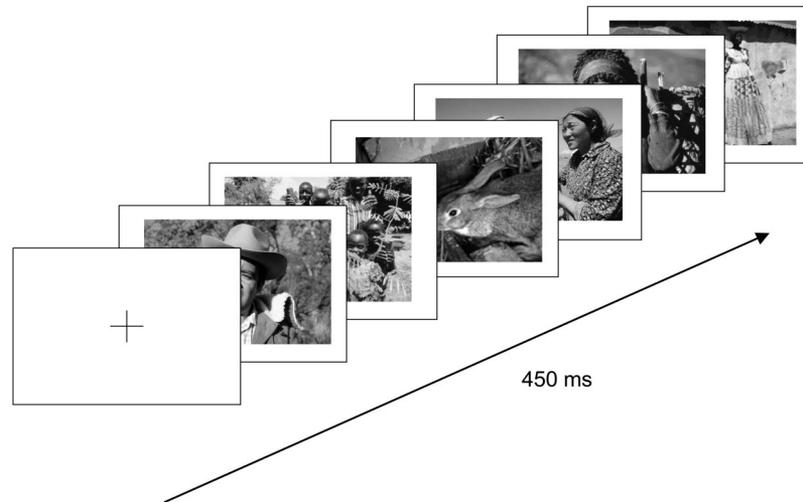
Figure 2. Example of a trial with an animal target among people distractors in Experiment 1. All images are from the Corel Stock Photo Library (Corel, 1996). Copyright 1996 by Corel. Reprinted with permission.

what vehicle they saw and whether they saw it to the right, center, or left side of the image. If they were unable to identify the target exactly they were asked to give whatever information they could, including the superordinate category, any features that they saw, and so on. After they typed in their responses or pressed a key indicating that they did not detect a target, a fixation cross reappeared to indicate the beginning of a new trial. Upon the completion of the RSVP sequence, participants had 1,000 ms to respond (i.e., to press the key if they had not already done so) and then to type in the identification, after which delay their response was considered as an error. There were two errors across all participants, and these were discarded from further analysis. We recorded reaction times to press the key for detection as well as response accuracy, both for detection and for identification and localization.

Each participant took part in one experimental session consisting of two conditions. In the *people as distractors* condition the distractor photographs all contained images of humans, whereas in the *no people* condition the distractor photographs contained scenes with no images of humans. An experimental session included 14 blocks of 12 trials each, in which trials with people as distractors and trials with no people as distractors were randomly mixed. At the beginning of the session, participants were also given a block of 12 practice trials whose data were not used. To prevent learning, we allowed each image to be seen only once by each participant. Image sequences within trials were randomized so that different participants were presented different image sequences in the different conditions. Condition effects on the behavioral measurements were assessed by analysis of variance (ANOVA) with Greenhouse–Geisser correction for nonsphericity.

Results and Discussion

Figure 3 shows the correct detection rates in Experiment 1. As we had hoped, the detection rates (78%) with the scenes with no people approximated those found by Li et al. (2002), although we used RSVP and they used single masked presentations. This allowed us to explore what other information is available when detection rates are at that level.

An ANOVA showed a significant effect of distractor type (people vs. no people), $F(1, 14) = 14.00$, $p < .002$. The main effect of animal versus vehicle targets was not significant, but there was a significant interaction between target type and distrac-

tor type, $F(1, 14) = 13.76$, $p < .002$, showing that people distractors interfered much more with the detection of animals than with the detection of vehicles. In fact, with the vehicles, the people distractors caused no decrement at all. The false-alarm rates on target-absent trials showed no significant differences when the distractor slides contained people (7% false alarms for animal targets and 8% for vehicle targets) and when they did not (6% false alarms for both animal and vehicle targets). The ANOVA on the reaction times showed an almost significant interaction of distractor type and target type, $F(1, 14) = 4.48$, $p < .053$. With animal targets, the slides with people as distractors slowed the reaction times on correct detection trials, giving a mean of 634 ms ($SE = 80$ ms), whereas the reaction times without people distractors averaged 559 ms ($SE = 80$ ms). With vehicle targets, the reaction times on correct detection trials showed no significant effect of distractor type, averaging 511 ms ($SE = 40$ ms) with people as distractors and 515 ms ($SE = 30$ ms) with no people as distractors.
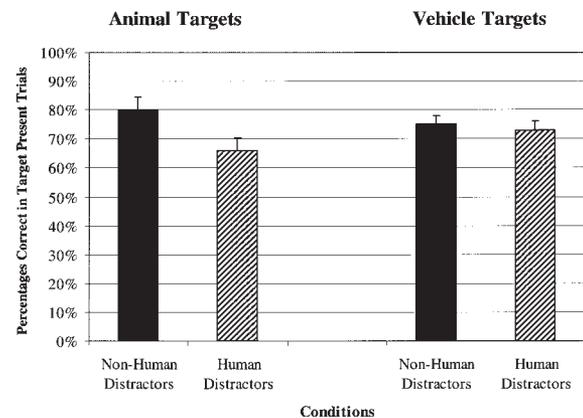

Figure 3. Percentages of correct detections on target-present trials for animal targets and for vehicle targets without and with humans present in the distractor scenes. Error bars are standard errors of means.

Thus, the presence of humans, with sets of features of which many are shared by other animals, severely impaired performance in the animal detection task, as predicted, but had no effect at all on the vehicle detection task. The interference cannot therefore be due to humans being intrinsically interesting and distracting.

To check whether later pictures were interfering with either detection or verbal reports of earlier ones in the sequence, inducing a kind of instant amnesia, we analyzed the serial position effects. Detection was completely unaffected, with means of 74%, 73%, 73%, and 74% for Positions 2, 3, 4, and 5, respectively. The same was true of correct identification responses, which averaged 44%, 40%, 42%, and 46%, respectively. These findings are important because they give no evidence that retroactive interference with memory for earlier pictures resulted from subsequent pictures seen before the participant had time to respond.

Correctly detected targets were also very poorly localized (to the left, center, or right of the scene). Of the overall 73% correct detections of animals, on average only 53% were correctly located (for which chance was 33%). Of the overall 74% correct detections of vehicles, 56% were correctly located. So on average, participants were correct in both detecting and locating the targets on only 39% of all trials. The poor performance suggests a failure to bind either the features to the features for identification or the features to their locations for localization.

We look next in more detail at the nature of the information on which these detections were based, by analyzing the typed reports on trials on which the target was correctly detected. First, let us look at the animal targets: In the no people condition, 80% were correctly detected, and of these, 44% were correctly identified. In the people condition, 66% were correctly detected, and of these, 42% were correctly identified, making up fewer than half of the detected targets and only one third overall of the targets presented. Note that in determining these percentages, we scored the animal names quite leniently, accepting, for example, *deer* for oryx or elk and *butterfly* for moth.

If participants rely mainly on features in detecting the targets, they should often be able to identify the superordinate category even when they are unable to identify the particular animal. Mammals do not have wings or beaks, and reptiles are seldom furry. Table 1 shows the distributions of responses with each set of

distractor slides. It includes the number of correct superordinate responses (such as *mammal* for lion or *insect* for beetle) and the number of errors taking the form of another animal from the correct category (e.g., *cow* for moose or *dog, ferret,* and *wolf* for fox). The total percentages of correct category responses (i.e., the number correctly identified, plus the correct superordinates, plus the same category responses) were 79% of the correct detections for the no people condition and 76% for the people condition. If we add the trials on which some features were correctly described, the totals are 95% and 92%, respectively. The numbers of incorrect responses that were also in the wrong category (e.g., *camel* for pelican or *swan* for panther) were only 5% for no people slides and 8% for people slides. These may correspond to false alarms made to some other slide in the series. Thus, when participants reported detecting a target, they almost always knew something about it. However, our results show that more than half the correct detections were based on partial information rather than full representations of specific animals. Note that this is unlikely to be due to the specific information not being potentially available. There was almost always at least one correct response given to each target. Only 3% had no correct identifications. The main effect of the people distractor slides was on the detection rates. Given a correct detection, the additional information available is quite similar with people and no people distractors.

