Act Quickly, Decide Later: Long-latency Visual Processing Underlies Perceptual Decisions but Not Reflexive Behavior

Jacob Jolij[1,2,3,4]**, H. Steven Scholte**[2]**, Simon van Gaal**[2]**,
Timothy L. Hodgson**[3]**, and Victor A. F. Lamme**[2]

Abstract

■ Humans largely guide their behavior by their visual representation of the world. Recent studies have shown that visual information can trigger behavior within 150 msec, suggesting that visually guided responses to external events, in fact, precede conscious awareness of those events. However, is such a view correct? By using a texture discrimination task, we show that the brain relies on long-latency visual processing in order to guide perceptual decisions. Decreasing stimulus saliency leads to selective changes in long-latency visually evoked potential components reflecting scene segmentation. These latency changes are accompanied by almost equal changes in simple RTs and points of subjective simultaneity. Furthermore, we find a strong correlation between individual RTs and the latencies of scene segmentation related components in the visually evoked potentials, showing that the processes underlying these late brain potentials are critical in triggering a response. However, using the same texture stimuli in an antisaccade task, we found that reflexive, but erroneous, prosaccades, but not antisaccades, can be triggered by earlier visual processes. In other words: The brain can act quickly, but decides late. Differences between our study and earlier findings suggesting that action precedes conscious awareness can be explained by assuming that task demands determine whether a fast and unconscious, or a slower and conscious, representation is used to initiate a visually guided response. ■

INTRODUCTION

Visual information plays an important role in guiding our actions. However, processing and extracting relevant information from visual input is a highly complex process that can take up a considerable amount of time. Recent studies have shown that it can take up to almost 400 msec before visual information is available for conscious report (e.g., Scharnowski et al., 2009; Heinen, Jolij, & Lamme, 2005). It is therefore not surprising that reacting to external events does not require awareness of those events. In the literature, there are numerous examples of dissociations between action and awareness. Goodale and Milner (1992), for example, describe a patient with visual agnosia (i.e., this patient has no conscious access to object identity) who is, nevertheless, perfectly able to manipulate objects. Another example is *blindsight*, the remarkable capability of some patients with a lesion to primary visual cortex to guess correctly about attributes of visual stimuli, such as color, orientation, or even facial expression (e.g., Weiskrantz, 1996).

Dissociations between action and awareness have been demonstrated in healthy individuals, too. For example,

when observers are instructed to pick up an object, the aperture between the thumb and the index finger is not influenced by distortions in the perceived size of the object, mimicking the situation in which conscious access to an object's features is accurate, but manipulation is still intact (Goodale & Milner, 1992). Complete dissociations between action and awareness have been demonstrated, too: Using TMS of primary visual cortex, blindsight has been reported in normal observers for color, orientation, and emotional expression (Boyer, Harrison, & Ro, 2005; Jolij & Lamme, 2005).

It seems that conscious awareness is not a prerequisite for action: Strictly speaking, there is no need to see something in order to respond to it. Presently, there is general consensus in the literature that there are different modes of processing in the visual system: visual processes that drive action and visual processes that result in conscious perception. To some extent, these processes are independent and rely on different neural structures. Visual processing for action is done in the dorsal cortical areas, and driven by fast magnocellular visual inputs, allowing for quick responses. Visual processing for perception, on the other hand, is done in more ventral cortical areas, and driven by parvocellular inputs, that are capable of representing more spatial details, but slower in processing visual information (Goodale & Milner, 1992).

[1]University of Groningen, The Netherlands, [2]University of Amsterdam, The Netherlands, [3]University of Exeter, UK, [4]Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland

However, within the "vision-for-perception" system, rapid stimulus processing preceding awareness is possible as well. Thorpe, Fize, and Marlot (1996) demonstrated that EEG responses to different categories of stimuli, such as animals versus inanimate objects, already deviate as early as 150 msec after stimulus presentation, showing that object classification can be accomplished extremely fast. Using an analysis of RT distributions to masked and unmasked stimuli, VanRullen and Koch (2003) demonstrated that a single feedforward sweep through the visual hierarchy is sufficient to trigger a behavioral response to the presence of a target. They presented masked and unmasked images of natural scenes for 26 msec, while subjects were engaged in go/no-go task: They had to respond only when the scene contained an animal. The analysis of the RT pattern revealed that the moment at which scenes containing an animal were discriminated from scenes not containing an animal (i.e., the first moment in time in which number of hits is significantly higher than the number of false alarms) was the same for both the masked and the unmasked stimuli, showing that the information extracted during the first 26 msec of visual processing can trigger a behavioral response related to a perceptual decision, independent of visual awareness.

Taken together, there is ample evidence that we can respond to visual events in our environment before we become aware of them—it almost seems that in order to explain human perceptual decision-making, awareness can be left out of the equation. Indeed, recent models of perceptual decision-making do not assign any special value to perceptual awareness. Instead, when confronted with a stimulus, the system simply responds when sufficient information is available in order to make a decision, based on internally set thresholds (e.g., Ratcliff & McKoon, 2008). Whether or not that information is represented in conscious awareness is irrelevant.

However, is this idea correct? Here we used texture checkerboards in order to investigate what information triggers behavior. Visual processing of texture checkerboards requires texture segregation, which is a two-stage process: First, borders of figures are detected, and subsequently, the figures are filled in. Border detection occurs around 80–90 msec, and is likely to be the result of lateral inhibition within cortical areas, whereas figure filling-in can take up to 200 msec and depends on re-entrant processing (Scholte, Jolij, Fahrenfort, & Lamme, 2008; Jehee, Roelfsema, Deco, Murre, & Lamme, 2007; Heinen et al., 2005; Roelfsema, Lamme, Spekreijse, & Bosch, 2002; Caputo & Casco, 1999; Lamme, 1995). This latter stage has been linked to perceptual awareness of texture stimuli, whereas the former stage could be sufficient in order to detect presence of a texture stimulus (Jehee, Lamme, & Roelfsema, 2007; Heinen et al., 2005; Lamme, 1995, 2003; Supèr, Lamme, & Spekreijse, 2001).

Interestingly, evoked potential components related to filling-in have been shown to increase in latency when decreasing perceived stimulus segregation strength, whereas earlier components reflecting border detection remain unaffected (Jolij et al., 2007; Caputo & Casco, 1999). We investigated the relation between texture processing and behavior in three experiments, in which we varied perceived segmentation strength of texture and the task at hand. In the first experiment, participants had to press a button whenever they detected a texture, irrespective of perceived segregation strength. EEG was measured during this experiment. In Experiment 2, we measured points of subjective simultaneity for two types of textures differing in perceived segmentation strength and the onset of a sound. In the last experiment, we used the two texture types as target stimuli in an antisaccade task.

We found that simple detection RTs to textures correspond strongly with latency of filling-in related component (Experiment 1). Points of subjective simultaneity for textures and a sound also showed a similar correspondence (Experiment 2). However, an analysis of single-trial EEG data revealed that after approximately 130 msec, sufficient information is available to reliably detect the presence of a texture, irrespective of perceived segregation strength. Data from the antisaccade experiment (Experiment 3) show that this early activity can be used to trigger reflexive behavior, in this case, erroneous express-saccades toward texture stimuli. Together, our findings suggest that responses to external events may, in some cases, not be triggered as soon as information is available. Instead, visually guided responses to external events seem to be initiated only when these external events are registered in visual awareness, unless it concerns more reflexive behavior.

