Effects of Task Requirements on Rapid Natural Scene Processing: From Common Sensory Encoding to Distinct Decisional Mechanisms

Nadège Bacon-Macé, Holle Kirchner, Michèle Fabre-Thorpe, and Simon J. Thorpe
CerCo, Université Toulouse 3, CNRS

Using manual responses, human participants are remarkably fast and accurate at deciding if a natural scene contains an animal, but recent data show that they are even faster to indicate with saccadic eye movements which of 2 scenes contains an animal. How could it be that 2 images can apparently be processed faster than a single image? To better understand the origin of this speed advantage in forced-choice categorization, the present study used a masking procedure to compare 4 tasks in which sensory, decisional, and motor aspects were systematically varied. With stimulus onset asynchronies (SOAs) above 40 ms, there were substantial differences in sensitivity between tasks, as determined by d' measurements, with an advantage for tasks using a single image. However, with SOAs below 30–40 ms, sensitivity was similar for all experiments, despite very large differences in reaction time. This suggests that the initial part of the sensory encoding relies on common and parallel processing across a large range of tasks, whether participants have to categorize the image or locate a target in 1 of 2 scenes.

Keywords: masking, decisional processes, manual and saccadic eye movement responses, natural scenes

A large number of experiments have now been performed on rapid scene categorization, and all have demonstrated that humans and monkeys are very efficient at detecting animals or other target categories in natural scenes (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Rousselet, Macé, & Fabre-Thorpe, 2003; Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001a). The general paradigm is based on a go/no-go task that involves the very brief presentation of a single image, to which participants have to respond manually by releasing a button if they detect a target (Thorpe et al., 1996). Human participants generally score over 90% correct, with mean reaction times ranging from 400 to 430 ms, and hits start to significantly outnumber false alarms at around 250–280 ms. More recently, a choice saccade task has been introduced in which two photographs are flashed simultaneously in both hemifields, and participants have to make a saccade toward the photograph that contains an animal (Kirchner & Thorpe, 2006). Whereas accuracy remains high (around 90% correct), saccadic reaction times are much shorter, with a mean value of 228 ms, and correct responses outnumber errors at latencies as short as 120 ms. Furthermore, an analysis of the low-level differences between target and distractor images has shown that they were unable to account for these very fast responses (Kirchner & Thorpe, 2006). The comparison of the above paradigms raises a number of issues, such as the nature of the pathways involved from perception to action and the temporal dynamics of visual processing. How can we explain the fact that with similar accuracy in target detection, reaction times can be so different from one task to another? In particular, how can we explain that shorter reaction times are obtained in a task in which the participant needs to process two simultaneous targets instead of just one? Moreover, although the choice saccade task seems to be a promising paradigm for the study of visual perception, we still need to understand where and how the localization process takes place in the whole stream of object recognition.

The comparison of reaction times in go/no-go and forced-choice tasks relates to the extensive literature in experimental psychology concerning mental chronometry and its theoretical analysis (Meyer, Osman, Irwin, & Yantis, 1988). According to these theories, performance depends on a sequence of processes that includes sensory encoding of the stimuli, storage of information in long-term memory, decision making, and programming of an appropriate motor response. Currently, a growing literature on electrophysiology and brain imaging focuses on the neural processes that take place between perception and action, corresponding to the mechanisms of neural selection and decision (Heekeren, Marrett, Bandettini, & Ungerleider, 2004; Schall, 2001, 2004; Seidemann, Zohary, & Newsome, 1998; VanRullen & Thorpe, 2001b; Vorberg, Mattler, Heinecke, Schmidt, & Schwarzbach, 2003). A consequence is that the perception–decision–action stream is no longer considered a simple sequential system but, rather, a continuous flow in which information accumulates over time until a decisional threshold is reached that will trigger the motor output (Gold & Shadlen, 2000; Hanes & Schall, 1996; Perrett, Oram, & Ashbridge, 1998; Romo & Salinas, 2003;

Shadlen & Newsome, 1996). At the functional and anatomical level, current models are now closer to a *distributed* or *parallel and asynchronous* system (Humphreys, Riddoch, & Price, 1997; McClelland, 1979; McClelland & Rogers, 2003; Rumelhart, Hinton, & McClelland, 1986). This suggests that the different requirements in the two tasks described above may have independent effects on performance. These differences can be listed under three points.

First, at the beginning of the processing sequence, the two tasks differ in terms of the physical data that enter the visual system: Participants are asked to process either one central image or two peripheral images. Concerning the eccentricity, it has been shown that peripheral presentations induce a drop in performance when participants are categorizing a single image at increasing eccentricities, but the effects are relatively small in the near periphery. Performance drops by 3.5% when the scene is presented at 3.6° from the center of the screen (Fize, Fabre-Thorpe, Richard, Doyon, & Thorpe, 2005). With respect to the number of images present, the processing of two simultaneously presented peripheral natural images has been shown to have surprisingly little cost in terms of accuracy or speed compared with the processing of a single target at the same location (Rousselet, Fabre-Thorpe, & Thorpe, 2002). Rousselet et al. (2002) argued that the slight drop in performance can be explained by a parallel model of processing in which each of the two images is processed by separate channels that pool their outputs. Their study specifically tested performance while participants had to decide in a go/no-go mode whether or not there was a specific target present in one or two images. The results cannot be directly compared with those of the choice saccade task because, in the latter, an animal is always present, either on the left or on the right of the center of the screen.

Second, at the end of the processing sequence, the anatomical pathways involved in motor preparation are clearly dissociated in both tasks: Saccadic eye movements will presumably rely on a network involving structures like the frontal eye fields, the lateral intraparietal area, and the superior colliculus, all of which have close connections with the oculomotor plant in the brain stem. However, manual responses will need to take a longer route, involving motor and premotor cortical areas, to finally reach the hand effectors through the spinal cord. It should be noted that there is evidence that these pathways are used almost optimally in the go/no-go task because training on specific images failed to affect the earliest reliable responses in this categorization task (Fabre-Thorpe et al., 2001). Similarly, in the choice saccade task, although saccadic reaction times decreased during the course of the study, this improvement of general motor skill did not interact with the type of image present (novel vs. repeated), again indicating that information flow from perception to action was close to optimal (Kirchner & Thorpe, 2006). Thus, at least part of the long reaction times seen in the categorization task could result from the longer neural path length required for manual responses. However, we cannot exclude the possibility that other differences at the level of visual analysis or decisional processes may also be involved.

In fact, there are also obvious differences in the very nature of the questions asked by the two basic tasks. In the standard go/no-go task, participants have to determine whether the image contains an animal (i.e., the question is "is there an animal?"), whereas in the choice saccade task, participants know in advance that an animal will be present in one of two images but have to indicate in which of the two scenes the animal appears (i.e., "is the animal on the left or on the right?"). To distinguish these different requirements, we refer to *categorization tasks* in the first case and to *discrimination tasks* in the second case. To what extent could this difference between the tasks explain the speed advantage of the saccadic discrimination task? There is evidence in the literature that different cortical streams are involved when the visual task requires object recognition (ventral "what" stream) and when it involves controlling goal-directed actions (dorsal "where" stream; Goodale & Milner, 1992; Ungerleider, 1985; Ungerleider & Haxby, 1994). Neurons in the dorsal stream are known to have short latency responses, considerably shorter than those in ventral stream areas such as the inferotemporal cortex. But the question "where is the animal?" is not trivial, for answering it might require binding of different features to a particular location (Evans & Treisman 2005). Such processing probably involves interactions between ventral and dorsal visual streams.

In a similar vein, the long tradition of focus on attentional deployment in visual search tasks indicates that some visual features, such as color or orientation, can be processed in parallel—that is, in a preattentive mode—whereas conjunctions of these features or slightly more complex stimuli, such as an *L* or a *T*, can only be recognized after attention has been focused on this particular location (Treisman & Gelade, 1980). Remarkably, natural scenes can be categorized even if focal attention is occupied elsewhere (Li, VanRullen, Koch, & Perona, 2002). At first glance, this would suggest that even complex scenes—as opposed to simple, single visual features—can be processed in parallel. However, when the same natural scenes are investigated in a visual search task, they require time-consuming, serial attention. Therefore, visual features and objects that can be processed preattentively in a dual-task paradigm will not necessarily yield parallel performance in visual search tasks. Natural scenes do not pop out in visual search tasks (VanRullen, Reddy & Koch, 2004), and the parallel processing seen with two images is already limited when four images are presented simultaneously (Rousselet, Thorpe, & Fabre-Thorpe, 2004b)

A fundamental issue concerns the nature of the information used to perform scene categorization tasks and whether this level of representation requires attentional mechanisms. Evans and Treisman (2005) proposed that while the initial processing may be sufficient for determining that an animal is present, detailed identification and localization may require additional attention-based processing. The possibility that relatively simple mechanisms may suffice for at least some forms of high-level visual tasks is supported by a number of recent studies. First, using a masking procedure, Grill-Spector and Kanwisher (2005) reported that as soon as participants could detect the presence of an object, they could also report its category. However, these authors also found that more precise object identification required additional processing time. Similarly, in other work from our group, Macé and colleagues (Macé, Bacon-Macé, Nespoulous, & Fabre-Thorpe, 2007; Fabre-Thorpe, Macé, & Joubert, 2005) found that reaction times in the go/no-go categorization task were in fact shorter for a superordinate category judgments (animal vs. nonanimal) than they were for more specific, basic-level category judgments (dog vs. nondog or bird vs. nonbird). A second argument comes from recent modeling work showing that scene categorization at levels similar to those of humans under masking conditions can be

achieved using relatively simple hierarchical classification systems, broadly compatible with a feedforward pass through the various levels of the ventral visual parthway (i.e., without the need for specific attentional mechanisms or binding; Torralba & Oliva, 2003). As a consequence, it is perfectly conceivable that there are limits to the sorts of processing that can be achieved with a first (preattentive) pass alone.