What sort of features might be involved in these rapid perceptual responses? We classified the feature responses to animal targets under a number of headings. In this analysis, we pooled the results with those of Experiment 3 using the same task and the same animal slides with a slower presentation rate. We also included features that were mentioned in addition to the identity or superordinate category (e.g., *yellow and black reptile, large, spotted cat*), although these did not appear in Table 1. There were a total of 124 color reports, 58 texture properties (comprising 48 mentions of *furry*, 6 of spots, 4 of *feathery*, 3 of scales, and a few other single properties like *smooth, shiny,* and *shaggy*), 55 body parts (including 12 mentions of legs, 11 of horns or antlers, 6 of tails, 4 of beaks, 4 of wings, 3 of ears, 3 of eyes, 2 of long necks, 2 of snouts, and one each of a number of other parts such as claws, fangs, hoofs, antennae, and shells), 31 mentions of size (usually *large* or *small*), and 10 shape reports (such as *pointy, round, long,*

Table 1

Percentages of Different Identification Choices for Correctly Detected Targets in Experiment 1: 75-Millisecond Exposure

Target and distractor type	Specific identity (e.g., *cow*)		Superordinate (e.g., *mammal*)		Same category (e.g., *deer*)		Features (e.g., *brown* or *large*)		Different category (e.g., *goose*)	
	M	SE	M	SE	M	SE	M	SE	M	SE
Animal targets										
No people	44	3	28	2	7	2	16	2	5	1
People	42	2	26	3	8	2	16	2	8	2
Vehicle targets										
No people	52	2	32	2	1	1	14	2	1	1
People	53	2	30	3	0		15	2	2	1

Note. Specific identity means that the identification was as accurate and specific as it would have been with unlimited exposure. So, for some pictures, it might, for example, be *bulldog* or *polar bear,* but for others in which the animal was small or less clear, it would be *dog* or *bear.*

or *thin*). These features were almost all correct. They give us some hints about what is perceived in these brief exposures. Note that often two or three features would be enough to give a good guess at the animal—for example *pink*, *stubby* suggests a pig, and *brown*, *furry*, *with antlers* suggests a stag. It seems possible that many of the correct identifications were based on similar feature information.

Results for the vehicle targets were similar. The identification and categorization responses were somewhat better than those for the animals, probably because the vehicle targets vary much less within categories. The responses were also harder to classify than those to the animals. There were only about six categories (plane, car, train, boat or ship, flying object [for balloon], land vehicle) and about 35 specific names (such as canoe, Oldsmobile, army helicopter, steam engine) compared with 68 for animals. We sometimes counted the category name as a correct identification if the more specific name was hard to select, for example, if the car was small or if there were many of them in the picture, but if the vehicle was large and clearly a particular type (like a Volkswagen Beetle, or a canoe, or a hot air balloon), we asked for the more specific name as a correct identification. Because of the somewhat arbitrary boundary between identification and classification, it is probably not useful to compare the identification rates with those for the animals. Thus, when there were no people present the detection rates were slightly higher for animals (80% for animals and 75% for vehicles, as shown in Figure 3), but the percentage of those detected targets that were also correctly classified was slightly higher for the vehicles. Overall, the pattern of results was similar for animals and for vehicles, and the same conclusions seem to be supported.

Experiment 2

Experiment 1 showed considerably lower detection rates for animal targets when the distractor slides contained humans. The interpretation we favored is that the presence of humans in the distractor scenes interfered because of the feature overlap between humans and animals and not because of their intrinsic interest or attention-catching quality, which would have impaired the vehicle targets as well. However, the experiments so far do not rule out the possibility that the interference might be due to semantic confusion. Humans are also animals, so rejecting them as nontargets (as participants were instructed to do) may have required extra thought or effort. To test for this possibility, in Experiment 2 we replaced the pictures with words. The targets were animals (e.g., *tiger*, *butterfly*, *worm*) and the distractors were either people (e.g., *child*, *doctor*, *thief*) or control words (e.g., *rose*, *city*, *engine*). Experiment 2 was otherwise identical to Experiment 1, with the exception that the target location was not reported, because all the words were presented at the center of the screen. Participants were required to search for an animal name among a sequence of words that in some instances denoted humans and in others denoted objects, places, and plants.

Method

Participants. Eight students from Princeton University (4 men, 4 women), all native English speakers between the ages of 18 and 30 years, with normal or corrected-to-normal vision, participated in this study as paid participants ($10 per hour). Informed consent was obtained before the beginning of the experiment. All the participants were tested with animal targets. The type of distractors (human or not) was varied within participants.

Stimuli. The stimuli were words equated for frequency taken from the Medical Research Council's Psycholinguistic database (Wilson, 1988). We selected from this bank 2,456 distractor words (all concrete nouns) and 288 animal target words (animal names), chosen to be as varied as possible. Animal target words included names of birds, fish, insects, mammals, and reptiles. There were restrictions on the word lengths of the animal names, because we equated these with the word lengths of the distractor words. The distractor words were also drawn from a wide range. The words naming people included words naming professions (e.g., *baker*, *professor*), relations (e.g., *aunt*), and other descriptive categories (e.g., *spinster*, *rogue*, *toddler*). The control words named food, plants, scenes, buildings, and other man-made objects. The words subtended 2.5° to 9.5° in width × 1° in height of visual angle and were presented at the center of the screen for 75 ms with no interval between word presentations.

Apparatus and procedure. The apparatus and procedure were identical to those in Experiment 1.

Results and Discussion

The mean percentages of correct detections for animal word targets was 87% both in the condition with distractor words denoting people and in the control condition with other words (the standard errors were 2.2% and 1.5%, respectively). There were no differences in false-alarm rates (4% in both conditions). The reaction times on correct detection trials also showed no substantial difference, with the mean animal detection latencies averaging 568 ms ($SE = 30$ ms) with people distractors and 566 ms ($SE = 40$ ms) with control words.

The major finding of this experiment is that even though animals and humans belong to the same semantic category, animal names are no harder to detect among people names than among other object names. Thus, semantic confusion cannot explain the difference we observed in Experiment 1. Visual feature overlap remains the most likely cause of the impairment.

Experiment 3

Without knowing what features mediate detection (if any), it is difficult to test our hypothesis. However, we can collect information on a range of manipulations to see if they fit the story or raise difficulties for it. One simple manipulation is to invert the scenes. This should make the task more difficult on almost any theory, but it might be expected to interfere more with semantic processing and identification than with processing of the physical features, at least for those that are orientation invariant. Thus, if the problem in search for animal targets with the people distractors arises from shared features, the relative difficulty of people distractors should remain as great when the scenes are inverted as when they are presented in their normal orientation. To the extent that the interference arises at the semantic level or after identification, the relative difficulty should be reduced.

A second goal in this experiment was to replicate the first experiment with a slower presentation rate to see how identification rates change relative to detection rates. Does the information also get more specific with an extra 35 ms, or do participants simply detect more targets?

We repeated the animal target conditions of Experiment 1, adding two conditions in which the images were all inverted. The exposure time of each image was increased from 75 ms to 110 ms, but the procedure was otherwise the same.

Method

Participants. Six students from Princeton University (4 men, 2 women), all native English speakers between the ages of 18 and 30 years, with normal or corrected-to-normal vision, participated in this study for course credit. Informed consent was obtained before the beginning of the experiment. All the participants were tested with animal targets. The type of distractors (human or not) as well as the orientation (inverted or upright) was varied within participants.

Apparatus and stimuli. The apparatus and stimuli were identical to those in Experiment 1.

Procedure. The procedure was the same as for Experiment 1 except that half the trials used inverted pictures. These were randomly mixed with upright ones within the experimental session. In the inverted condition, whole photographs were inverted 180°. Both the upright and the inverted conditions included sets of trials with people distractors and sets with no people, again randomly mixed within blocks. An experimental session included 14 blocks of 12 trials each, with a block of 12 practice trials at the beginning of the session whose data were not used. To prevent learning, we allowed each image to be seen only once by each participant. Image sequences within trials were randomized so that different participants were presented different image sequences in the different conditions. Condition effects on the behavioral measurements were assessed by ANOVA with Greenhouse–Geisser correction for nonsphericity.

Results and Discussion

Participants correctly detected the animal targets on 46% of trials in the inverted conditions and 81% in the upright conditions. There was a significant difference in the rate of correct detections among the inverted and upright images as expected, $F(1, 5) = 129.85, p < .00001$. There was also a significant effect of distractor type (51% with people and 76% with no people), $F(1, 5) = 114.19, p < .00001$. There were no significant differences in false-alarm rates for inverted versus upright ($M = 9\%, SE = 3\%$ vs. $M = 6\%, SE = 2\%$) or for people versus no people as distractors ($M = 6\%, SE = 2\%$ vs. $M = 9\%, SE = 3\%$).

The most interesting finding in this experiment is that the decrement caused by people distractors was about the same when the scenes were upright and when they were inverted (see Figure 4). The interaction of distractor type and inversion was not significant, $F(1, 5) = 1.63, p < .257$. This seems more consistent with the idea of feature-based detection than with the idea of full semantic processing and identification. It also suggests that the relevant features are to a substantial degree invariant with respect to orientation.