METHODS

Participants

All participants were healthy freshman psychology students from the University of Amsterdam (Experiments 1 and 2) and the University of Exeter (Experiment 3), with no reported history of neurological or psychiatric health problems, and normal or corrected-to-normal vision. They received either study-credit or €7 per hour for participation. Thirty-three participants (18–24 years, 21 women) participated in Experiment 1, 9 participants (18–19 years, 6 women) participated in Experiment 2, and 11 participants (18–34 years, all women) participated in Experiment 3. None of them participated in any of the other experiments of this study. All participants were naive observers and had never participated in texture discrimination experiments before. Written informed consent was obtained from all participants prior to the experiments, and all experiments were approved by the respective local Ethics Committees.

Visual Stimulation

In all experiments, we used stimuli on the basis of the stimuli used by Lamme, Van Dijk, and Spekreijse (1992),

who used a stimulus in which orientation of line segments was modulated in such a way that a texture checkerboard appeared and disappeared from a homogeneous texture background. However, we increased the spacing between the squares of the checkerboard, thus inducing a strong percept of texture-defined squares against a background. We manipulated the saliency of the texture squares by using two orientation differences between the squares and the background, being 90° or 20° (see Figure 1A).

Activity related to scene segmentation can be isolated by subtracting the homogeneous visually evoked potentials (VEPs) from the checkerboard VEPs (see Figure 1B). However, to isolate these components, all local elements of the textures (homogeneous and checkerboards) need to be balanced, that is, all local orientations should be presented an equal number of times as part of the background and as part of the texture squares in the foreground. In order to achieve this, we used line orientations of 35°, 55°, 125°, and 145°, yielding two pairs of 20° checkerboards (35° foreground and 55° background, and vice versa; 125° foreground and 145° background and vice versa), two pairs of 90° checkerboards (35° foreground and 125° background and vice versa; 55° foreground and 145° background and vice versa), and four homogeneous textures. All types of stimuli were presented an equal number of times.

Stimuli were generated on a PC and displayed on a 19-in. Iiyama monitor, with a refresh rate of 100 Hz and a screen resolution of 1024 by 768 pixels (21.8° by 16° visual arc). Subjects were seated 100 cm from the monitor. Each trial started with a homogeneous texture which was displayed for 300–800 msec. Then, the homogeneous texture was replaced by either another homogeneous texture, a 90° orientation contrast checkerboard or a 20° orientation contrast checkerboard, each with a 33% random probability. The stimuli were always chosen in such a way that all local line elements were replaced by new ones in order to prevent static parts in the display. VEPs were computed for the second textures only. The second texture remained on screen until the subject responded by pressing a button with the dominant hand, after which a new trial started. The size of the individual squares of the checkerboard was 2.5° by 2.5° visual arc. Although this results in relatively long presentation times for the texture stimuli compared to earlier work, we cannot expect that the longer presentation times will have a large effect on the processing of textures (compare e.g., Fahrenfort, Scholte, & Lamme, 2007, who have used very brief presentation times, and Scholte et al., 2008, using long presentation times, yielding very similar results).

During the temporal order judgment task (Experiment 2), white noise was played through speakers with different SOAs with respect to the onset of the second texture, ranging from 200 msec before the onset of the second texture to 200 msec after the onset of the second texture in 20-msec intervals. Noise lasted to the end of the trial
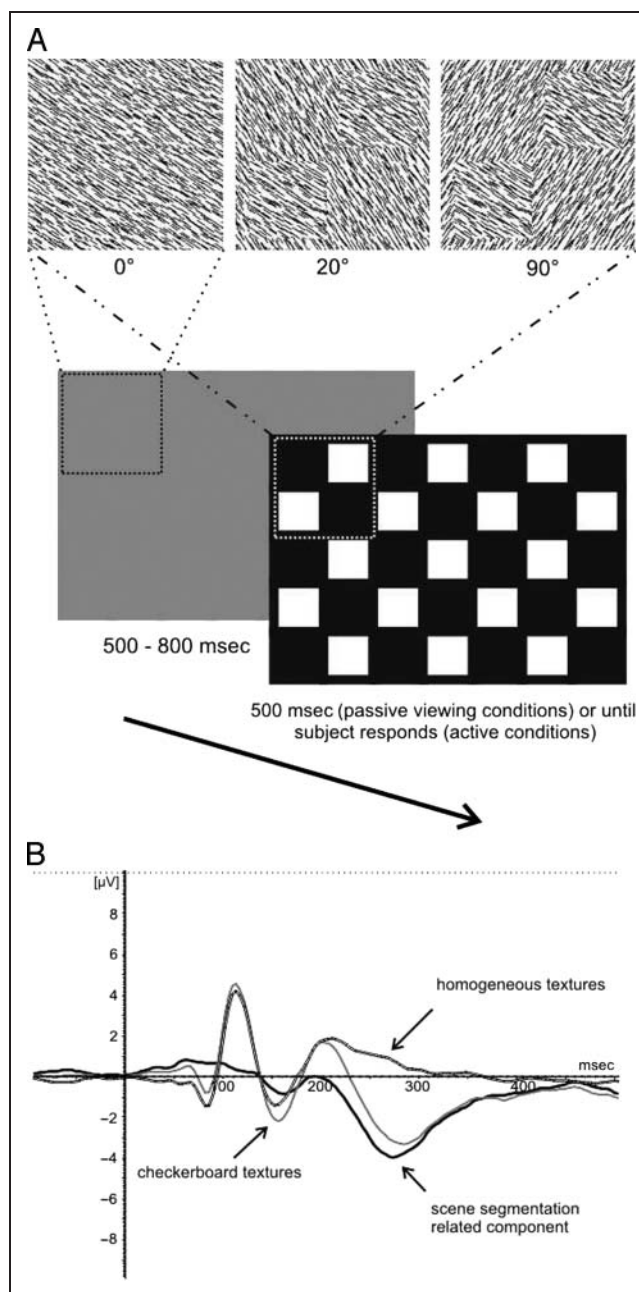

Figure 1. Typical trial run and VEP analysis. (A) Example of a typical trial run. After a homogeneous texture, presented for 300–800 msec, a 20° orientation contrast, a 90° orientation contrast, or a homogeneous texture could appear. Background and foreground elements were always replaced with new elements. In the passive condition, the texture was presented for 500 msec, whereas in the active condition, it was presented until the subject pressed a button. (B) Rationale of the VEP analysis. By averaging all homogeneous orientations (white) and by subtracting this signal from the average of all checkerboard presentations of a given orientation contrast (gray), we could isolate the segregation-specific signal (solid black). Here the activity in channel POz for a 90° orientation contrast texture is shown.

to avoid effects of sound offset (Jaskowski, 1996). The participant's task was to indicate whether the noise started before or after the onset of the second texture, but only when this texture contained a checkerboard. When the

second texture was a homogeneous texture, participants had to withhold their response. The next trial was started after the participant made a response. Before the experimental session, participants were trained to perform at ~100% at the −200 msec and 200 msec SOAs. At least 40 trials per texture–sound SOA per condition (20° or 90° orientation contrast) were collected per participant.