One way to address this question is to use a masking procedure to interrupt scene processing after given delays. In a previous study, we showed that human participants' categorization judgments are already highly reliable when a very contrasted (dynamic) mask interrupts scene processing after only 30 ms (Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe, 2005). Such data imply that processing at each stage in the visual system is remarkably rapid, with information accumulation following a first, feedforward sweep through the system (Bacon-Macé et al., 2005; VanRullen & Koch, 2003). However, the very short reaction times observed in the choice saccade task used by Kirchner and Thorpe (2006) were obtained with unmasked stimuli. In the present study, we again used a masking procedure with varied stimulus onset asynchronies (SOAs) between the presentation of the scene and the presentation of the dynamic mask to analyze the time course of visual processing in four different tasks. The tasks systematically differed both in terms of the type of motor output (oculomotor vs. manual) and the type of decision (presence or absence of an animal vs. choice of the scene that contains an animal).

Our previously published results using the masking procedure and the manual go/no-go task (Bacon-Macé et al., 2005) are reported here as Experiment 1 and compared with a manual yes–no task (Experiment 2), a saccadic forced-choice task (Experiment 3), and a manual forced-choice task (Experiment 4). Given that Experiments 1 and 2 were based on a single, central presentation of a scene, whereas Experiments 3 and 4 required simultaneous presentations of two images in the near periphery, the masking procedure used in the present study allowed us to investigate the chronometry of parallel processing of natural scenes. The present study tested whether parallel processing persists during the whole categorization process or is limited by attentional, decisional, or motor factors due to different task requirements.

Method

The general procedure for the four experiments is shown in Figure 1. In Experiments 1 and 2, participants had to perform a task in which one central image was briefly flashed on the screen and followed by a masking stimulus. Because the participant had to answer the question "is there an animal in the image?" we refer to these tasks as categorization tasks. In Experiment 3 and 4, two images were presented, one in each hemifield, and followed by a mask. In all trials, one of the two images contained the target and the question was, thus, "Which of the two scenes contains an animal?" These experiments are referred to as discrimination tasks.

Participants

A total of 31 participants with normal or corrected-to-normal vision took part in the experiments (15 female, 16 male; mean

age = 27.5 years). They all volunteered for the study and gave their written informed consent. Several participants took part in more than one experiment, but in each task, at least one third of the participants were totally naive with regard to the experimental procedure.

Stimuli

Natural scenes. In the four tasks, we used a large variety of achromatic natural scenes from the Corel database. Target photographs could contain mammals, birds, reptiles, or fish, and distractors were also varied, including landscapes, indoor or outdoor scenes, and man-made objects. Each image was seen only once in a given task. Horizontal images had been used in Experiment 1, but to allow simultaneous presentation of two photographs, we used a fixed set of vertical images in Experiments 2–4.

Mask. The masks were the same as those used in a previous study (Bacon-Macé et al., 2005). They were produced by filtering a white-noise pattern at four different spatial scales and increasing the contrast of the resulting images to 100%. Each mask was rotated and mirrored so that, in total, a pool of 16 masks was available. In the first task, the mask was composed of a sequence of 8 masks, with each spatial scale appearing twice in the sequence. However, because the masking effects were shown to be as strong with 4 masks as with 8 (Bacon-Macé et al., 2005), we used only 4 masks in the three other tasks, with each spatial frequency displayed once. Each mask was flashed for two frames, resulting in total mask durations of around 100 ms (Experiment 1) and 50 ms (Experiments 2–4).

Experimental Design

Participants were seated in a dark room, in front of a screen adjusted to an 800 × 600 pixel resolution with a 160-Hz refresh rate. Photographs of natural scenes were flashed for a single frame—that is, 6.25 ms—and were followed by a mask with a variable SOA. In each series, there were as many targets as distractors, and all SOA conditions (6.25, 12.50, 18.75, 31.25, 43.75, 81.25, and 106.25 ms) were randomly intermixed but in the same proportion. Furthermore, in the four tasks, we added one special condition in which only the mask was presented, without any scene being flashed before. This condition was unknown to the participants and occurred on roughly 1 out of 10 trials in each series.

Experiment 1: Categorization tasks with go/no-go manual responses. This protocol was similar to that used in a large number of earlier studies. A trial began with the display of a fixation cross in the center of a black screen for 600–900 ms at random. Then a target or distractor image (15.3° high × 10.4° wide, 800 × 600 pixels) was presented, followed by the mask. Participants had to release a button as quickly as possible if the image contained an animal and maintain contact with the button otherwise. Moreover, they were asked to try to release the button on 50% of the trials, whatever the masking condition. Each of the 16 participants (8 female, 8 male; mean age = 29.4 years) performed 16 series of 90 trials and were previously trained with two series. Unlike the three other experiments in this study, Experiment 1 contained an extra SOA condition of 25 ms. For this experiment, we used the data that had been collected previously. The detailed procedure and full data

CATEGORIZATION TASK: Is there an animal?
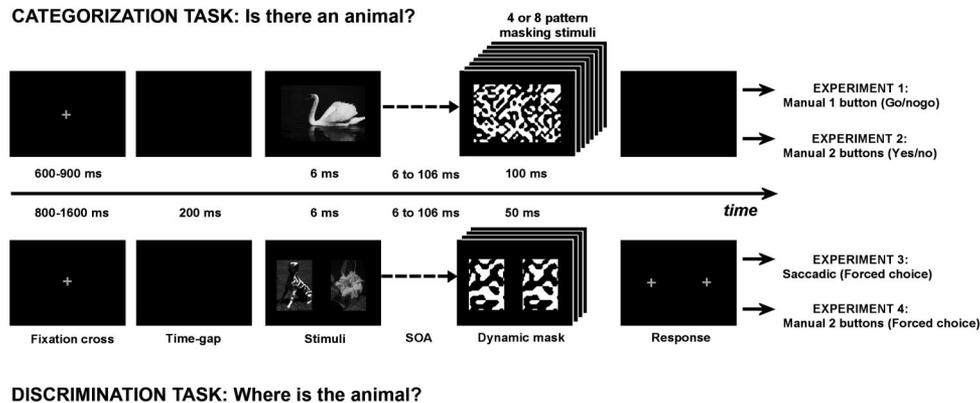
Figure 1. Experimental design in the four experiments. Each trial began with the display of a fixation cross for a variable delay. In three experiments (all but Experiment 1), we introduced a gap interval of 200 ms. In the categorization tasks, one central image was flashed and followed by a highly contrasted pattern mask after a variable stimulus onset asynchrony (SOA). In the discrimination tasks, two lateral images were presented, one of which was the target. Depending on the experiment, participants had to respond manually with one or two buttons or with a saccadic eye movement in the direction of the target. In all tasks, participants were told to be as accurate and fast as possible.

on this task have already been published (Bacon-Macé et al., 2005) and constitute a baseline for the three other, original experiments reported here and detailed below.

Experiment 2: Categorization task with yes–no manual responses. This task was comparable to the first experiment in that participants had to decide whether there was an animal in the centrally presented masked image. Unlike in the first experiment, participants had to provide a response on each trial: They maintained their index fingers on the two *Ctrl* buttons of a keyboard throughout the series and released the button on the side of their dominant hands to indicate that they had detected the presence of an animal in the scene or released the button on the opposite side if they thought the scene was a distractor. Motor responses were thus very comparable to those in the fourth experiment (see below). The set of vertical photographs (7.7° high × 11.5° wide, 305 × 445 pixels) used in this experiment was also used in Experiments 3 and 4. Twelve participants (5 female, 7 male; mean age = 26.5 years) performed 12 series of 80 trials after one block of training trials.

Experiment 3: Forced-choice discrimination with saccadic responses. A trial began with the display of the fixation cross at the middle of the screen for a random interval from 800 to 1,600 ms. To facilitate saccade initiation, we had the fixation cross disappear 200 ms before the presentation of the images (Fischer & Weber, 1993; Kirchner & Thorpe, 2006). Two vertical natural scenes were flashed in the left and right hemifields, with the center of the image at 6°. They were followed by two lateral masks, covering the surface of the images. Masks on both sides were identical. Each participant, whose head was stabilized by a forehead and chin rest, had the instruction to make a saccade toward the side where an animal had appeared. Possible saccade endpoints were indicated by two fixation points that were displayed for 1 s after the masks at the center of the images. Eight participants (4 female, 4 male; mean age = 27.8 years) performed 12 series of 64 trials. For

training, they performed two series without masking and a third series with masking.