Rousselet, Mace, and Fabre-Thorpe (2003) also tested the effects of inversion on the faces of animals or humans. They found that inversion did not significantly impair performance on a rapid detection task, whereas full identification was substantially impaired by inversion. Our results show a large drop in detection with inversion, as well as a drop in identification. Our targets were scenes rather than faces, and the features were therefore much more varied than would be found in faces and may have been more dependent on the context.

We can compare across Experiment 1 and the upright conditions of Experiment 3 to see the effect of presentation rate. Not surpris-
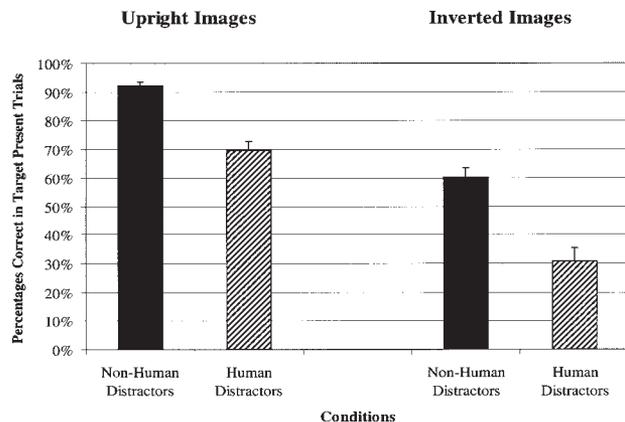

Figure 4. Percentages of correct detections on target-present trials for animal targets in the upright and inverted conditions without and with humans present in the distractor scenes. Error bars are standard errors of means.

ingly, the detection rate was higher when each picture was seen for 110 ms than when it was present for just 75 ms, $F(1, 12) = 55.24, p < .00001$. However, the mean proportion of the detected targets that was also correctly identified did not differ significantly, and neither did the correct categorizations (identification plus superordinates plus same category responses). This means that participants kept the same criteria for pressing the detection key across the two experiments and did not gain more specific information once the target was detected in the additional 35 ms that they had available in Experiment 3. There was, however, an interaction with people distractors and no people distractors, $F(1, 12) = 17.65, p < .001$, such that with longer exposures, identification improved for animals in scenes without people but actually decreased for animals in scenes with people. There was a corresponding increase in the correct superordinates with the slower presentation of people scenes, so that the total classification plus identification plus same category responses were the same with people and with no people. The presence of people seems to have reduced the specificity of the animal representations at the slower presentation rate, perhaps because the people themselves became more visible at the slower rate and so could interfere more.

Experiment 4

Experiments 1, 2, and 3 explored some aspects of category detection in visual scene perception and obtained data consistent with the idea that simple detection could be mediated by parallel sets of features at some stage prior to full binding and identification. In Experiment 4, we tested the more severe attentional constraints that arise when two targets must be identified within close temporal proximity. When participants search for two targets in an RSVP sequence of distractors, their capacity to detect the second target (T2) presented within 150 ms to 500 ms after the first target (T1) is severely impaired, a phenomenon known as the attentional blink (Broadbent & Broadbent, 1987; Chun & Potter, 1995; Raymond et al., 1992). The deficit in reporting the T2 in the RSVP sequence is the result of attending to the T1, because there is no difficulty in reporting T2 when it is the only target to be

detected (Joseph, Chun, & Nakayama, 1997; Raymond et al., 1992; Seiffert & Di Lollo, 1997). Experiment 4 was designed to test whether natural scene perception and categorization would show attention limits (an attentional blink) in this dual-task paradigm or be more generally independent of attention, in the sense that two targets embedded in natural scenes can be identified without interference when they appear within a short interval of each other.

Most accounts of the attentional blink place the interference at a stage of processing at which the representation of a visual event is consolidated and made available for conscious report. Within our framework, consolidation would correspond to the reentry check resulting in the formation of a bound object file. An alternative account is that both targets are fully identified perceptually but that access to working memory forms a bottleneck at which storing T1 prevents the additional storage of T2 within a certain interval of time (Giesbrecht & Di Lollo, 1998; Vogel & Luck, 2002; Vogel, Luck, & Shapiro, 1998). Working memory is unlikely to explain the limits we observed in Experiments 1 to 3. An immediate response to the targets was required, and the presentation ended as soon as the response key was pressed. Thus, the rate of perceptual processing must have set the limits. However, in attentional blink experiments, working memory may introduce further limits to capacity.

We asked participants to detect and identify two targets in a sequence of natural scenes in which we varied the temporal lag between the two targets. Performance with two targets was compared with a control condition in which participants were required to report only one target. We were interested also to see whether there is a difference in the magnitude of the attentional blink when two successive targets come from the same category and when they come from different categories. We compared performance when the targets were both animals or both vehicles and when there was one of each category. When two very different categories are used, the number of possible diagnostic features is greatly increased, which might make the task more difficult. Same category targets might also benefit from within-category priming from the features of T1 to those of T2.

To study the attentional blink, we used presentation rates that have typically been used in other attentional blink studies, so that we could compare our results with those obtained with other types of stimuli. The rate we chose (9.1 frames/s) was considerably slower than the rate in Experiments 1 and 2 (13.3 frames/s), for which we were more interested in comparing performance with that obtained with brief exposures in the single- and dual-task conditions by Li et al. (2002).

Method

The apparatus and stimuli were the same as those for Experiment 1 except that we increased the number of stimuli to 3,456 distractors and 576 targets. The images were 800 × 570 pixels, subtending roughly 20° visual angle in width and 13° visual angle in height. They were presented at the center of the screen for 110 ms each.

Participants. Thirteen students from Princeton University (7 men, 6 women), between the ages of 18 and 30 years, with normal or corrected-to-normal vision, participated in this study for course credit. Informed consent was obtained before the beginning of the experiment. All of the participants were tested in both the experimental and the control conditions.

Procedure. Participants initiated each trial by pressing a button on the keyboard. A small fixation cross then appeared in the center of the screen. After a 1,000-ms delay, an RSVP of 12 temporal frames began, each one centered at fixation. Each temporal frame consisting of one image lasted 110 ms, yielding a presentation rate of 9.09 frames per second. T1 appeared equally often in each of Positions 2–9, with these positions presented in randomized order. The identity of T1 depended on the block and could be either an animal or a vehicle. T2 could also be either an animal or a vehicle, depending again on the block. It was presented equally often two frames, four frames, six frames, or eight frames after T1, except that T2 never appeared in the last frame. All the possible pairs of positions for each lag were used equally often, which of course means that T2 appeared more often toward the end of the sequence than earlier (e.g., 3 times in Position 4 and 25 times in Position 11). There were no target-absent trials. The stimulus onset asynchrony (SOA) between the two targets varied between 220 and 880 ms, in steps of 220 ms. T2 was presented 24 times at each of the four SOAs.

There were six conditions, run in separate blocks with the order counterbalanced across participants: (a) respond to T2 only, animal target; (b) respond to T2 only, vehicle target; (c) two targets, same category, animals; (d) two targets, same category, vehicles; (e) two targets, different categories, T1 vehicle, T2 animal; (f) two targets, different categories, T1 animal, T2 vehicle. In the one-target control blocks, both types of targets were present (one from each target category), but the participants were instructed at the beginning of the block to report only the preassigned target category, which was always presented in the T2 position, and to ignore the unassigned target category. They were required to press a key when they detected the target and then later were required to report its identity. In the two-target conditions, the participant's task was to look either for two appearances of targets from the same category (two instances of vehicles or two of animals) or for two different category targets (an instance of an animal and an instance of a vehicle) among 10 distractor images. When the targets were from the same category, they were told which it would be. When they were from different categories, participants were not told which would be first, but they knew that the second would be from the other category. After each trial they were asked to identify the target (or targets) they saw, by typing the names using the keyboard. At the end of the RSVP sequence, the fixation cross reappeared to indicate the beginning of a new trial. Participants were allowed to take as much time as they needed to respond as accurately as possible.

Each participant took part in one session that included 96 trials in each of the six conditions, in separate blocks of trials. Before these experimental trials, participants completed 12 practice trials with 2 trials of each condition. The images used here were not repeated in the main experiment. To further reduce the effects of specific learning, we presented a target image once only for each participant, and it could not appear both as a target in one task and as a distractor in another task. The distractor images were presented twice for each participant but in a different order. Image sequences were randomized, so that different participants were presented different image sequences over the blocks of 96 trials. The order of the six conditions was counterbalanced across participants so that each condition was presented the same number of times in each position of the experiment (except that one order was presented twice, because there were 13 participants).