In Experiment 3, texture stimuli could only appear on the left or on the right side of the screen. No fixation dot was present in the target stimuli. The mask used in the experiment was made out of line segments of all orientations used in the experiments, that is, 35°, 55°, 125°, and 145° with a red fixation dot in the middle of the screen.

A typical trial run started with presentation of the mask for 250 msec. Afterward, the target stimulus was shown for 200 msec, followed by the mask again for 1000 msec. Participants were required to make a saccade away from the target as quickly as possible after target presentation, and return fixation to the fixation dot afterward.

Stimuli were presented using the SR Research Experiment Builder (SR Research, Osgoode, Canada) on a 17-in. Iiyama CRT monitor, running at a 100-Hz refresh rate with a resolution of 1024 × 768 pixels. Participants were seated approx. 1 meter away from the screen. The experiment was run in 13 blocks of 160 trials each. Stimulus location and orientation difference were randomized on a trial-to-trial basis. Participants did 2080 trials in total.

EEG Recording and Analysis

EEG was recorded using a 48-channel ActiveOne EEG (Biosemi, The Netherlands) system with active electrodes. Horizontal and vertical EOG was measured to control for eye blinks. The EEG signal was digitized and sampled at 256 Hz on a separate acquisition computer. Markers were sent with the EEG signal by the stimulation computer to allow for off-line segmentation. Per condition, three 10-min blocks were recorded; each block contained 450 trials. EEG data were analyzed using Analyzer (BrainVision Products GmbH, Germany). The raw signal was filtered between 1 and 15 Hz and then segmented in epochs of 1200 msec (200 msec prestimulus–1000 msec poststimulus) on the basis of markers sent with visual stimulation. Bad segments were rejected automatically when the maximum amplitude in a segment exceeded 100 μV. A baseline correction (baseline 20 msec) was performed for each segment. VEPs were computed for homogeneous and checkerboard trials.

To compute the scene segmentation specific signal, VEPs from homogeneous trials were subtracted from checkerboard trial VEPs. The resulting traces were averaged, thus resulting in a VEP representing neural activity related to scene segmentation (see Figure 1B). For both the 90° and the 20° orientation differences, and per subject, peak latencies in this subtraction signal were computed by finding the minimum amplitude in the interval 200–400 msec. Peak latencies for the P100 component

(for both 90 and 20 orientation contrasts) were computed by finding the maximum amplitude in the interval 90–130 msec in the checkerboard VEPs. To test for differences between VEPs in different conditions, we used paired t tests per sample, with an alpha of .05, Bonferroni-corrected for the number of channels and samples compared. We used SPSS version 16.0 for Windows (SPSS, Chicago) to compute Pearson correlations between peak and onset latencies and RTs. All statistics are based on the activity in channel Oz according to the International 10–20 System. This electrode was selected on the basis of previous studies (e.g., Scholte et al., 2008; Caputo & Casco, 1999; Lamme, Van Dijk, & Spekreijse, 1992).

RTs were measured using a push-button and were sampled via the parallel port of the stimulus computer for millisecond-precision analysis. These data were used to compute RTs. With every button press, a marker was sent to the EEG acquisition computer as well. Because the EEG acquisition computer was sampled at 256 Hz, these markers have a 4-msec temporal resolution. These markers were used to segment data based on RT. Participants were instructed to press the button with their dominant hand.

Because the brain does not use VEP averaging and subsequent statistical testing in order to detect the presence of texture stimuli, we ran a single-trial analysis using a nearest mean classifier approach in order to estimate the amount of information needed to detect the presence of a texture stimulus (Bandt, Weymar, Samaga, & Hamm, 2009). For this analysis, we used segmented EEG data from a single channel (Oz), filtered between 0.5 and 100 Hz of the 14 last participants in Experiment 1. This dataset was selected for technical reasons.

Individual trials were normalized so that the mean activity in each trial was zero. Templates for both individual checkerboard textures and the homogeneous textures were computed per participant, and individual trials were classified based on a template function that reflects the difference between checkerboard (C) and homogeneous (H) stimuli (i.e., $C − H$). For each trial x, a classifier score S was obtained by multiplying activity in each sample t with this template function, and taking the mean over the total number of samples T:

$$S(x) = (1/T) \sum (C(t) − H(t))\, x(t)$$

To avoid autocorrelation, we adopted a "leave-one-out" approach for computing the templates: For each trial N, this trial was excluded from the templates used to classify it. To determine the moment at which sufficient information is available to make above-chance perceptual decisions, we systematically increased the number of data points per trial fed into the algorithm. The moment of above-chance level target detection was determined per target type (20° or 90°) by testing when d', averaged over all participants, exceeded 0.

Eye Tracking

Eye movements in Experiment 3 were recorded with a head-mounted eye-tracking system (Eyelink II, SR Research, Osgoode, Canada) at a sample rate of 500 Hz. Saccadic RTs were computed off-line by taking the moment of the first eye movement exceeding a velocity of 30 deg/sec. Any saccades with a saccadic RT smaller than 75 msec were discarded, as were saccades with a saccadic RT greater than mean $\pm 3 \times SD$. Trials in which participants blinked were excluded from the analysis. Saccade direction was classified on the basis of the saccade termination point; subsequently, saccades were classified as antisaccades or prosaccades on the basis of information of the target position.

RESULTS

Experiment 1: Differences in RT Correspond to Latency Differences in Long-latency VEP Components

In the RT experiment (Experiment 1), reducing orientation contrast from 90° to 20° increased RTs from 391.1 to 427.6 msec [$t(32) = 21.4$, $p = .000$; see Figure 3A]. Detection performance was not affected by reducing orientation contrast and was 97% for both conditions.

The subtraction potentials to both the 20° and 90° orientation contrasts showed the characteristic pattern observed in texture segmentation tasks: a small positivity around 80–100 msec, followed by a large negative deflection between 200 and 300 msec, localized in the posterior electrodes. These two timeframes are linked respectively to boundary detection and surface segregation (Scholte et al., 2008; Vandenbroucke, Scholte, van Engeland, Lamme, & Kemner, 2008).

Peak latencies of the scene segmentation-related VEP components increased from 293.4 to 322.8 msec [$t(32) = 7.7$, $p = .000$; see Figure 2A]. Correlations between RTs and latencies of scene segmentation signals of individual subjects are highly significant: .51 ($p = .002$) and .68 ($p = .000$) for the 20° and 90° orientation contrast checkerboard textures, respectively (Figure 2B).

The relation between manual RT and peak latency of scene segmentation-related VEP components showed up even more explicitly in a reanalysis of the RT/EEG data. We split trials in two groups of the 50% fastest versus 50% slowest trials and recomputed the VEPs for the 20° and 90° checkerboards. Please note that for this analysis we did not compute the segregation specific signal, but only took VEPs for checkerboard stimuli into account, because only in these trials was a manual response given. The checkerboard VEPs were significantly different only from 230 to 300 msec ($p < .05$, corrected for 256 samples), showing that differences in RTs are explained by differences in relatively long-latency processes, and not by differences in earlier cortical processing (Figure 2C).