Experiment 4: Forced-choice discrimination with manual responses. The experimental design was similar to the third task, with the same set of images but a different response mode. Throughout the series, participants had to maintain each of their index fingers on the two *Ctrl* buttons of a keyboard (Hasbroucq, Mouret, Seal, & Akamatsu, 1995). They had to indicate the side on which an animal had appeared by releasing the *Ctrl* button on the appropriate side. Participants were told to be as quick and accurate as possible. Eight participants (3 female, 5 male; mean age = 26.0 years) performed 12 series of 64 trials, after having performed 1 training series.

Globally, the number of trials in the different experiments was made equivalent for statistical analysis, except for the first experiment, which presented a higher number of trials because of the associated electroencephalograph (EEG) recordings (Bacon-Macé et al., 2005). The additional training series in Experiment 3 were performed to habituate nonexpert participants to making saccadic eye-movement responses.

Response Recording

In the first task, manual responses were detected using a button with an infrared diode system, and this signal was directly recorded by the stimulation software. In the third task, eye position was recorded by two horizontal electroculogram (EOG) electrodes (Neuroscan NuAmps; Compumedics, Hamburg, Germany) with a sampling rate of 1 kHz, a notch at 50 Hz, and a low-pass filter at 90 Hz. Before analysis of the signal, a baseline correction was applied on a range starting from 400 ms preceding the stimulus onset to the beginning of image presentation. A saccadic response was validated for a trial if the subtracted signal between left and right EOG electrodes reached an amplitude of at least ±35 μV. The nearest signal inflection preceding this point was then con-

sidered the onset of the response. In the second and fourth tasks, manual responses were detected with the two *Ctrl* buttons of a keyboard connected to the stimulation software (Presentation; Neurobehavioral Systems, Albany, CA).

Analysis

Effects on performance within and between tasks were assessed using analysis of variance (ANOVA) with a Greenhouse–Geisser correction for nonsphericity. If necessary, post hoc analyses were performed using *t* tests with a Bonferroni correction or Wilcoxon tests.

Results

Accuracy

In this section, the accuracy of the participants is analyzed for the four tasks as a function of SOA. We first determined the accuracy in terms of total percentage correct, but we also analyzed the results by calculating *d'*, which is known to be a better parameter for comparing tasks that involve both one- (go/no-go, yes–no) and two-alternative (forced-choice) detection paradigms (Green & Swets, 1966; Kroll, Yonelinas, Dobbins, & Frederick, 2002).

Percentage correct. In all tasks, there was one chance in two to give a correct answer on a given trial. Chance level is thus calculated at 50% and significance determined by a chi-square test.

ANOVA showed that accuracy in the discrimination tasks was higher than accuracy in the categorization tasks, $F(3, 307) = 9.28$, $p < .001$, and, interestingly, did not depend on the particular motor response to be produced. When all SOA conditions were pooled together, accuracy in the discrimination tasks averaged 77% with manual responses (individual range = 67.41%–79.61%) and 76.1% with saccadic responses (72.29%–82.78%). In contrast, in the categorization tasks, participants scored an average 72.8% in the go/no-go task (64.82%–79.20%), which was not significantly different from the accuracy in the yes–no task: 73.4% (66.63%–77.02%).

As expected, accuracy varied with the SOA conditions (see Figure 2A). In all experiments, the percentage correct began at chance level with a 6-ms SOA, which indicates that visual processing was totally blocked, but rapidly increased until reaching a plateau above a 44-ms SOA. Participants were already above chance level ($p < .001$) with a 12-ms SOA in three of the tasks, and only the manual yes–no experiment failed to reach statistical significance in this SOA condition. The maximum value was obtained with the two last SOA conditions (81 ms and 106 ms), in which accuracy came close to or exceeded 90% correct. This accuracy is comparable to the accuracy obtained in a previous go/no-go experiment using gray-level stimuli without masking (Delorme, Richard, & Fabre-Thorpe, 2000). Interestingly, the four tasks had a common global accuracy profile with increasing SOA, although post hoc analyses revealed some significant differences between Experiment 1 (categorization with a manual go/no-go task) and the three other experiments in the 19-ms SOA condition, $F(3, 43) = 12.58$, $p < .001$.

In the two discrimination experiments, targets were presented equally often on the left or right side of the display, and partici-pants, regardless of their preferred hand, did not show any significant response bias toward one or the other side (percentages of responses on the left and right side: 48.7% vs. 51.3% in the saccadic responses, 51.1% vs. 48.9% in the manual responses). Moreover, in the condition in which no target image was presented before the mask, the response rate for each hemifield was close to 50%, indicating that participants responded randomly regardless of the side (49.6% vs. 50.4% in the saccadic responses, 48.5% vs. 51.5% in the manual responses).

In the two categorization tasks, participants were told that there were about the same number of target and distractor trials in a series, and they were encouraged to respond on half the trials. Despite this instruction, the number of *go* or *yes* responses (hits) varied considerably with the difficulty of the task (manual go/no-go, $F(6, 111) = 239.62$, $p < .001$; manual yes–no $F(6, 83) = 30.59$, $p < .001$), as shown in Table 1. In both experiments, response rates were stable for the three longest SOA conditions (44, 81, and 106 ms) but strongly decreased in all of the other, more challenging SOA conditions ($p < .001$ [Wilcoxon tests and Bonferroni corrections]). Furthermore, in the more difficult SOA condition of 6 ms, response rates (either hits or false alarms) were comparable to the response percentages when no target image was presented before the mask (response percentage in the 6-ms SOA/no-target condition: 22.3% vs. 19.6% in the go/no-go experiment, 31.1% vs. 25.1% in the yes–no experiment).

Globally, in the categorization tasks, participants showed strong inhibition when they had to initiate a response with limited information about the stimulus. In fact, two factors can explain a decrease in performance: either an increase in the number of false alarms or a decrease in the number of hits. Although response latencies did not increase much (see below), the number of hit responses strongly decreased at short SOAs. In contrast, in the discrimination tasks, only an increase in the number of false alarms could produce a decrease in the percentage of correct responses. Thus, accuracy measures, although very instructive, cannot be considered on their own to analyze performance variation as a function of task difficulty. In the next section, we provide details of the *d'* analysis, which is known to be a bias-free measure of sensitivity (Green & Swets, 1966; Macmillan & Creelman, 1991).

d' analysis. In the one-alternative paradigm, the detection sensitivity measure *d'* is calculated as $z_H - z_{FA}$, where H is the hit rate and FA the false alarm rate (Macmillan & Creelman, 1991). In the two-alternative paradigm, the equation needs to take into account the summation of the variability for two distributions (Macmillan, 2002). As a result, the *d'* measure must consider a correction factor and is calculated as $(1/\sqrt{2})(z_H -- z_{FA})$. These values were calculated for the four experiments, separately for each SOA condition and for each individual participant. For some participants, rates were 1 (in particular, some participants scored 100% correct in the easiest conditions), which would lead to an infinite value for the *z*-score. In such cases, we used a rate of 0.99 instead.

Figure 2B shows the *d'* values as a function of SOA for all four experiments. Across all SOA conditions, *d'* values were significantly different between experiments, $F(18, 307) = 4.528$, $p < .001$, but post hoc analyses indicated that the paradigms that used the same number of alternatives (either categorization or discrimination) did not differ statistically. Thus, discrimination task *d'*s did not differ between the manual ($M = 1.2$ [0.68–1.50]) and
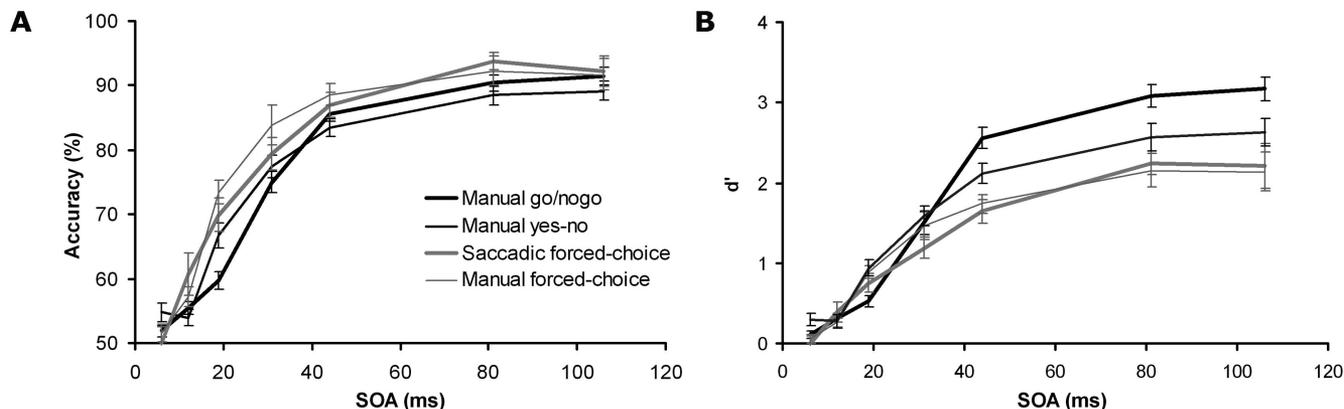
Figure 2. Mean performance as a function of stimulus onset asynchrony (SOA) in the four experiments. A: Mean accuracies, with chance level at 50%. In the categorization tasks, the percentages were calculated over all correct responses and, thus, reflect both hits to targets and correct rejections with distractors. B: Mean d's in the go/no-go and yes–no tasks and mean corrected d's in the forced-choice tasks. Error bars represent standard errors of the means.

saccadic response paradigms ($M = 1.2$ [0.97–1.68]). Similarly, in the categorization tasks, the yes–no d' did not differ from the d' calculated for the go/no-go task ($M = 1.5$ [0.96–1.76] vs. $M = 1.6$ [1.02–2.12]).