Results and Discussion

We look first at T1 accuracy in the two-target conditions, which averaged 91% ($SE = 2\%$) correct identifications for vehicles and 97% ($SE = 2\%$) for animals. A repeated measures ANOVA on the T1 responses showed a significant effect of target type, with animals more accurate than vehicles, $F(1, 12) = 9.99$, $p < .008$. Because of the blocked presentations, participants knew the T1 category when they were congruent and did not know it when they

were incongruent. This allows us to look at the effects of prior knowledge of the category on accuracy of T1 identification. The ANOVA showed no effect, suggesting that participants were able to detect a target equally well when it could be either of two categories and when it belonged to a known category.

The main interest is in performance on T2, which is normally subject to the attentional blink. An attentional blink is defined as a drop in performance on the T2 task when T1 is also required relative to the control condition when T1 is not required. To count as a blink, performance should also recover as the lag between T1 and T2 increases.[1] Figure 5 shows the mean percentages of correct identifications of T2 as a function of T1–T2 SOA. For the two-target responses, the detections of T2 are conditional on T1 having been correctly detected.

In this experiment the scoring criteria were somewhat different from those in Experiments 1–3. The number of target pictures was increased from 168 to 576, which meant that, even given unlimited time, a smaller proportion of the pictures could be clearly identified at the more specific level. We therefore accepted more category responses as correct identifications than we had in Experiment 1 (e.g., *fish* or *car* rather than *goldfish* or *Volkswagen*). Comparisons across the two experiments are therefore not appropriate, but they are also not important to our arguments.

To assess the attentional blink deficit, we entered the percentages of correct identifications of T2 (conditional on T1 being correct when both were required) into a three-factor repeated measures ANOVA. The three factors were the lag (the SOAs between T1 and T2), the condition (single task and dual task, taking the means of same and different category targets), and the target type (animal or vehicle). Statistical analysis confirmed the graphical evidence of the dual-task performance deficit by showing significantly lower target identifications in the dual-task than in the single-task condition for both target types. There was a significant effect of lag, $F(3, 36) = 41.86$, $p < .00001$, and a significant effect of dual- versus single-task condition, $F(1, 12) = 129.05$, $p < .000001$, as well as a significant interaction of lag and condition, $F(3, 36) = 38.12$, $p < .00001$. In the dual-task conditions, performance improved as the lag between T1 and T2 increased from an SOA of 220 ms to 880 ms, whereas performance in the single-task condition remained constant. The attentional blink lasted substantially longer in this experiment than is typical with letters and other simpler stimuli, for which it is usually over within 500 to 600 ms (Chun & Potter, 1995; Raymond et al., 1992). Animals were identified better than vehicles, $F(1, 12) = 26.25$, $p < .0001$, but there was no significant interaction of target type with lag or with condition ($F < 1$), suggesting that the same attentional effects appear with both target types.

To assess the effect of congruency between T1 and T2 on the ability to identify the T2 targets, we conducted another ANOVA on just the two dual-task conditions. The three factors were the lag (the SOAs between T1 and T2), the congruency (same category and different category targets), and the target type (animal or vehicle as target). The analysis indicated an attentional blink deficit that is more severe when participants had to switch categories than when they looked for two targets from the same category. We see this in the significant lag, $F(3, 36) = 41.32$, $p < .00001$, and congruency main effects, $F(1, 12) = 23.20$, $p < .00001$, and in the interaction of lag and congruency, $F(3, 36) = 2.97$, $p < .045$. Animals were again better identified than were vehicles, $F(1, 12) = 11.64$, $p < .005$, and there was a significant interaction between target type and congruency, $F(1, 12) = 5.76$, $p < .033$, with a larger effect of congruency when T2 was a vehicle than when it was an animal.

This experiment clearly establishes that attention limits do arise at some stage when one views natural scenes. Participants showed good performance in detecting the first of two targets, whether or not the category was known in advance. However, the pronounced deficit in identifying T2 when it appeared within a few hundred milliseconds of T1 shows marked capacity limits in postdetection processing of targets in natural scenes that are at least equal to those obtained with alphanumeric and other simpler stimuli. Identifying T1 after it has been detected and consolidating the memory for later report interferes strongly with detecting and identifying T2. Marois, Yi, and Chun (2004) also found an attentional blink in detecting and classifying a natural scene (as indoor or outdoor) after identifying a face as T1 in an RSVP sequence of scrambled scenes.

A second interesting finding in our data is the increased blink when T2 belongs to a different category from T1. Two possible accounts would be that setting up a different set of relevant features to be monitored takes time or, conversely, that keeping the same set (in the same category conditions) allows feature level priming to facilitate performance.

Experiment 5

To clarify the role of category switching in T2 detection, we repeated the same conditions as in Experiment 4 except that all the conditions were randomly mixed within blocks, so that participants were always searching for both categories. We did not include a single-target condition in this experiment.

Method

Participants. Nine students from Princeton University (4 men, 5 women), between the ages of 18 and 30 years, with normal or corrected-
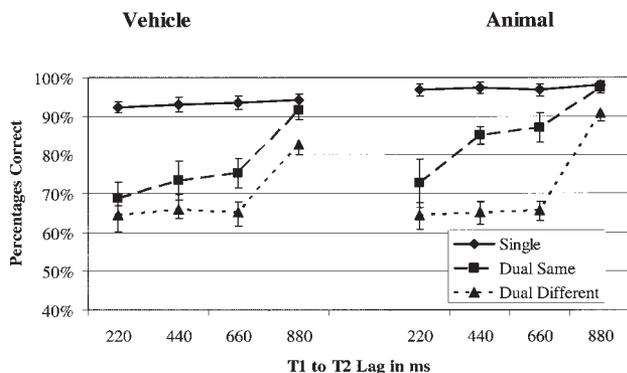

Figure 5. Correct report of the second target (T2) given correct report of the first target (T1) as a function of lag, together with results for the control condition (T2 only) in which no response was required to T1. Error bars are standard errors of means.

---

[1] We did not present T2 immediately after T1. There was always at least one intervening image, so the better performance due to grouping that is often observed at T1 + 1 did not appear in this experiment.

to-normal vision, participated in this study for course credit. None had taken part in the previous attentional blink experiments.

Stimuli and procedure. The apparatus, stimuli, and procedure were the same as for Experiment 4 except that there was no single-task control condition and the congruent and the incongruent targets in the two dual-task conditions were randomly mixed instead of blocked. The scoring criteria were the same as for Experiment 4.

Results and Discussion

In this experiment, T1 was reported correctly on 94% ($SE = 2\%$) of the animal target trials and 86% ($SE = 2\%$) of the vehicle target trials, $F(1, 8) = 17.26$, $p < .003$. Report of T2 conditional on T1 being correct averaged 77% for animals and 67% for vehicles, $F(1, 8) = 32.37$, $p < .00001$. There were no significant interactions between target type and any other factor, so we pooled across animals and vehicles in the remaining analyses. An ANOVA on the T2 results showed a significant main effect of lag, $F(3, 24) = 147.70$, $p < .001$, and a significant effect of congruence (T2 category same as T1 or different), $F(1, 8) = 9.66$, $p < .014$, but no significant interaction of lag and congruence ($F < 1$).

Figure 6 shows the probability of reporting T2 correctly given that T1 was correct, separately for Experiments 4 and 5. An analysis of the dual-task conditions of Experiments 4 and 5 together showed no significant main effect of experiment, $F(1, 20) = 2.04$, $p < .169$, but a significant interaction between lag and experiment, $F(3, 60) = 15.99$, $p < .0001$. The main effects of lag, $F(3, 60) = 140.06$, $p < .00001$, and of congruence, $F(1, 20) = 28.33$, $p < .0001$, were both highly significant, but the interactions between lag and congruence and between experiment and congruence were not significant, $F(3, 60) = 2.04$, $p < .118$, and $F(1, 20) = 2.16$, $p < .158$, respectively. Finally, the interaction among experiment, lag, and congruence approached significance, $F(3, 132) = 2.38$, $p < .078$.

The main difference between the two experiments is the shape of the attentional blink, shown in the interaction between experiment and lag. Whereas in Experiment 4 there seemed to be a floor effect, keeping performance at a minimum of around 65% even at the shortest lags, in Experiment 5 the interference continued to increase linearly as the interval was reduced. Thus, the blink was much more severe at the shortest lags when participants did not
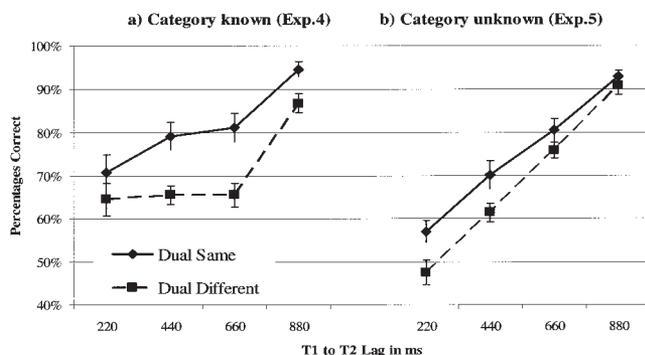
know which category to expect than when they did, whether it turned out to be the same category as T1 or a different one. When they knew that the category would switch, participants may have been able partially to prepare in advance, whereas when they had no knowledge of what to expect, identifying and consolidating the representation of T1 seems to have blocked detection of T2 for the first 250 ms on a substantial proportion of trials.