Finally, a single-trial analysis using a nearest mean classification algorithm (Bandt et al., 2009) revealed that sufficient information is available in unfiltered single trials to reliably detect presence of a texture after 113 msec for the 90° targets [$t(13) = 2.38$, $p < .05$], and after 117 msec for the 20° targets [$t(13) = 2.16$, $p < .05$]. Only after 238 msec was the classification performance better for 90° targets than for 20° targets [$t(13) = -2.33$, $p < .05$], suggesting that before that moment, the amount of information present in a single trial does not differ for both types of targets (Figure 3).

These results show a clear relation between scene segmentation related components in the VEP and visually guided behavior: Reducing orientation contrast had similar effects on the latency of segregation-specific VEP components and on RTs. The analysis of slow versus fast trials shows that fast trials are accompanied by a latency decrease of scene segmentation related VEP components, whereas slow trials are characterized by an increased latency of scene segmentation related VEP components. In all experiments and analyses, earlier components of the VEP remained unaffected: Latencies of the P100 did not show any relation with RT, thus suggesting that in texture discrimination tasks, visually guided behavior depends on processes taking place approximately 175–275 msec after stimulus presentation. As argued earlier, these processes reflect scene segmentation, a process critically dependent on recurrent interactions between higher and lower visual areas, and possibly related to visual awareness (Scholte et al., 2008; Heinen et al., 2005). As previous work has shown that scene segmentation does not depend on attention, it is unlikely that changes in attentional capture can explain the pattern of results we observed (Scholte, Witteveen, Spekreijse, & Lamme, 2006; Schubö, Meinecke, & Schröger, 2001).

Experiment 2: Differences in Subjective Simultaneity Correspond to Differences in VEP Latency

Points of subjective simultaneity (PSSs) for the onset of a sound and the onset of checkerboards with 20° and 90° orientation contrast were determined in Experiment 2. Differences in PSSs are believed to reflect differences in perceptual latency (see Jaskowski, 1996). If the long-latency VEP components we report in Experiments 1 do reflect the perceived segmentation of a texture stimulus, one would expect that comparing results from the 20° and 90° textures would yield comparable effects on the VEP components and PSSs.

We found a difference of 18 msec between PSSs, suggesting that a 20° orientation contrast checkerboard is perceived 18 msec later than a 90° orientation contrast checkerboard (Figure 4). This difference was only marginally significant, however; Monte Carlo analyses yielded a p value of .09. Post hoc analyses showed that four of nine individual subjects did show a significant difference of

Figure 2. Differences in
RT correspond to differences
in visual processing speed.
(A) RT distributions and
subtraction VEPs (channel Oz)
for the 20° (gray lines) and 90°
(black lines) checkerboards.
Population averages are
indicated by dotted arrows.
Differences in mean RTs and
mean peak latencies are almost
identical. (B) Correlations
between RTs and VEP latencies.
Latency of scene segmentation
related VEP components of
each subject is plotted against
manual RT. Circles show
individual data points, lines
show best linear fit. (C) VEPs
for the 50% fastest (black
trace) and 50% slowest (gray
trace) trials, 90° orientation
contrast textures. Channel
Oz is shown. Black arrow
indicates onset of the significant
difference (p < .05, corrected
for 256 samples). Earlier
components do not differ,
whereas later VEP components,
related to scene segmentation,
are slower in "slow" trials
than in "fast" trials.
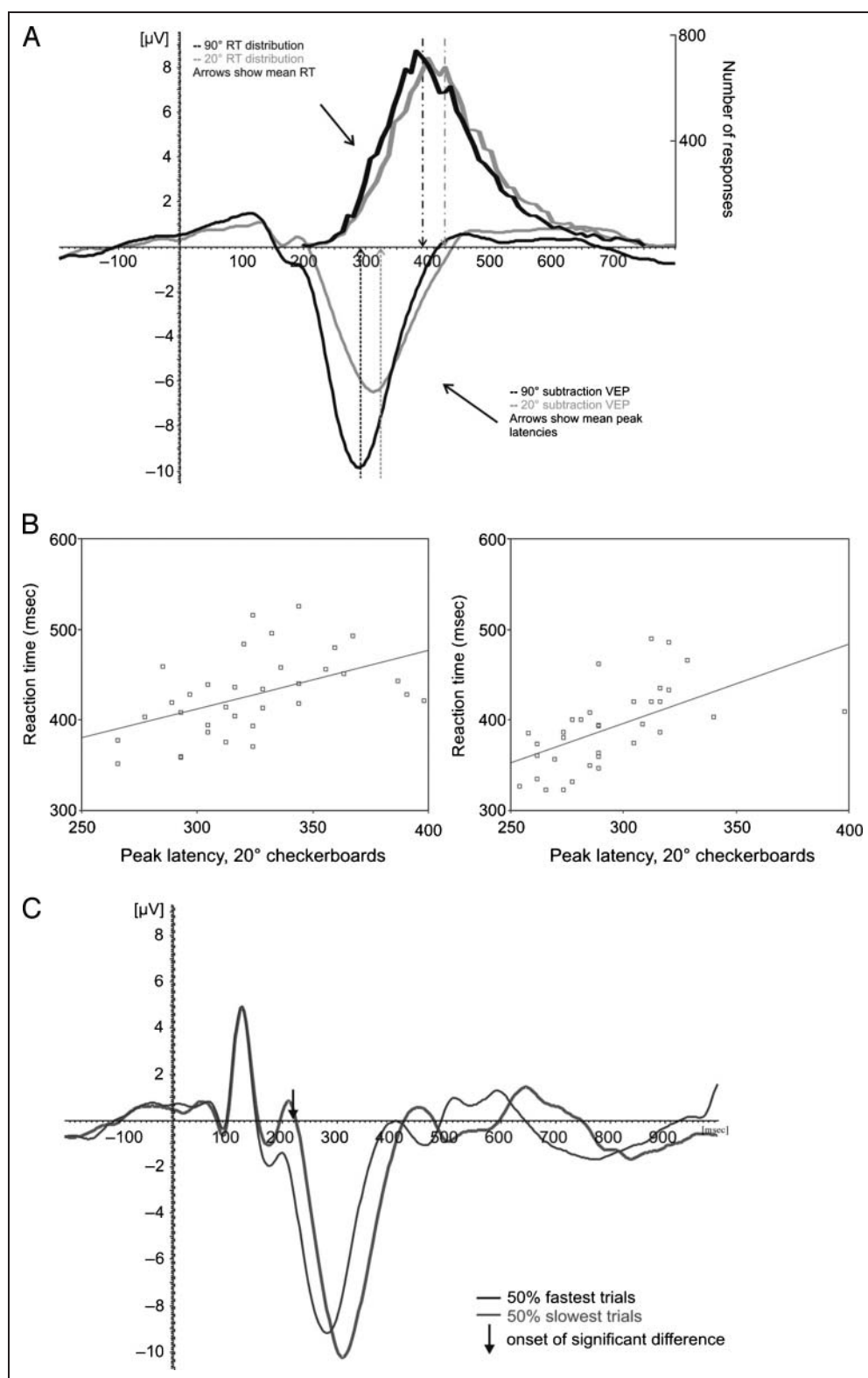
around 30 msec in PSSs; data of two subjects did not show
any difference; the data of the remaining two subjects
were too noisy to reliably estimate the PSS. The differ-
ences in PSSs (ca. 18 msec) we report here correspond well
to the differences in latencies in scene segmentation re-

lated components we report in Experiment 1 (24 msec),
suggesting that differences in PSSs between 20° and 90°
orientation contrast checkerboards are to be attributed
to differences in latencies of scene segmentation related
processing.