More interestingly, Figure 2B clearly shows two separate phases as d' increases with SOA. First, for SOAs up to and including 31 ms, the four curves are quite close, and globally, differences between experiments did not reach significance. The only exception was the 19-ms SOA condition, in which, as described above for accuracy, the manual go/no-go task had a slightly inferior d' value, $F(3, 43) = 12.582$, $p < .01$, compared with both the saccadic discrimination and the manual yes–no paradigms. However, with SOAs above 31 ms, d' values varied considerably between experiments (44 ms: $F(3, 43) = 10.867$, $p < .001$; 81 ms: $F(3, 43) = 7.78$, $p < .001$; 106 ms: $F(3, 43) = 6.857$, $p < .001$), and it is interesting to note that the highest d' values were seen in the manual go/no-go task. The d' in the yes–no task was intermediate, but both discrimination tasks had values that were substantially lower. The comparison between the two discrimination tasks and the go/no-go task reached significance from an SOA of 44 ms onwards ($p < .006$). Therefore, whereas the categorization paradigms had a higher sensitivity than the discrimination paradigms with SOAs equal to or longer than 44 ms, d' showed very similar values in all four experiments for SOAs of 31 ms and below, suggesting that early visual processing might be very similar in the four tasks.

Reaction Times

The response mode (eye or hand) and the type of paradigm (categorization or discrimination) both had a strong impact on reaction times (RTs). Mean RTs were significantly different across the four tasks, whatever the SOA, $F(3, 307) = 147.03$, $p < .001$. In the three manual tasks, mean RTs were all above 400 ms (see Table 1). By contrast, saccadic RTs were much shorter, with a mean value of only 250 ms (individual range = 193–339, SOA range = 245–255). Moreover, the speed of response also depended on whether the task involved a discrimination or a categorization

paradigm. Mean RTs were shorter in both discrimination tasks regardless of the effector (saccadic: 250 ms [307–572]; manual: 404 ms [391–426]) than they were in the two categorization tasks, either with go/no-go (422 ms [362–509; 399–453]) or yes–no responses (477 ms [367–626; 470–486]).

Furthermore, in the two discrimination tasks, mean RTs did not fluctuate much with SOA, although there were large differences between participants (see Table 1 and Figure 3A). In contrast, in the categorization tasks, we observed an increase in RT for the shortest SOAs: In the go/no-go experiment, the difference in mean RTs between short (<44 ms) and long (≥ 44 ms) SOAs reached significance ($p < .01$), with a maximum difference of 54 ms. In the yes–no experiment, the same decreasing RT profile was observed on correct yes–no responses at increasing SOAs, and although it did not reach significance, the difference reached a maximum of 40 ms.

Figure 3B shows the RT distributions in the four experiments. Minimal RTs were calculated by chi-square tests as the first 10-ms time bin for which correct responses significantly exceeded errors. This value is an estimation of the minimal processing duration from visual input to motor output. Globally, comparing minimal RTs in the four experiments led to three main conclusions. First, processing speed critically depends on the effector, as saccadic discrimination responses can be initiated more than twice as quickly as manual discrimination responses (120 ms vs. 270 ms). Second, the task requirement is also critical: discrimination tasks (saccadic or manual) lead to shorter RTs than categorization tasks (go/no-go or yes–no). Finally, the number of effectors is also critical, as participants were faster in the go/no-go task than in the yes–no task (422 ms vs. 477 ms), $F(3, 307) = 147.033$, $p < .001$.

Speed–Accuracy Trade-Off

In the previous sections, we analyzed performance separately in terms of accuracy and response speed, but these parameters are often closely correlated because of speed–accuracy trade-offs. Usually, a shift to shorter RTs results in an increase in errors, whereas focusing on accuracy leads to longer RTs (Meyer, Irwin,

Table 1
Numbers of Responses and Reaction Times (RTs) in All Experiments

| Experiment and response type or RT data	Stimulus onset asynchrony (ms)							
	6	12	19	31	44	81	106	All
Number of responses								
1: Manual go/no-go categorization								
Hit	310	443	636	1,070	1,232	1,249	1,250	6,190
False alarm	262	303	387	431	320	214	189	2,106
Miss	970	838	644	210	48	31	30	2,771
Correct Rejection	1,018	976	893	849	960	1,066	1,091	6,853
2: Manual yes–no categorization								
Hit	259	296	433	571	638	640	648	3,485
False alarm	188	237	193	183	156	89	83	1,129
Miss	462	425	287	142	83	77	73	1,549
Correct Rejection	529	480	527	543	563	634	631	3,907
3: Saccadic discrimination								
Hit	355	425	490	572	627	671	657	3797
False alarm	356	281	212	151	94	44	56	1194
4: Manual discrimination								
Hit	392	434	558	638	673	701	695	4,091
False alarm	368	326	202	122	87	59	63	1,227
RT (ms)								
1: Manual go/no-go categorization								
M	439	453	442	421	401	399	399	422
SD	68	75	53	46	41	39	41	48
Minimum	340	347	368	350	326	332	322	362
Maximum	535	575	545	501	474	466	481	509
2: Manual yes-no categorization								
M	497	494	485	464	457	458	459	473
SD	104	92	80	68	62	54	56	72
Minimum	376	362	374	374	367	373	378	372
Maximum	750	673	644	592	546	527	541	611
3: Saccadic discrimination								
M	255	252	250	245	246	250	250	250
SD	43	39	47	45	47	55	60	48
Minimum	209	211	200	189	192	173	173	193
Maximum	339	327	338	326	338	347	357	339
4: Manual discrimination								
M	407	402	396	386	381	381	384	391
SD	78	73	64	61	55	55	48	61
Minimum	304	294	305	304	305	317	318	307
Maximum	528	489	474	463	458	457	455	472

Osman, & Kounios, 1988; Pachella, 1974). Such trade-offs have been found in several of our previous experiments with natural scenes (Kirchner & Thorpe, 2006; Thorpe et al., 1996). However, by limiting the amount of available information with a masking procedure, one can ask whether the compromise between speed and accuracy is still present. It is possible that if all the available information has already been processed, there would be little to gain from delaying the motor response.

We thus calculated the correlation between RT and accuracy in the four tasks on the basis of the mean data of each participant. Figure 4 shows the speed–accuracy correlation index as a function of SOA, separately for each task. In the manual go/no-go task, there was no speed-accuracy trade-off, whatever the SOA. In contrast, in the discrimination tasks, a correlation was present for SOAs longer than 44 ms, whereas it was reduced or even absent with SOAs shorter than 31 ms. Finally, in the yes–no task, a speed–accuracy trade-off was present, although it was less marked

than in the discrimination tasks. This suggests that when participants have to choose between two possible response alternatives (*left* or *right*, *yes* or *no*), the visual system can improve accuracy by slowing down the response. But this only seems to be true when the mask disrupts visual processing relatively late, allowing the visual system to process further the input of available information. In contrast, with short SOAs, it may not be useful to allow time for more information processing if all of the available information has already been extracted.