Experiment 6

The results of Experiments 4 and 5 clearly show that the identification of a category target requires attention and competes with detection of a second target appearing within the next 800–1,000 ms. Our hypothesis is that the interference arises when the features of T1 must be bound and an object representation formed in order for the target to be identified and subsequently reported. One way to test this hypothesis is to bypass the detailed perceptual processing of T1 and require only its detection—a process that we suggest can be achieved simply by registering the presence of unbound target features. In Experiment 6, the T1 task was changed to require only detection. We used the same design and stimuli as in Experiment 4 except that the participants were asked to report the identity of only T2. To ensure that they did detect T1 before identifying T2, we introduced some catch trials (20%) in which T1 was omitted and in which participants were not supposed to respond at all. To the extent that this attempt to eliminate the feature binding and identification of T1 could be successful, we predicted that the attentional blink on T2 might disappear. However, we realized that the natural spontaneous result of detecting a target is to identify it and check that it is indeed a member of the target category, so it was not clear how far our attempt to eliminate the late processing of T1 would succeed.

Method

Participants. Ten students (5 men, 5 women), between the ages of 18 and 30 years, from the paid volunteer pool at Princeton University participated in this study ($10 per hour). All had normal or corrected-to-normal vision. None had been in the previous attentional blink experiments. Informed consent was obtained before the beginning of the experiment. Three of the participants were replaced because they failed to meet the criterion of fewer than 15% false alarms on catch trials when only one target was presented.

Stimuli and procedure. The apparatus, stimuli, and procedure were the same as those for Experiment 4, except that participants responded only to T2 and only if T1 was present. To ensure that they did detect T1, we introduced catch trials (20%) on which T1 was omitted and responses were to be withheld. The scoring criteria were the same as for Experiment 4.

Results and Discussion

Performance on the catch trials was very good: False alarms to T2 when there was no T1 occurred on only 1.3% of catch trials. The mean correct identifications of T2 are shown in Figure 7.

As shown in the figure, performance was not flat across lags. There was a visible lag effect but no significant effect of target congruence. The only significant effects were the main effect of lag, $F(3, 27) = 55.62$, $p < .0001$, and a three-way interaction of

Figure 6. Correct report of the second target (T2) given correct report of the first target (T1) as a function of lag in Experiment 4 with the category known (a) and in Experiment 5 with the category unknown (b). Error bars are standard errors of means.
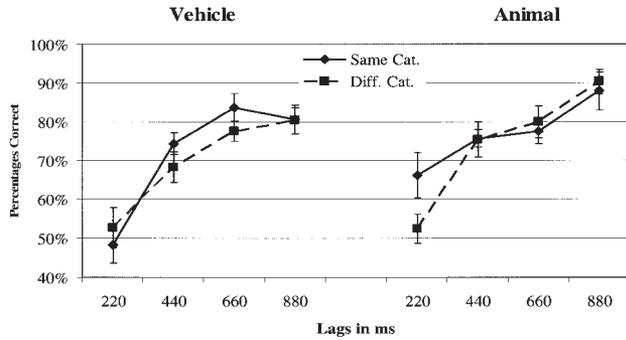
Figure 7.　Results of Experiment 6. Correct report of the second target given that the first target was present, for vehicle and for animal targets and for the first target in same or different categories. Error bars are standard errors of means. Cat. = category; Diff. = different.

lag, congruence, and target type, $F(3, 27) = 3.31$, $p < .035$. The three-way interaction seems to be due to a reduced attentional blink at the shortest lag to animal targets on same category trials relative to the three other conditions.

The main interest was in the comparison of performance on T2 when T1 was identified and when it was simply detected. To examine this, we compared performance with that in Experiment 4, which matched Experiment 6 in all other respects. An ANOVA between experiments showed no significant main effect of experiment ($F < 1$) but significant interactions of experiment and lag, $F(3, 63) = 12.29$, $p < .0001$, and of experiment and congruence, $F(1, 21) = 8.70$, $p < .008$. There were also significant main effects of congruence, $F(1, 21) = 19.60$, $p < .0001$, and of lag, $F(3, 63) = 85.28$, $p < .001$.

Figure 8 shows a much faster recovery time for the attentional blink when T1 was simply detected than when it had to be identified. The blink remains as deep as in the other experiments, but it is over much sooner. The shape of the lag effect is convex rather than concave. The remaining interference may be explained either by the fact that participants could not avoid identifying T1 or by the participants' need to confirm that they had indeed detected a target. The interaction with target congruence reflects the much smaller difference between same and different category
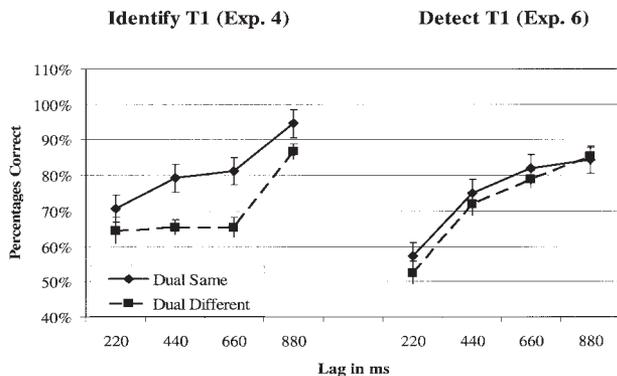

Figure 8.　Correct identification of the second target (T2) when the first target (T1) was also identified (Experiment 4) or when T1 was simply detected (Experiment 6). Error bars are standard errors of means.

targets in Experiment 6 than in Experiment 4. The category of T1 is clearly less important when T1 is merely detected than when it must also be identified.

Asymptotic performance is lower in this experiment than in the previous ones (85% compared with more than 90%). This may be because participants sometimes missed T1 and therefore did not respond to T2, thinking it was T1. In the previous experiments, they were supposed to identify any target that they detected.

Experiment 7

In Experiment 6, we showed a reduced attentional blink when participants had only to detect T1 rather than to identify it, but there was still a substantial decrement in correct identifications of T2 at the shortest lag. In Experiment 7, we tested what would happen when both targets were detected but neither had to be identified. According to our hypothesis, participants might be able to do this task solely using parallel feature detection. If so, this should further reduce the competition for attentional resources and eliminate the blink.

Method

Participants.　Ten students (6 men, 4 women), between the ages of 18 and 30 years, from the paid volunteer pool at Princeton University participated in this study ($10 per hour). All had normal or corrected-to-normal vision. None had been in the previous attentional blink experiments. Informed consent was obtained before the beginning of the experiment.

Stimuli and procedure.　The apparatus, stimuli, and procedure were the same as those for Experiment 5, except that participants were asked only to detect T1 and T2, by pressing a button on the left immediately when they detected T1 and a button on the right immediately when they detected T2. Reaction times for the detection responses were also measured. To test for false positives, we made 30% of the trials catch trials with only one target; T2 was omitted. Participants were given 12 trials of practice. Trials on the four different combinations of T1 and T2 target types (vehicle–vehicle, animal–animal, animal–vehicle, and vehicle–animal) were randomly mixed.

Unlike the previous experiments that included identification responses, this one gave us no way to verify that the keypress responses were true detections rather than false alarms. The fact that there were very few false alarms in the earlier experiments is reassuring. We can also check that the reaction times are plausible. We used the following criteria for accepting responses as genuine: (a) Two responses were made. (b) The response to T1 was made with the left key and the response to T2 was made with the right key. (c) The reaction time was not less than 200 ms or more than 1,100 ms after the onset of the target, except that for T1 responses they could be longer than 1,100 ms if they met all the other criteria. We chose 200 ms and 1,100 ms because those were the shortest and longest reaction times in Experiment 1 that resulted in a correct identification. In the two target conditions, we found 4% of cases in which the reaction time to T1 exceeded the upper limit and was not followed by T2, and we labeled them as T2 detections instead, but did not score them correct because there was no prior correct detection of T1.