Experiment 3: Reflexive Behavior May Be Triggered by Early Visual Processing

In the antisaccade experiment, participants failed to inhibit the tendency to make an eye movement toward the stimulus in 12.7% of the trials, and made a saccade to the target stimulus. We computed RTs for antisaccades and erroneous prosaccades for the 20° and 90° orientation difference stimuli. Antisaccades based on 90° orientation contrast textures (mean RT 364.39 msec) were 27.19 msec faster than on 20° orientation contrast textures (mean RT = 391.58 msec) [$t(10)$ = 8.37, p = .000], a difference strikingly similar to the latency difference in long-latency texture processing related evoked potential components for these stimuli (approx. 30 msec) reported in Experiment 1.

For the prosaccades, subjects were 47.91 msec faster in making a saccade toward the 90° orientation contrast stimuli (mean RT = 327.24 msec) than to the 20° orientation contrast stimuli (mean RT = 375.14 msec) [$t(10)$ = 4.12, p = .000]. However, the distribution of RTs (Figure 5A) reveals that the RT distribution of erroneous saccades toward 20° orientation contrast stimuli is, in fact, bimodal, and that the peaks of the RT distributions of erroneous saccades toward 20° and 90° orientation contrast stimuli overlap. Indeed, paired t tests reveal no significant differences between the RT distributions (25 msec bins) up to 325 msec, but between 350 and 400 msec significantly more prosaccades are made toward 20° targets than to 90° targets. Apparently, there are two types of erroneous saccades—one generated on the basis

Figure 3. Single-trial classification analysis for 20° and 90° orientation difference textures. (A) Classification performance of the nearest mean classifier algorithm for 20° (gray line) and 90° textures (black line) based on a single-trial analysis of segmented EEG data expressed in d' units. Vertical lines indicate onset of above-chance discrimination for 20° and 90° textures (gray and black lines, respectively). (B) Template functions (i.e., normalized VEPs) used for the classification algorithm for 20° textures (gray line), 90° textures (black line), and homogeneous textures (dotted line). Vertical lines again indicate onset of above-chance discrimination for targets.
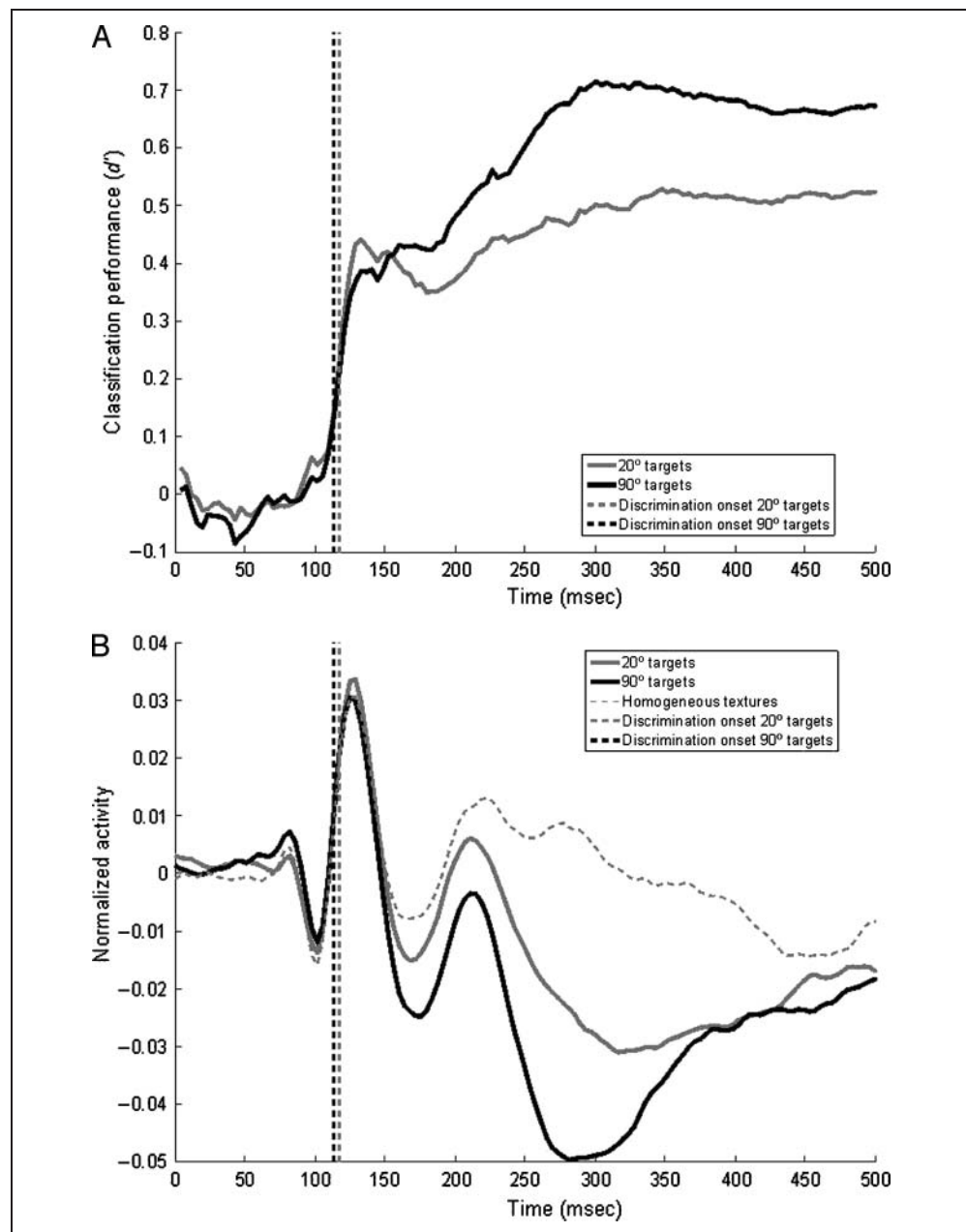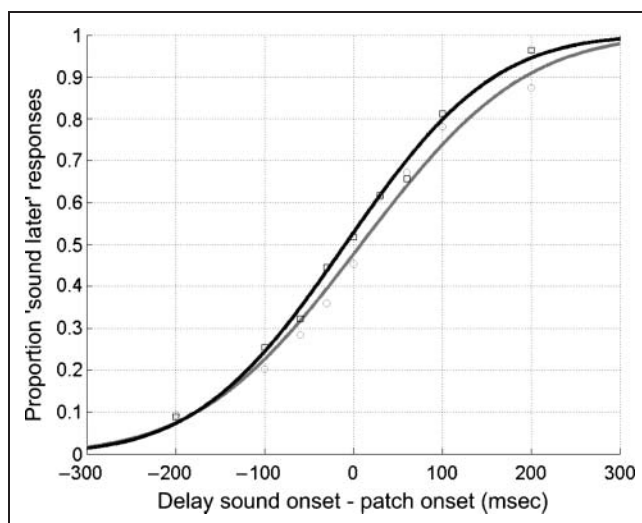
Figure 4. Differences in PSSs for 20° and 90° orientation differences. Black line (probit fit) and symbols show the proportion of "sound later" responses for 90° checkerboards; gray line (probit fit) and symbols show the proportion of "sound later" responses for 20° checkerboards.

of an early signal, and a second, smaller subset, based on a later signal.