Discussion

The present study analyzed human performance in four different tasks to study rapid visual processing of natural scenes under high temporal constraints that result from a dynamic mask. To compare the four paradigms, it is important to identify clearly the differences that characterize them, which brings out three major disso-
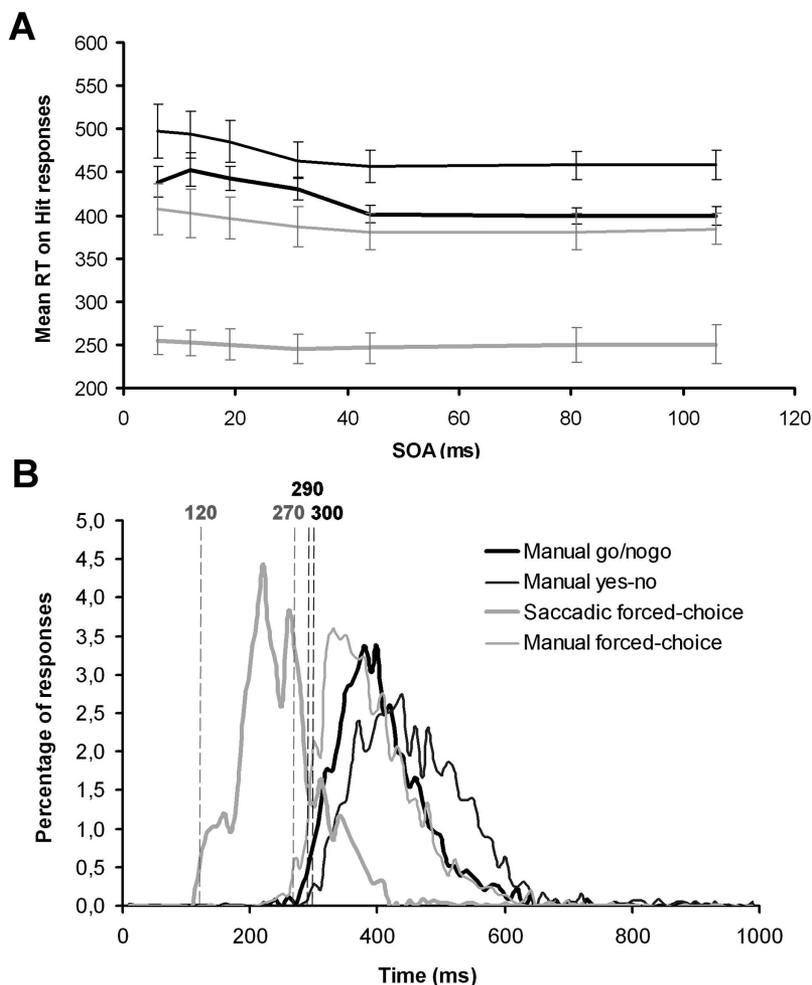
Figure 3. Reaction times (RTs) in the four experiments. Only values for hits were considered in the yes–no task. A: Mean RTs as a function of stimulus onset asynchrony (SOA). Error bars represent standard errors of the means. B: Overall RT distributions, calculated as the difference between percentages of correct and incorrect responses for each time bin (10-ms bin width) in the four experiments. Dotted lines indicate the minimum time (first significant bin; i.e., the first bin in which correct responses significantly outnumber errors).

ciations, as schematized in Figure 5. The first dissociation can be seen at the *perceptual stage* and directly concerns the processing that can be performed on the raw input to the visual system: In the paradigms used in Experiments 1 and 2, only one image was flashed centrally on the screen, whereas the paradigms used in Experiments 3 and 4 required the parallel processing of two different images. Second, the tasks differed at the *decisional level*. In Experiments 1 and 2, participants were asked to indicate whether an animal was present; the judgment thus was categorical in nature. In Experiments 3 and 4, however, participants could rely on the assumption that an animal was always present but had to determine on which side the animal appeared. Thus, categorization tasks (Experiments 1 and 2) required a decision that was based on the analysis of a single stimulus, whereas discrimination tasks (Experiments 3 and 4) involved the processing (and possibly the comparison) of two simultaneously presented stimuli to decide which one contained the animal. Finally, at the very end of the perception–action stream, we varied the *response mode*, which

either involved a saccadic eye movement (Experiment 3) or manual responses (Experiments 1, 2, and 4).

The motor stage showed large differences depending on which effectors were used (saccadic eye movement or manual response) and whether the response involved one response given by one effector (go/no-go categorization task with manual response), two different responses given by a single effector (saccadic discrimination task), or the same response given by two different effectors (manual responses with the two hands in the discrimination task and in the yes–no categorization task). Although these differences might be thought to have a critical impact on sensorimotor processing, we have shown here that performance measures were comparable in all tasks, in particular with SOAs of 31 ms or less.

Early Sensory Processing Is Independent of the Task

When the mask was presented shortly after the stimulus, visual processing was strongly affected, as demonstrated by low perfor-
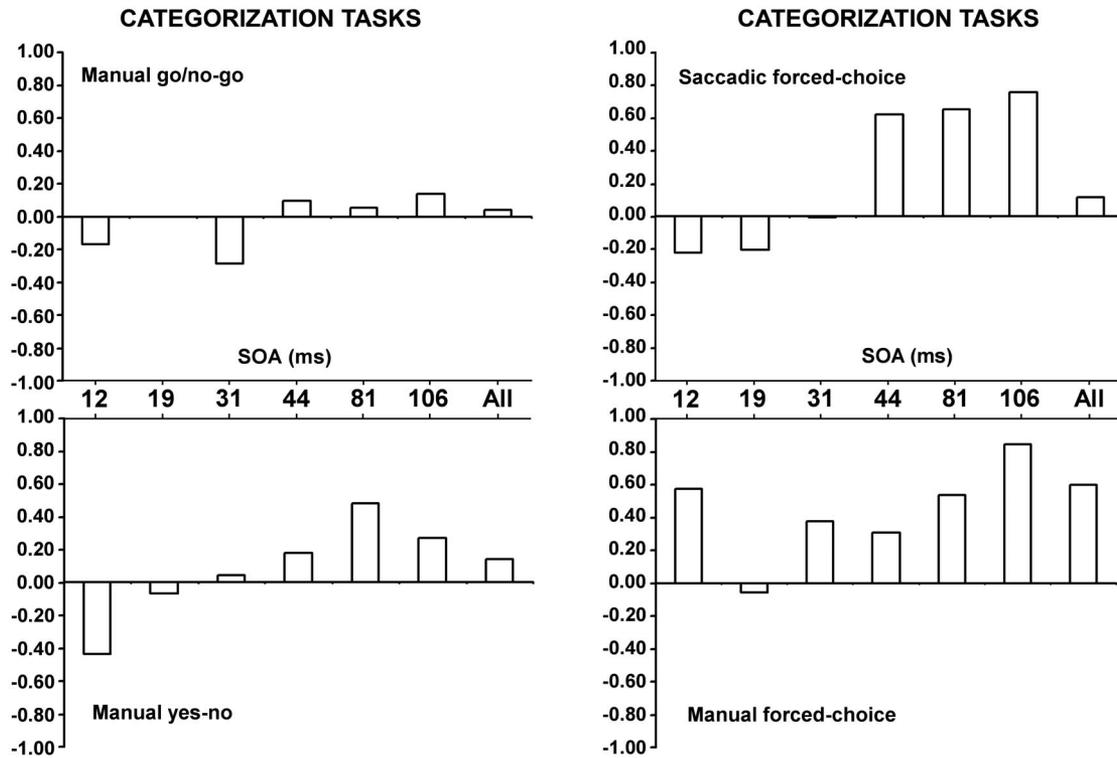
CATEGORIZATION TASKS

Manual go/no-go

SOA (ms)

12 19 31 44 81 106 All

CATEGORIZATION TASKS

Saccadic forced-choice

SOA (ms)

12 19 31 44 81 106 All

Manual yes-no

Manual forced-choice

Figure 4. Correlation index (*r* [*y*-axes]) as a function of stimulus onset asynchrony (SOA; from 12–106 ms) and for all SOAs grouped together (bar in the right side). For each SOA condition, the Pearson correlation was calculated between the mean accuracy and the mean reaction time of each participant. Each bar thus represents a correlation index that takes into account as many points as there are participants in a given experiment.
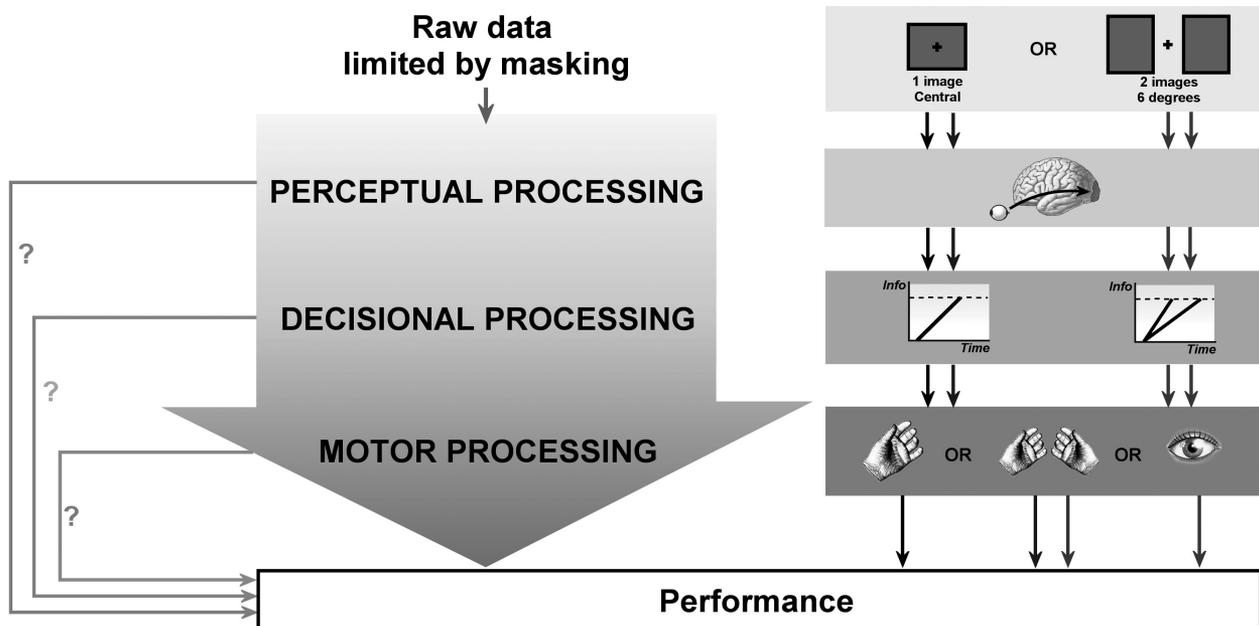
Raw data limited by masking

PERCEPTUAL PROCESSING

DECISIONAL PROCESSING

MOTOR PROCESSING

?