Results and Discussion

T1 was correctly detected on 91% ($SE = 4%$) of trials, with a mean reaction time of 580 ms ($SE = 30$ ms). Detection rates for T2 averaged 78% ($SE = 4%$), with a mean reaction time of 686 ms ($SE = 20$ ms). Participants made false-alarm responses to T2 when only T1 was present on 18% of catch trials. Figure 9 shows
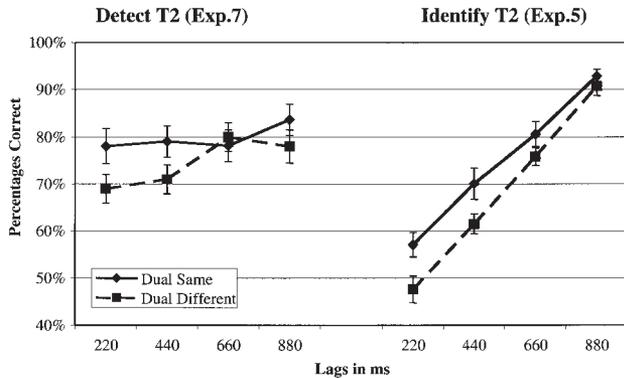
Figure 9. Correct detection of the second target (T2) when the first target was also detected (Experiment 7) or when both the first target and T2 were identified (Experiment 5). Error bars are standard errors of means.

performance on T2 with congruent and incongruent category targets. As shown in the figure, performance was almost flat across lags. The effect of lag on detection rates fell just short of significance, $F(3, 27) = 2.90$, $p < .053$. In addition, there were significant main effects of congruence, $F(1, 9) = 5.36$, $p < .046$, and of target type (animal or vehicle), $F(1, 9) = 15.98$, $p < .003$. Again, as in Experiments 4 and 5, we see the effect of congruence, but as in Experiment 5, it does not significantly interact with lag, $F(3, 27) = 1.81$, $p < .169$.

The main interest was in the comparison of performance on T2 when both targets were to be identified (in Experiment 5) and when they were simply to be detected. To examine this, we compared performance across the two experiments, which matched in all other respects. An ANOVA between experiments showed a significant main effect of experiment, $F(1, 17) = 4.61$, $p < .046$, of congruence, $F(1, 17) = 14.37$, $p < .001$, and of lag, $F(3, 51) = 66.49$, $p < .0001$, as well as a significant interaction of experiment and lag, $F(3, 51) = 29.95$, $p < .0001$. Figure 9 shows a much steeper effect of lag in Experiment 5 when the targets had to be identified and a lag effect that is attenuated or actually eliminated when both targets were simply detected.

There are two important findings in Experiment 7: (a) The attentional blink is eliminated with same category targets and only marginally present with different category targets. Thus, the bottleneck in processing arises when the targets must be identified and not when they can simply be detected on the basis of target features. (b) Performance on T2 does not reach the same high asymptote as it does in Experiment 5. Detections level off at around 80% instead of more than 90%, and they are accompanied by much higher false-positive rates. Both findings fit well with the hypothesis that detection can be based on the presence of disjunctive sets of features, detected in parallel and without binding. This explains both the near absence of the attentional blink and the lowered accuracy on T2. Detection rates on T1 were high, as they were in Experiment 5, but it is possible that they included some mistaken responses to T2 instead. If this happened, the T1 omissions would instead show up as omissions on T2.

General Discussion

Although preattentive processing allows many features of the visual scene to be processed in parallel, it is generally assumed that focused attention is required for the formation of high-level visual representations. The goal in our research was to explore the information that is extracted from an image during the first pass of visual processing and to see how far the high-level interpretation can go before attention is invoked. We used two different paradigms to explore attention limits at two possible stages of processing. In Experiments 1–3, we minimized the memory load by asking for immediate detection and identification responses to a single target item in an RSVP sequence. In Experiments 4–7, we explored limits to postdetection processing with a two-target attentional blink paradigm.

Single-Target Experiments (1–3)

The results in the first three experiments suggest that early aspects of natural scene categorization may reflect the parallel detection of disjunctive sets of features rather than the binding and individuation of high-level objects. Rather than claiming we have proved this hypothesis, we offer it as an alternative to accepting the detection data as proof of a late selection model, and we present data that make it a plausible alternative.

Although we did not use a dual-task paradigm in these first three experiments, the presentation rate was very high, and within each natural scene, multiple objects were presented. So detection rates of 80% (when no people were present) could be taken to suggest full attention-free processing, consistent with claims by Rousselet et al. (2002) and by Li et al. (2002). To counter this argument, we wish to make the point that high detection rates need not imply full identification and binding. When probed for what they had seen, participants identified fewer than half the detected targets and often failed to localize them correctly. The alternative account we propose—that detection can be mediated by sets of disjunctive features activating whatever high-level nodes they are compatible with—therefore cannot be dismissed or ignored. Consistent with this view, we showed that scenes with people in them lowered animal detection considerably, as expected from the fact that people share many perceptual features with animals. Also consistent is our finding that inverting the pictures left the people decrement intact, although it reduced the detections and identifications and must have substantially reduced the semantic information available.

In Experiment 2 we excluded the possibility that the impairment with people distractors was due to semantic confusion caused by the fact that people are closely related conceptually to nonhuman animals. When the stimuli were words and not images, there was no significant interference in detecting animal targets among distractor words denoting people. Thus, the decrement with animal targets was most likely due to the visual feature overlap with humans, consistent with the idea that detection depends on sets of disjunctive features. People share many features with other animals, forcing a more detailed perceptual analysis.

The Nature of Perceptual Features

So far we have assumed the existence of feature detectors characterizing natural categories. How plausible is this assumption? Feature detectors can be functionally defined as units that explicitly discriminate particular properties relatively independently of others and that do it in parallel across an area of space.

The features detected in perceptual processing are often assumed to be elementary visual properties like colors, orientations, directions, or speeds of movement. However, the definition of these elementary units of visual processing in any particular situation is really an empirical question that should be tested with a set of converging operations (Treisman & Gormican, 1988). Evidence can be drawn from behavioral experiments, computer simulations, and neural recordings either from single units or from brain imaging. Behavioral tests that we have used include showing that the candidate features can mediate both parallel search and texture segregation (implying that the features are replicated across a spatial area); examining whether they migrate independently in illusory conjunctions when attention is diverted; and testing whether they can be separately attended (implying in both cases that they are independently coded). Levin et al. (2001) used correlational analyses to show that features such as rectilinearity or part typicality can facilitate search for a target animal among distractor artifacts.

Feasibility tests of a feature detection account include the development of neural network models that are capable of using specified feature sets to classify complex scenes. Campbell, Thomas, and Troscianko (1997) created a multilayer perceptron that learned to correctly classify 91% of the image area into 11 different categories including various objects in the scene and their components, using features such as texture, size, shape, and position. Ullman et al. (2002) used a training set of 138 frontal faces and 40 side views of cars to test a recognition scheme based on parts or fragments rather than properties. Their computer program extracted optimal building blocks for successful classification that turned out to be of intermediate size: 11%–16% of the original image.

Finally, neural evidence from single-unit recordings can test whether candidate features are in fact detected by specialized populations of neurons. In categorizing natural scenes, the relevant features are likely to be relatively complex. Humans and other animals have evolved to detect the natural components of real world objects. Gross, Rocha-Miranda, and Bender (1972) showed that single units in the infero-temporal area respond to monkey hands or faces, and Tanaka (1996) found many units in the infero-temporal area that respond to what look like elementary components of natural objects. There is also considerable plasticity. The visual system may develop new sets of feature detectors for whatever classifications prove useful. There are indications that training can influence visual processing for particular stimuli, even at relatively early stages in the visual system (Ahissar & Hochstein, 1997; Karni & Bertini, 1997; Tovee, Rolls, & Ramachandran, 1996). At higher levels involving object categorization, Vogels (1999a, 1999b) recorded from monkey infero-temporal cells during a task of tree classification, ruling out low-level properties such as color, texture, or size. Sigala and Logothetis (2002) recorded from monkey infero-temporal neurons responding selectively to object features that constituted the relevant dimensions for categorizing particular object sets (dorsal fins and tail shape for fish, eye height and eye separation for schematic faces). Recording at still higher levels, from the monkey frontal cortex, Freedman, Riesenhuber, Poggio, and Miller (2002) changed the reward contingencies from distinguishing cats from dogs to distinguishing creatures that varied on orthogonal dimensions—dog–catness versus cat–dogness. The same units learned to respond selectively to one or the other of these new categories.

There seems then to be substantial evidence that feature detectors are involved in classifying natural objects and that new sets can be formed for newly relevant classifications. It is highly unlikely that one is born with detectors for wheels or engines to mediate the detection of vehicles in our experiments, but these components have become highly relevant in the modern mechanized world. In our attentional blink experiments (unlike the go/no-go detection of Thorpe et al., 2001), the attentional blink deficit increased in depth when the targets were from the more artificial means-of-transport category rather than from the animal category. It is likely that feature sets can be learned and automatized to different degrees.