The distribution of antisaccade RTs, on the other hand, shows two distinct peaks around 350–400 msec for antisaccades based on 90° and 20° orientation contrast stimuli, with significantly more antisaccades away from 20° stimuli (gray area in Figure 2A; $p < .01$) for bins later than 400 msec, suggesting that these saccades are based on a relatively long-latency visual signal. The different RT patterns for antisaccades and prosaccades show that they are based on different visual representations. The absence of a difference in peak RTs for 20° and 90° textures suggests that most prosaccades are based on detection of borders, around 100 msec. This is corroborated by the fact that most prosaccades to the texture stimuli are made toward the texture squares closest to fixation, which would have captured attention and evoked eye movements (Figure 5B).

On the other hand, the striking similarity between the latency difference in the peak latencies for antisaccades toward 20° and 90° stimuli and the VEP components reflecting filling-in of 20° and 90° stimuli strongly suggests that antisaccades, and a small portion of erroneous prosaccades is based on the representation of a filled-in figure. Apparently, making an antisaccade requires more than simply inhibiting a prosaccade and reprogramming the eye movement: Antisaccades actually require deeper visual processing of the stimulus on which they are based than prosaccades.

DISCUSSION

In three experiments, we have investigated the relation between behavior, perception, and neural processing using texture stimuli. Both simple RTs and PSSs increase by al-

most the same amount of time as the latency of long-latency VEP components when the perceived strength of segmentation is decreased, strongly suggesting that both responding to and conscious perception of a texture require the processes underlying these VEP components. However, reflexive behavior seems to be based on earlier stages of information processing, as indicated by the absence of a difference in RT for erroneous prosaccades to 20° and 90° targets. A single-trial analysis of EEG data shows that sufficient information is indeed processed approx. 130 msec after stimulus onset to reliably detect the presence of a target. Summarizing, we can state that long-latency visual processes, linked to visual awareness, do play an important role in detecting and responding to visual events in our study.

Interestingly, this suggests that instead of accumulating evidence from the earliest moment of sensory processing, neural decision-making systems seem to rely on visual representations that are formed at longer latencies. In the literature, there is growing support for the accumulator model of perceptual decision-making (see, e.g., Ratcliff & McKoon, 2008, for a recent review). In the accumulator model, sensory evidence is accumulated over time until an internal threshold is reached and a motor response is initiated. A recent neuroimaging study in humans demonstrated that a network of occipital, parietal, and prefrontal areas is involved in this process. Critically, neurons in the intraparietal sulcus seem to function as accumulators of perceptual evidence, but exactly what feeds into this accumulating system, in terms of visual information, remains unspecified (Kayser, Buchsbaum, Erickson, & D'Esposito, 2010). Our experiments seem to suggest that only visual information that has been processed quite extensively feeds into the decision-making process. However, this seems to contrast with earlier findings, suggesting that perceptual decisions can be made extremely fast. How can we explain this discrepancy?

As argued, a large amount of information appears to be extracted from visual input during the earliest stages of visual processing (see e.g., Thorpe et al., 1996). Recent studies suggest that visual information processing underlying rapid image classification is based on the statistical properties of visual input: The majority of variability in EEG responses up to approx. 100 msec after stimulus presentation can be explained by the distribution of contrast values of an image. Interestingly, this underlying statistical structure contains a large amount of ecologically meaningful information about the general lay-out of a visual scene—for example, whether a scene is cluttered, or contains just one single object. It has been proposed that the statistical structure of visual input provides the system with sufficient information to do coarse categorization tasks. However, if such statistical structure is absent, deeper processing is required for perceptual decision-making (Scholte, Ghebreab, Waldorp, Smeulders, & Lamme, 2009; VanRullen & Koch, 2003; Lamme & Roelfsema, 2000; Thorpe et al., 1996).

A similar conceptual distinction between coarse categorization versus more precise stimulus identification has

been suggested in the context of feedforward versus recurrent processing in object-sensitive visual areas. Neurons in temporal cortex, for example, respond to faces, in general, 150 msec after stimulus presentation; however, approximately 50 msec later, the same neurons respond to individual faces (Sugase, Yamane, Ueno, & Kawano, 1999). Neural network models of texture segregation tasks show a similar behavior: Global discriminations between shapes can be done on the basis of a single feedforward sweep from lower to higher layers in the network, analogous to the feedforward transfer of visual information from lower-tier to higher visual areas. However, in order to allow for finer discriminations, for example, between two exemplars of the same category, feedback processing from higher to lower visual areas seems to be required in order to obtain sufficient spatial detail (Jehee et al., 2007; Hochstein & Ahissar, 2002). In addition, feedback processing may also provide a mechanism to highlight information that is relevant for a certain task (Roelfsema, Lamme, & Spekreijse, 1998).

In general, we could state that visual information processing may be characterized as a two-stage process: During the first stage, coarse information based on low-level features and image statistics is extracted. This information may be sufficient to trigger perceptual decisions, depending on the task at hand. However, in order to do finer