?

?

Performance

1 image Central

OR

2 images 6 degrees

Info

Time

Info

Time

OR

OR

Figure 5. Schematic representation of the different stages within the visual sensorimotor processing stream. Differences that characterize the four experiments are indicated for each stage.

mance. A logical hypothesis that has been proposed previously is that the amount of information available from low-level properties of the stimulus is limited because of interference between the stimulus and the mask that takes place early in the cortical pathway, probably before area V4 (Bacon-Macé et al., 2005). It is even plausible that a substantial part of the masking could occur as early as the retina. The d' analysis indicated similar sensitivity in all experiments when the mask was presented shortly after the stimulus, with an SOA of 31 ms or less, suggesting that the amount of information that can be extracted from the stimuli at these SOAs may be similar between experiments. Whatever the task or the response mode, the visual system seems to integrate stimulus information with the same efficiency, and disrupting the processing leads to the same limitations. When the mask is close to the image, information about the stimulus enters the visual system but is wiped out by the information about the mask as it goes along the visual pathway. Only information that can be extracted very rapidly can escape interference from the mask and be used further downstream (Bacon-Macé et al., 2005; Romo & Salinas, 2003; VanRullen & Thorpe, 2002). We therefore propose that in all four experiments, maximum performance was limited by the amount of visual analysis that could be achieved before masking occurred at early sensory stages. This idea is reinforced by the absence of a speed–accuracy trade-off for SOA values shorter than 30–40 ms, which suggests that the visual system was working in a "just-in-time" mode and could not manipulate the available information to refine potential interpretations of the stimulus.

It is important to point out the fact that sensitivity was similar for all four tasks in this range of SOA, even though the number of stimuli to be processed varied. This result reinforces the view that, at least in the specific case of images presented simultaneously to the two visual hemifields, processing can be performed in parallel without the need for focused attention (Rousselet et al., 2002). Other data favoring the view that the animal detection task can be performed in the virtual absence of focalized attention has come from a dual-task experiment in which animal categorization could be done at almost no cost in peripheral vision while participants were fully attending to a central task (Li et al., 2002). More recent data show that other high-level visual tasks, including face-gender processing (Reddy, Wilken, & Koch, 2004) and even face identification (Reddy, Reddy, & Koch, 2006), can also be done in the near absence of attention.

Depending on Task and Response Mode, Decisional and Motor Processes Are Dissociated With Longer SOAs

With SOAs of 44 ms and longer, accuracy improved and reached a plateau in the four experiments—above 83% correct. The interference between the stimulus and the mask information occurred later, and the visual system had more time to process the stimulus. Interestingly, from this point onward, d' measures show that sensitivity began to differ between the experiments. Specifically, the discrimination experiments had a lower sensitivity than the categorization experiments (see Figure 2B). This effect was not visible in the raw accuracy measures when no correction was applied in the discrimination

tasks to compensate for the processing of two images at the same time (Macmillan, 2002).

Categorization tasks have a higher d' compared with discrimination tasks. Several nonmutually exclusive hypotheses can explain the d' reduction in the discrimination tasks compared with the categorization tasks (see Figure 5). A first point concerns the raw data entering the system, because at least some drop in performance could be expected with the image eccentricity in the discrimination tasks (6°). In fact, the impact of eccentricity on performance has already been determined in previous categorization studies in which, at eccentricities of 4°–13° from the center, a decrease of 3%–4% was accompanied by an increase in mean RTs of about 15 ms (Fize et al., in press; Thorpe, Gegenfurtner, Fabre-Thorpe, & Bulthoff, 2001).

Second, the analysis of mean RTs showed that the latencies in the discrimination experiments were remarkably stable from one condition to another, even with the shortest SOAs, unlike the categorization experiments, in which RTs increased with decreasing SOA (see Figure 3A and Table 1). Thus, not only were RTs shorter in the discrimination tasks, they also seem to have been unrelated to the task difficulty. Typically, in the discrimination paradigms, whether manual or ocular, it could be that the final decision is based on more automatic response triggering. Because in the discrimination tasks all trials were go trials, the gating of the motor response to be produced might have been reached faster. This could constitute one explanation for the short RTs in discrimination as compared with categorization tasks. Furthermore, this hypothesis might also account for the lower sensitivity in discrimination tasks, because the visuomotor system may sometimes produce a slightly premature response relative to the selection process, leading to greater inaccuracy.

This hypothesis was further supported by the analysis of speed–accuracy trade-offs (see Figure 4), which indicated that in the discrimination experiments and for SOAs longer than 44 ms, the slowest participants tended to be more accurate than the fastest participants. This suggests that when additional information can be extracted, a response strategy can be applied to modulate the compromise between speed and accuracy. In contrast, in the categorization experiments, this speed–accuracy trade-off was either reduced (yes–no task) or totally absent (go/no-go task), even with the longest SOAs. How can we explain that there was a speed–accuracy trade-off in the discrimination but not in the categorization tasks and that it only appeared at longer SOAs? In both categorization tasks, participants had a low response rate with short SOAs, and despite the instruction to respond on half of the trials in each series, they were strongly inhibited in the more difficult conditions (the response rate decreased to around 20% with an SOA of 6 ms in both tasks). A first explanation for the absence of speed–accuracy trade-offs in these tasks would be that the participants were very concerned by the instruction given to force the rate of manual responses, which could have prevented them from applying other strategies such as a modulation of speed and accuracy. The existence of such an inhibitory effect is particularly plausible given that speed–accuracy trade-offs have already been observed in previous experiments using the go/no-go task but no masking (Delorme et al., 2000; Thorpe et al., 1996). In the discriminations tasks, however, participants were asked to respond on every trial because an animal was always present. To explain speed–accuracy trade-offs in such choice tasks, psychological

models of decision making have proposed that these trade-offs should be related to moving the decision boundaries that are located at the end of a sensory–decisional accumulation process closer together so that responses will be made more quickly but will be more prone to errors (Ratcliff & Rouder, 1998; P. L. Smith & Ratcliff, 2004). Our data are in agreement with this prediction, as RTs in the discrimination tasks were shorter than those in the categorization tasks. However, on the basis of these models, we were expecting a correlation between RTs and accuracy (Palmer, Huk, & Shadlen, 2005), whereas RTs were constant for all SOA values, despite the fact that accuracy ranged from chance level all the way to 90% correct. Therefore, it is all the more remarkable that we observed speed–accuracy trade-offs within the long SOA conditions. It thus seems that decision boundaries were low, so accuracy was lower than it was in the categorization tasks, but at the same time flexible so that they allowed some variability when more and more sensory information was available at longer SOAs.

Finally, an interesting secondary issue that remains is whether a decision is reached after independent and parallel processing of each image or the decision is based on a comparison process. Given the response speed, it has previously been proposed that the two images are processed independently, each image in one ventral system, until one of these processes reaches the decision boundary so that a response to the opposite side can be triggered (Kirchner & Thorpe, 2006). However, an alternative hypothesis supposes that the visual system takes advantage of the presentation of the two images by directly comparing the outcome of the processing results in the two hemifields. This comparison might involve higher level areas such as (dorsolateral) prefrontal cortex and, in the case of eye movements, the frontal eye fields. With regard to the difference in accuracy between discrimination and categorization tasks, the integration hypothesis would predict that the rate of information accumulation on two simultaneous but integrated stimuli may be greater, leading to shorter RTs, than in case of a single-image task (categorization) in which participants cannot use any other strategy than comparing the target with internal representations. A disadvantage of such comparison mechanisms would be that when the visual system has to process noisy information, as would be the case with masking, the sensory information reaching the boundary might be less reliable, so responses would be more likely to be made in error.

Discrimination tasks, whether manual or ocular, have similar sensitivities. Another interesting point comes from the similar sensitivities observed in the two discrimination tasks, despite the fact that one used the manual and the other used the ocular motor system. Remarkably, although accuracy and d' were both similar, mean RTs were very different, as saccadic responses could be triggered almost 150 ms before manual responses (250 ms vs. 391 ms). It is difficult to define exactly which anatomical pathways were involved in these experiments, but the contribution of the ventral pathway, which is known to be implicated in object recognition (Felleman & Van Essen, 1991; Goodale & Milner, 1992; Tanaka, 1993) seems highly likely and comparable in both tasks, especially at the sensory level. But in view of the extremely short RTs, one can wonder whether some processing steps could have been bypassed to allow the first meaningful responses to be initiated so quickly. A plausible cortical route for manual response preparation might go from the ventral pathway to the spinal cord via areas including the premotor and motor cortex (Boussaoud, di

Pellegrino, & Wise, 1995; Foxe & Simpson, 2002; Rizzolatti, Luppino, & Matelli, 1998). The 150-ms increase in RTs in the case of manual responses can partly be justified by the numerous stages from visual to motor areas and longer transmission delays from one stage to another, together with the additional delays required to go through the spinal cord to the hand. However, the similar sensitivity in the manual and saccadic tasks at each SOA step suggests that the amount of sensory information was similar and that the paradigm alone (here, discrimination) determined the level of information processing necessary to make a response. Whether the mode of response is manual or ocular may constitute a factor that operates at a later stage, leaving initial processing unchanged.