One key question is whether these single units would still respond in the absence of attention. So far as we know, there have been few attempts to test this with cells that are clearly responding to features or parts of objects rather than to identified wholes. Rolls, Aggelopoulos, and Zheng (2003) recorded responses to whole objects and showed that the neural response drops off quite sharply as the relevant object is moved away from fixation, but this could be due to the fall-off in acuity. Sheinberg and Logothetis (2001) also recorded from cells in monkey temporal lobes during a search for familiar objects embedded in complex scenes and concluded that these cells responded only once the target had been detected. As Sheinberg and Logothetis said, their activity seems "tightly coupled to the process of transforming perceived wholes into learned reactions" (p. 1349) rather than to the perceptual processing leading up to detection. In our framework, these target detection cells would appear to reflect the output of the binding process rather than any input to it.

Another important question arises: "If a beak can be an elementary feature, why not a whole bird?" (A. Hollingworth, personal communication, March 17, 2005). Our answer is threefold: (a) There is a substantial increase in complexity between a typical beak and a typical bird. Although beaks vary in shape, the majority of them fall within a restricted range of parameters, and our hypothesis would be that the atypical ones might be missed. Wings change dramatically between the folded and the spread states; perhaps one detects them only when spread. The visual system probably uses a large set of disjunctive features, so that it does not rely on any particular fixed set of features to detect a target animal.

(b) Because there are fewer features than objects and many are shared within categories, the visual system can potentially achieve considerable economy. Just as writing uses a set of 26 letters to produce all possible words, the visual system can extract similarities and shared parts across different sets of objects rather than facing the combinatorial explosion involved in representing all possible conjunctions of parts and properties.

(c) The neural evidence supports the existence of cells responding to components or features of intermediate complexity, but there is little evidence at levels up to the infero-temporal area for whole object detectors other than detectors for the biologically special faces. Where these are found, for example in the prefrontal cortex, they may be coding the output of the binding process rather than the input to it. Again, the key question is whether they would respond when attention is directed elsewhere.

Attentional Blink Experiments (4–7)

In the second set of experiments (4–7), we explored attention limits by looking at the effects of detecting or of identifying a first target on the processing of a second. Although in Experiments 4 and 5, T1 was again easily identified, T2 was missed on a large proportion of the trials on which it closely followed T1. Thus, at least some stage of processing T1 must have demanded attention. We suggest that simply detecting T1 is based on automatic feature registration but that to identify the animal or vehicle, some further analysis is required, perhaps consisting of feature binding and the creation of an object file to represent the integrated object. The identification of T1 interferes with perception and/or memory for T2 for a period of up to 1 s.

The attentional blink increased both in depth and in duration when T2 was drawn from a different category from T1. There are a number of possible explanations for this two-category deficit: one is that the set of diagnostic features is much larger for two categories than for one and that maintaining this dual set takes more resources. We can probably rule out this account because it would predict the same deficit on T1 when its category was unknown relative to when it was known. This deficit was not in fact observed: T1 was detected equally well when its category was unknown (when vehicle–animal trials and animal–vehicle trials were randomly mixed) and when its category was known (when vehicle–vehicle and animal–animal trials were presented in separate blocks). Thus, simply maintaining the features of both categories in readiness was not more demanding than maintaining only one set.

Another possible account of the different category deficit on T2 is that category switching, analogous to task switching, demands additional resources to suppress the first set of features and retrieve the second (Allport, Styles, & Hsieh, 1994; Potter, Chun, Banks, & Muckenhoupt, 1998). Potter et al. (1998) showed that both a visual dual-task decrement and category-switching between digits and letters could contribute independently to the attentional blink. Yet another explanation is that when the categories are the same, the features of T1 might prime T2, facilitating its detection. We tested the voluntary task-switching account in Experiment 5 by removing the prior knowledge of the T2 category, presenting the target types randomly mixed within blocks instead of segregated. This uncertainty substantially increased the size of the attentional blink at short intervals for both same and different category targets and reduced the benefit of repeating the same category for T2, relative to Experiment 4. Thus, when participants knew that no switch would be needed, they benefited considerably from maintaining the same set of features ready for T2, and when a switch was required, the attentional blink increased. The remaining effect of congruence even when the category was unknown may reflect a passive priming component that benefits performance when the categories match, even without active preparation. When the category was known in advance (in Experiment 4), both active maintenance and passive priming could play a role, producing a somewhat larger same category benefit at intermediate lags (although this three-way interaction did not quite reach significance, $p = .078$).

In Experiments 6 and 7 we eliminated the requirement to identify the T1 and asked participants simply to detect its presence. When T2 still had to be identified, this substantially shortened the

attentional blink, although it did not eliminate it. It may be difficult to refrain completely from checking or from identifying T1 when its presence is relevant to the task. However, when both T1 and T2 were simply detected, the blink was almost eliminated. This helps to pinpoint the bottleneck in the attentional blink at the stage that differs between the identification and the detection task. Our account is that detection can be based on one or more target features, whereas identification usually requires binding. Although feature detection can occur in parallel and without demands on attention, binding cannot. It seems that, in contrast to Experiment 6, participants in Experiment 7 were for the most part able to avoid the identification check after detecting T1, perhaps because T2 was to be detected as well as T1, reducing any confusion about which target required which kind of processing. The fact that accuracy was quite low, with 18% false positives and only around 80% correct detection of T2, is also consistent with our account. Performance based on the detection of features with no binding and identification check would likely be less accurate than identification achieved through focused attention. The detection rates are similar to those with shorter presentation times in Experiment 1, but false-positive rates are much higher because in Experiment 1, participants were required to identify the target as well and therefore had to bind the detected features and could rule out false-positive responses to unbound or inconsistent features.

A Theoretical Framework

Some theorists see the blink as affecting primarily the late stages of perceptual processing, such as binding, consolidation, and conscious access; others see the blink as having no effect on perceptual processing and acting only to prevent storage in visual working memory. The data from Experiments 4 to 6 do not rule out the second account. The finding that the blink disappears when there is no need to name either target is consistent with an account in terms of access to visual working memory as well as with our account in terms of binding, because the identities must be stored for subsequent report. However, the limits to performance in Experiments 1 to 3 are unlikely to depend on working memory, since the responses were made immediately on detection.

The framework we propose offers a unified account for both sets of data. It elaborates on the account generally accepted for the attentional blink, as recently summarized by Marois et al. (2004). Perceptual processing is seen as comprising a first stage of perceptual analysis that is "largely unconscious and allows for the rapid, global, and highly efficient categorization of items and events in a visual scene," whereas a "second 'attentional' stage is necessary for the thorough identification, consolidation, and conscious report of visual events" (Marois et al., 2004, p. 465). Our goal was to specify in more detail what these two stages consist of and, in particular, to qualify the claim for efficiency in the first stage.

We assume that perception results from a recursive process with different degrees of elaboration occurring over different time intervals and in different tasks. An initial rapid pass through the perceptual hierarchy registers many features in parallel, both local and global. These features activate all the nodes with which they are compatible in a recognition network as well as possibly inhibit nodes with which they are incompatible. Thus, high-level semantic information about both the objects and their context may become

rapidly available before any detailed localization and binding of features (Hochstein & Ahissar, 2002). Because the access is nonselective, illusory conjunctions are primed as well as correct conjunctions. In addition, a global representation of the scene with the layout of global boundaries and potential objects becomes available, and semantic constraints between contextual features and target features may help reduce the uncertainty even before binding is achieved. Polar bears are found with icebergs, and deer with woods or fields.

A reentry process with focused attention then serially selects the local objects of potential interest and binds their features to form object files that can be consciously experienced (Kahneman et al., 1992). The presence of a particular category of objects may be detected at the initial stage of coding through the priming of nodes for objects in that category. The presence of the detected target is then normally checked by focusing attention onto the location of one of its features, allowing the other features in that location to be bound and the object consciously identified by matching to the active nodes in the recognition network. If the original stimulus is no longer available, the category may still be inferred from the features that were detected.

Evidence consistent with the proposed framework comes from a variety of sources. Perhaps most telling, Levin et al. (2001) showed that search for animals among artifacts or artifacts among animals was not much impaired when the parts of the relevant images were jumbled, suggesting that detection could be triggered by features without regard to the overall structural relations. Inverting the faces of animals or humans does not significantly impair performance on a rapid detection task, whereas full identification is substantially impaired by inversion (Rousselet et al., 2003). Thus, a fast visual categorization of complex stimuli appears to precede the formation of complex high-level representations. Further evidence consistent with an initial feature registration followed by a binding process registering the configuration comes from Liu, Harris, and Kanwisher (2002), who used magnetoencephalography to investigate stages of processing of faces. They distinguished two evoked components, the M100, which is more sensitive to face parts, responding to a scrambled face with the local features preserved, and a later M170, responding to a face configuration in which the local features were replaced by standard ovals, preserving only the relative locations of the features. Hopf, Vogel, Woodman, Heinze, and Luck (2002) also found evidence from electromagnetic studies in humans for two stages of object identification, relying on two different attention allocations: fast object detection resulting from allocation of attention to task-relevant features that then directs spatial attention to the object's location, allowing its positive identification. Raymond (2003) found that an attentional blink was triggered only when a new object representation had to be formed and not when an old object was updated with a new feature. Finally, in another attentional blink experiment, Marois et al. (2004) using fMRI were able to separate the brain areas involved in a first, rapid, attention-free detection of a natural scene (i.e., parahippocampal regions of the medial temporal lobe) from the areas activated in a second, attention-demanding, conscious response to the target (i.e., the frontal cortex).