Figure 5. (A) Averaged distribution of pro- and antisaccadic RTs. *y*-axis gives the number of responses given in a given 20-msec bin, normalized per participant. Dotted lines show prosaccades, solid lines show antisaccades; gray lines show responses to 20° orientation contrast stimuli, black lines show responses to 90° orientation contrast stimuli. Shaded area denotes a significant difference between the number of responses in a given bin for 20° vs. 90° stimuli, *p* < .05 (corrected for multiple comparisons). (B) Averaged distribution of saccade termination points. *y*-axis gives the number of responses in a given 40 pixel bin, collapsed over both leftward and rightward saccades, *x*-axis gives horizontal position. Further legend as in (A). Shaded areas indicate horizontal position of texture squares.
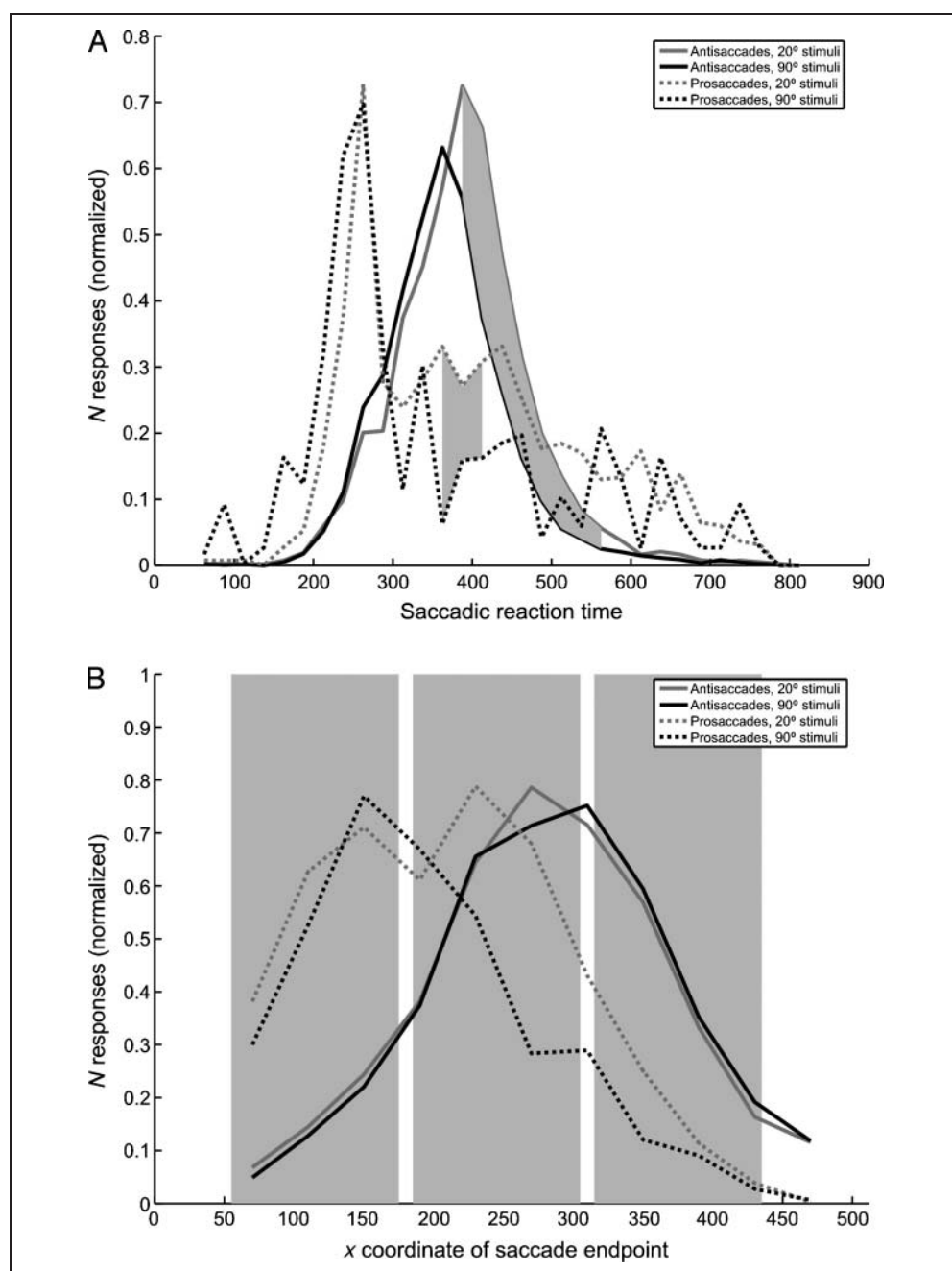
discrimination, such as within-category discriminations, additional processing is required, and perceptual decision-making will require more time (VanRullen, Reddy, & Koch, 2004; Hochstein & Ahissar, 2002). Although this model seems plausible, and may be able to explain why in some tasks perceptual decisions are based on coarse representations, it does not completely explain the behavior of participants in our experiments. In our study, we may assume that information extracted up to approximately 130 msec after stimulus presentation should be sufficient to trigger a behavioral response (Experiment 1), and indeed, we find that reflexive saccades may, in fact, be triggered by this early cortical processing (Experiment 3). However, initiation of a simple button press, or an antisaccade, appears to be based on longer-latency visual processing, despite the fact that coarse information is present in the system.

We theorize that the selection of what information feeds into the perceptual decision-making systems is not simply a matter of using *any* information that is available. Instead, there seems to be a preference for using information that represents the visual environment with high spatial detail when such information, in general, is available (see also Ratcliff & Smith, 2010; Jolij, 2008; Jolij & Lamme, 2005). In our experiments, all stimuli were presented long enough to be clearly seen by the participants. Although participants were instructed to respond as fast as possible in the RT and antisaccade tasks, we did not press for faster responses by penalizing slow responses. Therefore, it is very well possible that participants employed a strategy in which they waited until they consciously perceived the target, and then pressed the button or made an eye movement. Given the very good performance on both tasks, this is quite likely. However, in the studies by VanRullen et al. (2004), briefly presented targets were used, which in some trials were even masked. Therefore, overall visibility was much lower. In fact, they even report that some participants in their study could reliably detect targets in the absence of visual awareness, resembling a blindsight-like condition. We propose that the poor overall quality of visual information may have triggered the perceptual decision network to start accumulating evidence as soon as sensory information became available, instead of waiting for a more precise representation that does not become available as a result of masking (Fahrenfort et al., 2007; Lamme, Zipser, & Spekreijse, 2002).

In a study on blindsight in normal participants using TMS, Jolij and Lamme (2005) report a very similar contrast between availability of unconscious (or preconscious) information for perceptual decision-making. In this study, TMS of primary visual cortex was used to induce artificial blindsight in normal observers. Interestingly, subjects showed the blindsight capability only in a context in which the stimuli were generally hard to see, and the subjects were mostly guessing even without magnetic stimulation. When stimulus visibility was increased, and subjects could base their response on their conscious percept, the blind-

sight capability disappeared, probably because they ignored unconsciously processed information. This suggests that, in a context where conscious representations are available, these are preferred over unconscious representations to initiate a response, most likely due to the fact that conscious representations have a higher accuracy (Jolij, 2008).

Interestingly, not all visually guided behavior seems to be governed by such a mechanism. In Experiment 3, we show that reflexive prosaccades to texture targets are most likely to be based on an early border detection signal, and not on a longer-latency signal, whereas antisaccades are based on long-latency visual processes. This fits with the idea that task demands also play a role in selecting what information to use in perceptual decision-making. Reflexive saccades are evoked by suddenly appearing targets that may be highly relevant for an organism; the motor program to execute such a saccade can be programmed and executed quickly. A false alarm in such a case may not be too harmful: An eye movement has no consequences in the "external world."

However, the relative harmlessness of an erroneous eye movement may not be sufficient to explain the pattern of results concerning the effects of unconscious visual representations on visually guided behavior observed in the literature. In several studies, Schmidt and Schmidt (2009) and Schmidt, Niehaus, and Nagel (2006) investigated the effects of unconsciously (or preconsciously) processed information on pointing movements: They found that the initial trajectory of a pointing movement is guided by feedforward-processed visual information. Interestingly, both the results obtained by Schmidt and coworkers, and the results we obtained in Experiment 3 share one critical feature: Both prosaccades and pointing movements do not require a spatial transformation or remapping of a response, and are completely stimulus-driven.