Categorization tasks show a small difference of sensitivity. Figure 2B shows a difference between the go/no-go and the yes–no experiments for SOAs longer than 44 ms. Not only did the go/no-go task have a better d' at this SOA range, RTs were also shorter than those in the yes–no task. Differences between these two sorts of tasks have already been reported in the context of lexical decision experiments (Hino & Lupker, 2000; Perea, Rosa, & Gomez, 2002; Siakaluk, Buchanan, & Westbury, 2003). They are usually explained by supposing that there are more competitive stages related to the decisional and motor mechanisms in the yes–no task. Indeed, at the decisional level, participants have to produce only one answer. However, although in the yes–no experiment, participants were required to release one of two buttons, which implies that two distinct motor programs had to be prepared and initiated, the response selection process was apparently simpler in the go/no-go task, in which participants responded with one button and only in the case of an affirmative response.

Finally, it is important to note that previous studies that have aimed at comparing discrimination and categorization paradigms have typically failed to find any difference in d' after applying the correction for the processing of two simultaneous stimuli in a forced-choice task (Green & Swets, 1966; Khoe, Kroll, Yonelinas, Dobbins, & Knight, 2000; Yonelinas, Hockley, & Murdock, 1992). We were thus surprised to observe such marked differences between these tasks at the longest SOAs, when perception was at its best. Interpretations based on traditional signal detection theory have been subject to debate because they do not always provide a reliable measure of the accuracy in memory tests (Kroll et al., 2002). Other models involving modifications of classical signal detection theory have been recently proposed and found to provide a better fit for the observed receiver-operating characteristics. They are based on additional measures of performance such as confidence judgments and, thus, could not be applied here. However, it is noteworthy that the studies in question mainly used recognition memory tests, which are quite different from our analysis of sensory processing in the visual system. Moreover, the debate has mainly focused on the number of sources of information needed to explain competitive responses, which is not representative of our experiments, in which the source of information was probably unique, coming from the sensory processing in the ventral pathways (Banks, 2000; Dunn, 2004; D. G. Smith & Duncan, 2004).

Attention and Parallel Processing

One of the most significant differences in this set of experiments opposes the categorization experiments, in which only one picture

had to be visually processed, with the discrimination experiments, which required the processing of two simultaneously presented images. As discussed before, it is still difficult to determine either the level of processing performed when two pictures are presented or the pictures' potential interaction, but it appears very clear that at least at the shortest SOAs, sensitivity was the same in the four experiments, indicating that the time course of processing is similar whether participants are analyzing one central or two peripheral pictures. Previous work from our group (Rousselet et al., 2002) has argued that despite their complexity, natural scenes can nevertheless be processed in parallel when presented separately in each hemifield. Rousselet et al. (2002) also recorded the cerebral activity using EEG and observed that signals were similar in the two conditions over the occipital areas, unlike frontal sites, in which a difference between signals clearly appeared from 190 ms. They argued that the competition mechanisms in the processing of two simultaneously presented pictures may occur at quite a high level of integration, supporting psychological models that situate the competitive process at a late selection stage (Chun & Potter, 1995; Duncan, 1980). Our results also point in the same direction. In the discrimination tasks, the two pictures were flashed in each hemifield, but nevertheless, the initial 30 ms or so of processing appeared to be common with the single-image presentation. The only differences appeared later on, at longer SOAs, with a lower sensitivity for the discrimination tasks that engaged the presentation of two stimuli.

These data are particularly relevant for the literature on attention and to the models that have been proposed to explain the electrophysiological or clinical data showing a competition process between different stimuli (Desimone & Duncan, 1995; Luck, Girelli, McDermott, & Ford, 1997; Treisman, 1998). It has notably been observed that when several stimuli are present within the same neuronal receptive field, the initial response is remarkably robust regardless of the number of stimuli, but when compared with the response to a single picture, the response to several stimuli shows a degradation that can only be observed after a certain delay (Chelazzi, Duncan, Miller, & Desimone, 1998; Chelazzi, Miller, Duncan, & Desimone, 1993, 2001). This has been interpreted as a demonstration that parallel processing occurs in the early stages of the recognition process, before a more competitive, attention-demanding stage takes place. In the biased-competition model proposed by Duncan and colleagues, competition would be mediated by bottom-up or top-down modulations, depending on the nature of the task or the physical properties of the stimulus (Desimone & Duncan, 1995; Duncan, Humphreys, & Ward, 1997).

In the present study, the very nature of the tasks differed, with categorization experiments on the one hand and discrimination experiments on the other hand. Although it might be expected that discriminations could potentially require competitive and/or comparative processes (the significant information content for the task were mutually exclusive in the two simultaneously presented pictures), we were nevertheless led to the same conclusions, and observed that the task-related factors intervened at a later stage. This point can be related to the observation that many features of natural scenes can be processed in parallel although the level of representation of these first-extracted features may be limited (Evans & Treisman, 2005). Evans and Treisman argued that although the first-extracted cues can be sufficient for detection, full identification of the scene could still require focused attention.

Studies using natural scenes as stimuli are not simple to integrate into the long tradition of research about the attentional requirements of visual search or dual tasks. Remarkably, although scenes with animals can be processed preattentively and lead to high degrees of performance in categorization tasks when focal attention is occupied elsewhere in a dual-task experiment (Li et al., 2002), they do not pop out in a traditional visual search task (VanRullen et al., 2004). The recent distinction between parallel and preattentive processes may constitute a turning point in the understanding of fast visual processing of natural scenes. The existence of neurons in the ventral pathway that code for intermediate or even high representations of features might allow the rapid extraction of significant amounts of information about a scene before some higher forms of competition occur. This late competition could depend heavily on the number of stimuli in the scene and the distance between objects that together determine the interference within the receptive fields. It is noteworthy that in the present and previous studies, the stimuli were presented at a relatively large distance from each other, one in each hemifield (VanRullen et al., 2004). This may constitute a special configuration in which the neuronal populations coding the pictures remain distinct throughout virtually the whole length of the ventral stream (Rousselet, Thorpe, & Fabre-Thorpe, 2004a; VanRullen, Reddy, & Fei-Fei, 2005; VanRullen et al., 2004).

A Possible Difference Between Eye Movements and Manual Responses

Given the remarkable similarity in performance that we found between the manual and saccade-based discrimination tasks, it is important to ask whether there may be additional hidden penalties associated with the shortened RTs that we found with the eye-movement version of the task. Why would the brain add an additional 150 ms to the mean RT without reason? Recent work has shown that online corrections of hand movements can be achieved in as little as 120 ms in response to a change in a visual stimulus, implying that the input–output linking in online manual control can be particularly short in some circumstances (Angel & Higgins, 1969; Beaubaton & Hay, 1986; Brenner & Smeets, 2004). There are some reasons to believe that eye-movement-based decisions may not have the flexibility that comes with the longer latency manual responses. Biologically, it is more pertinent to perform accurate manual responses than accurate saccades. Indeed, there is no real penalty for the performance of an incorrect shift of gaze, apart from the need to make a corrective saccade, and it could be of vital importance for survival to perform fast ocular movements toward objects that are of potential interest. However, an incorrect manual response can be much more dangerous, and thus, additional processing of incoming information could be very useful. Nevertheless, although there may be differences in the degree of flexibility of saccade and manually based decisions, this should not detract from the underlying message of the present study—namely, the idea that the initial processing dynamics of natural image processing are remarkably similar for a range of different experimental paradigms.

Conclusion

Comparison of four experimental paradigms shows a clear dissociation in the effect of masking on performance depending on

how early processing is disrupted, with a critical value around 40 ms. When information processing was disrupted early on, sensitivity was similar for all four experimental protocols, whatever the paradigm or the response mode. This is of considerable interest as it suggests that information integration processes are the same in the initial phase of sensory encoding. Such results also support feedforward models of information transmission, which have previously been proposed to explain the remarkable speed with which humans can respond in such categorization and discrimination tasks (Bacon-Macé et al., 2005; Delorme, Rousselet, Macé, & Fabre-Thorpe, 2004; Fabre-Thorpe et al., 2001; Kirchner & Thorpe, 2006; VanRullen & Koch, 2003).

Finally, one of the main results of the present study is the clear validation of paradigms relying on eye-movement responses to analyze early perceptual processing. We have shown sensitivity to be similar whether the task involves manual or ocular responses, at least in the first 40 ms. But the fact that saccadic RTs were almost half as long allows temporal constraints to be further refined and supports the idea that early processing of visual information relies on an automatic mode despite the apparent complexity of processing required by natural scenes.