Rousselet, Thorpe, and Fabre-Thorpe (2004) in a recent review grappled with the same problem as we do in this article, namely that "the ventral pathway seems to be able to perform complex analyses on several objects simultaneously, but only during a short time period. Subsequently only one or a very few objects are explicitly selected and consciously perceived" (p. 363). Their solution is different, however, because they assume that these perceptual objects are simultaneously bound and identified, whereas we explore the possibility that the data could reflect the parallel registration of sets of unbound features, with the bottleneck arising in the attempt to bind them. They argued that spatial binding could be achieved rapidly and in parallel in the ventral visual pathway and propose a number of neural mechanisms in areas up to and including the infero-temporal cortex that could specify conjunctions of parts of shapes or objects. In their view, a bottleneck arises after the initial identification of the objects because of competition or lateral inhibition, although it is not entirely clear why competition should arise only at this late stage. We agree that many relational or emergent properties may be directly coded; in fact, they would be needed to specify the kind of high-level features we propose. But the behavioral evidence that illusory conjunctions do frequently occur when attention is diverted cannot be ignored (e.g., Treisman & Paterson, 1984; Treisman & Schmidt, 1982). There does in fact seem to be a binding problem. Our article sketches one way of achieving the rapid but crude semantic assessments of complex visual scenes despite the limitations imposed by this binding bottleneck. Once detected, if the target object is still present, one can draw on attentional resources to form an object file to make it consciously accessible (Kahneman et al., 1992). If not, one may be able to tell simply from the features which animal, vehicle, or other object of interest it is likely to be, but one will also be vulnerable to errors.

References

Ahissar, M., & Hochstein, S. (1997, May 22). Task difficulty and the specificity of perceptual learning. *Nature, 387,* 401–406.

Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umiltà and M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 421–452). Cambridge, MA: MIT Press.

Biederman, I. (1972, July 7). Perceiving real-world scenes. *Science, 177,* 77–80.

Braun, J. (2003). Natural scenes upset the visual applecart. *Trends in Cognitive Sciences, 7,* 7–9.

Broadbent, D. E., & Broadbent, M. H. P. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. *Perception & Psychophysics, 42,* 105–113.

Campbell, N. W., Thomas, B. T., & Troscianko T. (1997). Automatic segmentation and classification of outdoor images using neural networks. *International Journal of Neural Systems, 8,* 137–144.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 109–127.

Corel. (1996). Corel stock photo library [CD-ROM]. Ottawa, Ontario, Canada: Author.

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorization of natural scenes does not rely on colour cues: A study in monkeys and humans. *Vision Research, 40,* 2187–2200.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General, 129,* 481–507.

Enns, J., & Rensink, R. A. (1991). Preattentive recovery of three-

dimensional orientation from line drawings. *Psychological Review, 98,* 335–351.

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience, 13,* 171–180.

Fabre-Thorpe, M., Richard, G., & Thorpe S J. (1998). Rapid categorization of natural images by rhesus monkeys. *NeuroReport, 9,* 303–308.

Freedman, D., Riesenhuber, M., Poggio, T., & Miller, E. (2002). Visual categorization and the primate prefrontal cortex: Neurophysiology and behavior. *Journal of Neurophysiology, 88,* 929–941.

Giesbrecht, B. L., & Di Lollo, V. (1998). Beyond the attentional blink: Visual masking by object substitution. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1454–1466.

Gross, C., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology, 35,* 96–111.

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

Hopf, J., Vogel, E., Woodman, G., Heinze, H., & Luck, S. (2002). Localizing visual discrimination processes in time and space. *Journal of Neurophysiology, 88,* 2088–2095.

Joseph, J. S., Chun, M. M., & Nakayama, K. (1997, June 19). Attentional requirements in a "preattentive" feature search task. *Nature, 387,* 805–807.

Kahneman, D., Treisman, A., & Gibbs, B. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 175–219.

Karni, A., & Bertini, G. (1997). Learning perceptual skills: Behavioral probes into adult cortical plasticity. *Current Opinion in Neurobiology, 7,* 530–535.

Lavie, N. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics, 56,* 183–197.

Levin, D. T., Takarae, Y., Miner, A. G., & Keil, F. (2001). Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. *Perception & Psychophysics, 63,* 676–697.

Li, F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences, USA, 99,* 9596–9601.

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

Mack, A., & Rock, I. (1998). *Inattentional blindness: Perception without attention.* Cambridge, MA: MIT Press.

Marois, R., Yi, D. J., & Chun, M. (2004). The neural fate of consciously perceived and missed events in the attentional blink. *Neuron, 41,* 465–472.

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.

O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999, March 4). Change-blindness as a result of "mudsplashes." *Nature, 398,* 34.

Potter, M. C. (1975, March 14). Meaning in visual search. *Science, 187,* 965–966.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 509–521.

Potter, M. C., Chun, M. M., Banks, B. S., & Muckenhoupt, M. (1998). Two attentional deficits in serial target search: The visual attentional blink and an amodal task-switch deficit. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 974–992.

Ramachandran, V. S. (1988, January 14). Perception of shape from shading. *Nature, 331,* 163–166.

Raymond, J. E. (2003). New objects, not new features, trigger the attentional blink. *Psychological Science, 14,* 54–59.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary

suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18,* 849–860.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8,* 368–373.

Rolls, E. T., Aggelopoulos, N. C., & Zheng, F. (2003). The receptive fields of inferior temporal cortex neurons in natural scenes. *Journal of Neuroscience, 23,* 339–348.

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience, 5,* 629–630.

Rousselet, G. A., Mace, 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.

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? *Trends in Cognitive Sciences, 8,* 363–370.

Seiffert, A. E., & Di Lollo, V. (1997). Low-level masking in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1061–1073.

Sheinberg, D., & Logothetis, N. (2001). Noticing familiar objects in real world scenes: The role of temporal cortical neurons in natural vision. *Journal of Neuroscience, 21,* 1340–1350.

Sigala, N., & Logothetis, N. K. (2002, January 17). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature, 415,* 318–320.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1,* 261–267.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19,* 109–139.

Thorpe, S., Fize, D., & Marlot, C. (1996, June 6). Speed of processing in the human visual system. *Nature, 381,* 520–522.

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

Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & Bülthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *European Journal of Neuroscience, 14,* 869–876.

Tovee, M. J., Rolls, E. T., & Ramachandran, V. S. (1996). Rapid visual learning in neurons of the primate temporal visual cortex. *NeuroReport, 7,* 2757–2760.

Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 40*(A), 201–237.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology, 6,* 171–178.

Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London, Series B, 353,* 1295–1306.

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

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95,* 15–48.

Treisman, A., & Paterson, R. (1984). Emergent features, attention, and object perception. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 12–21.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14,* 107–141.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5,* 682–687.

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

VanRullen, R., & Thorpe, S. J. (2001b). The time course of visual pro-
    cessing: From early perception to decision-making. *Journal of Cognitive
    Neuroscience, 13,* 454–461.

Vogel, E. K., & Luck, S. J. (2002). Delayed working memory consolida-
    tion during the attentional blink. *Psychonomic Bulletin & Review, 9,*
    739–743.

Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological
    evidence for a postperceptual locus of suppression during the attentional
    blink. *Journal of Experimental Psychology: Human Perception and
    Performance, 24,* 1656–1674.

Vogels, R. (1999a). Categorization of complex visual images by rhesus
    monkeys. Part 1: Behavioural study. *European Journal of Neuroscience,
    11,* 1223–1238.

Vogels, R. (1999b). Categorization of complex visual images by rhesus
    monkeys: Part 2. Single cell study. *European Journal of Neuroscience,
    11,* 1239–1255.

Wilson, M. D. (1988). The MRC psycholinguistic database: Machine
    readable dictionary, Version 2. *Behavioural Research Methods, Instru-
    ments and Computers, 20,* 6–11.

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless
    bundles of basic features. *Vision Research, 37,* 25–44.