A manual response on a response box, or even more complex, cognitively guided behavior, such as antisaccades on the other hand, requires such a spatial transformation. Pressing a button on a response box when you see a stimulus does require a specific stimulus–response mapping that bears no direct relation with the stimulus, contrary to a prosaccade or a pointing movement. It is well possible that such a response remapping requires cognitive effort and deeper visual processing of a stimulus to avoid errors that may induce slowing in the cognitive control system (Ridderinkhof, Van Den Wildenberg, Segalowitz, & Carter, 2004). However, this hypothesis needs to be experimentally verified.

Summarizing, we believe that our results are in line with the hypothesis that the normal modus operandi of human observers is to base a visually guided response on a more elaborated, and most likely conscious, representation of the environment, even though response generation on the basis of quickly formed unconscious representations is very well possible. However, these latter representations are only used to guide behavior when necessary, for example, when a fast response is required (e.g., in fight–flight

situations or in reflexive saccades), or when a conscious representation is not available as a result of a lesion, or in experimental manipulations such as TMS or masking.

Acknowledgments

REFERENCES

Bandt, C., Weymar, M., Samaga, D., & Hamm, A. O. (2009). A simple classification tool for single-trial analysis of ERP components. *Psychophysiology, 46,* 747–757.

Boyer, J. L., Harrison, S., & Ro, T. (2005). Unconscious processing of orientation and color without primary visual cortex. *Proceedings of the National Academy of Sciences, U.S.A., 102,* 16875–16879.

Caputo, G., & Casco, C. (1999). A visual evoked potential correlate of global figure–ground segregation. *Vision Research, 39,* 1597–1610.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2007). Masking disrupts reentrant processing in human visual cortex. *Journal of Cognitive Neuroscience, 19,* 1488–1497.

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

Heinen, K., Jolij, J., & Lamme, V. A. F. (2005). Figure–ground segregation requires two distinct periods of activity in V1: A transcranial magnetic stimulation study. *NeuroReport, 16,* 1483–1487.

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

Jaskowski, P. (1996). Simple reaction time and perception of temporal order: Dissociations and hypotheses. *Perceptual and Motor Skills, 82,* 707–730.

Jehee, J. F., Lamme, V. A. F., & Roelfsema, P. R. (2007). Boundary assignment in a recurrent network architecture. *Vision Research, 47,* 1153–1165.

Jehee, J. F., Roelfsema, P. R., Deco, G., Murre, J. M., & Lamme, V. A. F. (2007). Interactions between higher and lower visual areas improve shape selectivity of higher level neurons-explaining crowding phenomena. *Brain Research, 157,* 167–176.

Jolij, J. (2008). From affective blindsight to affective blindness: When cortical processing suppresses subcortical information. In F. Columbus (Ed.), *Neural pathways: New research* (pp. 205–208). New York, NY: Nova Science Publishers.

Jolij, J., Huisman, D., Scholte, H. S., Hamel, R., Kemner, C., & Lamme, V. A. F. (2007). Processing speed in recurrent networks correlates with general intelligence. *NeuroReport, 18,* 39–43.

Jolij, J., & Lamme, V. A. F. (2005). Repression of unconscious information by conscious processing: Evidence from affective blindsight induced by transcranial magnetic stimulation. *Proceedings of the National Academy of Sciences, U.S.A., 102,* 10747–10751.

Kayser, A. S., Buchsbaum, B. R., Erickson, D. T., & D'Esposito, M. (2010). The functional anatomy of a perceptual decision in the human brain. *Journal of Neurophysiology, 103,* 1179–1194.

Lamme, V. A. F. (1995). The neurophysiology of figure–ground segregation in primary visual cortex. *Journal of Neuroscience, 15,* 1605–1615.

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences, 7,* 12–18.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The different modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23,* 571–579.

Lamme, V. A. F., Van Dijk, B. W., & Spekreijse, H. (1992). Texture segregation is processed by primary visual cortex in man and monkey: Evidence from VEP experiments. *Vision Research, 32,* 797–807.

Lamme, V. A. F., Zipser, K., & Spekreijse, H. (2002). Masking interrupts figure-ground signals in V1. *Journal of Cognitive Neuroscience, 14,* 1044–1053.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20,* 873–922.

Ratcliff, R., & Smith, P. M. (2010). Perceptual discrimination in static and dynamic noise: The temporal relation between perceptual encoding and decision making. *Journal of Experimental Psychology: General, 139,* 70–94.

Ridderinkhof, K. R., Van Den Wildenberg, W. P., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition, 56,* 129–140.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature, 395,* 376–381.

Roelfsema, P. R., Lamme, V. A. F., Spekreijse, H., & Bosch, H. (2002). Figure–ground segregation in a recurrent network architecture. *Journal of Cognitive Neuroscience, 14,* 525–537.

Scharnowski, F., Rüter, J., Jolij, J., Hermens, F., Kammer, T., & Herzog, M. H. (2009). Long-lasting modulation of feature integration by transcranial magnetic stimulation. *Journal of Vision, 9,* 1–10.

Schmidt, T., Niehaus, S., & Nagel, A. (2006). Primes and targets in rapid chases: Tracing sequential waves of motor activation. *Behavioral Neuroscience, 120,* 1005–1016.

Schmidt, T., & Schmidt, F. (2009). Processing of natural images is feedforward: A simple behavioral test. *Attention, Perception, and Psychophysics, 71,* 594–606.

Scholte, H. S., Ghebreab, S., Waldorp, L., Smeulders, A. W., & Lamme, V. A. F. (2009). Brain responses strongly correlate with Weibull image statistics when processing natural images. *Journal of Vision, 9,* 1–15.

Scholte, H. S., Jolij, J., Fahrenfort, J. J., & Lamme, V. A. F. (2008). Feedforward and recurrent processing in scene segmentation: Electroencephalography and functional magnetic resonance imaging. *Journal of Cognitive Neuroscience, 20,* 2097–2109.

Scholte, H. S., Witteveen, S. C., Spekreijse, H., & Lamme, V. A. F. (2006). The influence of inattention on the neural correlates of scene segmentation. *Brain Research, 1076,* 106–115.

Schubö, A., Meinecke, C., & Schröger, E. (2001). Automaticity and attention: Investigating automatic processing in texture segmentation with event-related brain potentials. *Brain Research, Cognitive Brain Research, 11,* 341–361.

Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature, 400,* 869–873.

Supèr, H., Lamme, V. A. F., & Spekreijse, H. (2001). Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nature Neuroscience, 4,* 304–310.

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

Vandenbroucke, M. W., Scholte, H. S., van Engeland, H., Lamme, V. A. F., & Kemner, C. (2008). A neural substrate for atypical low-level visual processing in autism spectrum disorder. *Brain, 131(Pt 4),* 1013–1024.

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

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

Weiskrantz, L. (1996). Blindsight revisited. *Current Opinion in Neurobiology, 6,* 215–220.