References

Angel, R. W., & Higgins, J. R. (1969). Correction of false moves in pursuit tracking. *Journal of Experimental Psychology, 82,* 185–187.

Bacon-Macé, N., Macé, M. J., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research, 45,* 1459–1469.

Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science, 11,* 267–273.

Beaubaton, D., & Hay, L. (1986). Contribution of visual information to feedforward and feedback processes in rapid pointing movements. *Human Movement Science, 5,* 19–34.

Boussaoud, D., di Pellegrino, G., & Wise, S. P. (1995). Frontal lobe mechanisms subserving vision-for-action versus vision-for-perception. *Behavioural Brain Research, 72,* 1–15.

Brenner, E., & Smeets, J. B. (2004). Colour vision can contribute to fast corrections of arm movements. *Experimental Brain Research, 158,* 302–307.

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology, 80,* 2918–2940.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993, May 27). A neural basis for visual search in inferior temporal cortex. *Nature, 363,* 345–347.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (2001). Responses of neurons in macaque area V4 during memory-guided visual search. *Cerebral Cortex, 11,* 761–772.

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.

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

Delorme, A., Rousselet, G. A., Macé, M. J., & Fabre-Thorpe, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Brain Research: Cognitive Brain Research, 19,* 103–113.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review, 87,* 272–300.

Duncan, J., Humphreys, G., & Ward, R. (1997). Competitive brain activity in visual attention. *Current Opinion in Neurobiology, 7,* 255–261.

Dunn, J. C. (2004). Remember–know: A matter of confidence. *Psychological Review, 111,* 524–542.

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1476–1492.

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., Macé, M. J.-M. & Joubert, O. (2005). Dog or animal? What comes first in vision? [Abstract]. *Perception, 34*(Suppl.), 8.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1,* 1–47.

Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences, 16,* 553–567.

Fize, D., Fabre-Thorpe, M., Richard, G., Doyon, B., & Thorpe, S. J. (2005). Rapid categorization of foveal and extrafoveal natural images: Associated ERPs and effects of lateralization. *Brain and Cognition, 59,* 145–158.

Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans: A framework for defining "early" visual processing. *Experimental Brain Research, 142,* 139–150.

Gold, J. I., & Shadlen, M. N. (2000, March 23). Representation of a perceptual decision in developing oculomotor commands. *Nature, 404,* 390–394.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15,* 20–25.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* Huntington, NY: Robert E. Krieger.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition. *Psychological Science, 16,* 152–160.

Hanes, D. P., & Schall, J. D. (1996, October 18). Neural control of voluntary movement initiation. *Science, 274,* 427–430.

Hasbroucq, T., Mouret, I., Seal, J., & Akamatsu, M. (1995). Finger pairings in two-choice reaction time tasks: Does the between-hands advantage reflect response preparation? *Journal of Motor Behavior, 27,* 251–262.

Heekeren, H. R., Marrett, S., Bandettini, P. A., & Ungerleider, L. G. (2004, October 14). A general mechanism for perceptual decision-making in the human brain. *Nature, 431,* 859–862.

Hino, Y., & Lupker, S. J. (2000). Effects of word frequency and spelling-to-sound regularity in naming with and without preceding lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 166–183.

Humphreys, G. W., Riddoch, M. J., & Price, C. J. (1997). Top-down processes in object identification: Evidence from experimental psychology, neuropsychology and functional anatomy. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 352,* 1275–1282.

Khoe, W., Kroll, N. E., Yonelinas, A. P., Dobbins, I. G., & Knight, R. T. (2000). The contribution of recollection and familiarity to yes–no and forced-choice recognition tests in healthy subjects and amnesics. *Neuropsychologia, 38,* 1333–1341.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46,* 1762–1776.

Kroll, N. E., Yonelinas, A. P., Dobbins, I. G., & Frederick, C. M. (2002). Separating sensitivity from response bias: Implications of comparisons of yes–no and forced-choice tests for models and measures of recognition memory. *Journal of Experimental Psychology: General, 131,* 241–254.

Li, F. 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.

Luck, S. J., Girelli, M., McDermott, M. T., & Ford, M. A. (1997). Bridging the gap between monkey neurophysiology and human perception: An

ambiguity resolution theory of visual selective attention. *Cognitive Psychology, 33,* 64–87.

Macé, M. J. M., Bacon-Macé, N., Nespoulous, J. L., & Fabre-Thorpe, M. (2007). *What's seen first in the brain: The animal or the bird?* Manuscript submitted for publication.

Macmillan, N. A. (2002). Signal detection theory. In H. Pashler (Ed.), *Steven's handbook of experimental psychology, Vol. 4: Methodology in experimental psychology* (3rd ed., pp. 43–90). New York: Wiley.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge, England: Cambridge University Press.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287–330.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience, 4,* 310–322.

Meyer, D. E., Irwin, D. E., Osman, A. M., & Kounios, J. (1988). The dynamics of cognition and action: Mental processes inferred from speed–accuracy decomposition. *Psychological Review, 95,* 183–237.

Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology, 26,* 3–67.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. W. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41–82). Hillsdale, NJ: Erlbaum.

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision, 5,* 376–404.

Perea, M., Rosa, E., & Gomez, C. (2002). Is the go/no-go lexical decision task an alternative to the yes/no lexical decision task? *Memory & Cognition, 30,* 34–45.

Perrett, D. I., Oram, M. W., & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: An account of generalisation of recognition without mental transformations. *Cognition, 67,* 111–145.

Ratcliff, R., & Rouder, J. F. (1998). Modeling response times for two-choice decisions. *Psychological Sciences, 9,* 347–356.

Reddy, L., Reddy, L., & Koch, C. (2006). Face identification in the near-absence of focal attention. *Vision Research, 46,* 2336–2343.

Reddy, L., Wilken, P., & Koch, C. (2004). Face–gender discrimination is possible in the near-absence of attention. *Journal of Vision, 4,* 106–117.

Rizzolatti, G., Luppino, G., & Matelli, M. (1998). The organization of the cortical motor system: New concepts. *Electroencephalography and Clinical Neurophysiology, 106,* 283–296.

Romo, R., & Salinas, E. (2003). Cognitive neuroscience: Flutter discrimination: Neural codes, perception, memory and decision making. *Nature Reviews Neuroscience, 4,* 203–218.

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., Macé, M. J., & 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. (2004a). How parallel is visual processing in the ventral pathway? *Trends in Cognitive Science, 8,* 363–370.

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004b). Processing of one, two or four natural scenes in humans: The limits of parallelism. *Vision Research, 44,* 877–894.

Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1: Foundations.* Cambridge, MA: MIT Press.

Schall, J. D. (2001). Neural basis of deciding, choosing and acting. *Nature Reviews Neuroscience, 2,* 33–42.

Schall, J. D. (2004). On building a bridge between brain and behavior. *Annual Review of Psychology, 55,* 23–50.

Seidemann, E., Zohary, E., & Newsome, W. T. (1998, July 2). Temporal gating of neural signals during performance of a visual discrimination task. *Nature, 394,* 72–75.

Shadlen, M. N., & Newsome, W. T. (1996). Motion perception: Seeing and deciding. *Proceedings of the National Academy of Sciences, USA, 93,* 628–633.

Siakaluk, P. D., Buchanan, L., & Westbury, C. (2003). The effect of semantic distance in yes/no and go/no-go semantic categorization tasks. *Memory & Cognition, 31,* 100–113.

Smith, D. G., & Duncan, M. J. (2004). Testing theories of recognition memory by predicting performance across paradigms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 615–625.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences, 27,* 161–168.

Tanaka, K. (1993, October 29). Neuronal mechanisms of object recognition. *Science, 262,* 685–688.

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

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

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network, 14,* 391–412.

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

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

Ungerleider, L. G. (1985). The corticocortical pathways for object recognition and spatial perception. In C. Chagas, R. Gattass, & C. Gross (Eds.), *Pattern recognition mechanisms* (Vol. 54, pp. 21–37). Vatican City: Pontifical Academy of Sciences.

Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Current Opinion in Neurobiology, 4,* 157–165.

VanRullen, R., & Koch, C. (2003). Visual selective behavior can be triggered by a feed-forward process. *Journal of Cognitive Neuroscience, 15,* 209–217.

VanRullen, R., Reddy, L., & Fei-Fei, L. (2005). Binding is a local problem for natural objects and scenes. *Vision Research, 45,* 3133–3144.

VanRullen, R., Reddy, L., & Koch, C. (2004). Visual search and dual tasks reveal two distinct attentional resources. *Journal of Cognitive Neuroscience, 16,* 4–14.

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

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

VanRullen, R., & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Research, 42,* 2593–2615.

Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T., & Schwarzbach, J. (2003). Different time courses for visual perception and action priming. *Proceedings of the National Academy of Sciences, USA, 100,* 6275–6280.

Yonelinas, A. P., Hockley, W. E., & Murdock, B. B. (1992). Tests of the list-strength effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 345–355.