

# The earliest EEG signatures of object recognition in a cued-target task are postsensory

Jeffrey S. Johnson

Center For Neuroscience, UC Davis, Davis, CA, USA



Bruno A. Olshausen

Center For Neuroscience, UC Davis, Davis, CA, USA, &  
Redwood Neuroscience Institute, Menlo Park, CA, USA



Recent experiments have demonstrated early target minus nontarget differences in the human event-related potential (ERP) during visual object recognition tasks. It is unclear whether these differences reflect high-level visual processes, effectively indexing the speed of object recognition, or whether they arise from postsensory decision processes, leaving the actual time of object recognition uncertain. Here we report three sets of ERP experiments designed to determine what processes underlie the target minus nontarget difference signals seen in visual cued-target paradigms. We demonstrate that the same difference signals are present when the target match is made to word stimuli as well as to object stimuli, suggesting that the disparate mechanisms involved in letter string and object processing are not directly responsible for the signals. We also find that the amplitude of these signals can be reduced by increasing trial difficulty in three different ways: image difficulty, level of semantic categorization, and overall task demands. In many respects, the difference signal is similar to the postrecognition P300. Together, these results suggest that the target minus nontarget difference does not reflect object recognition per se, but rather postsensory decision processes.

**Keywords:** object recognition, visual cortex, electrophysiology, ERP, ERPimage, P300

## Introduction

Theorists and modelers of the visual system have long been impressed with the relative speed with which it appears to perform the computationally difficult task of scene analysis. Images that are only briefly presented can nonetheless evoke selective neural activity (Keyser, Xiao, Földiák, & Perrett, 2001; Grill-Spector, Kushnir, Hendler, & Malach, 2000) or even recognition itself (Intraub, 1999; Breitmeyer, 1984). Extracellular spike recordings in macaque (Nowak & Bullier, 1997) and intracranial electroencephalogram (EEG) recordings in humans (Allison, Puce, Spencer, & McCarthy, 1999) confirm that the response latencies in high-level visual cortex are remarkably low – as fast as 130-140 ms in the facial recognition areas of human fusiform gyrus. But these latencies alone may not be able to tell us how quickly the visual system completes its job, because the onset of activity in high-level visual areas does not of itself necessarily imply sufficient stimulus processing to decide on the identity of objects in a scene.

Another recent method used to characterize the speed of visual processing has been the search for target-related differences in the human event-related potential (ERP), particularly during the viewing of natural objects. Differences in the ERP seen after the presentation of objects that vary only in their target status are likely to result from a decision on the identity of those objects. If so, they provide an index of the time when object recognition is sufficiently complete to make that decision. We refer to this time index as a neural signature. In general, a neural signature of a process may result from the process itself, or it may arise

from a subsequent process, and provides evidence that the original process is either ongoing or complete.

The tasks that have been designed in search of neural signatures of object recognition fall into two major classes: single-category tasks, where the subject is given a single target category (e.g., animal) before a long block of images, and cued-target tasks, where the subject is given a new target category before each image. The primary purpose of a cued-target task is to ensure that, across subjects, the same stimuli are presented as both targets and nontargets to eliminate potential low-level statistical differences between them; in other words, the task is balanced. Single-category tasks may also be balanced by alternating cue blocks. Under balanced conditions, both single-category tasks (VanRullen & Thorpe, 2001b) and cued-target tasks (Johnson & Olshausen, 2003) have confirmed the existence of early target-related signals whose onset (at least in the cued-target case) correlates with the subject's subsequent reaction time. On the fastest trials, these differences arise as early as 150 ms, but on slower trials differences are not in evidence until 300 ms or more after presentation of the stimulus.

Although the above studies described the time course of the signal onset of a neural signature of object recognition, they were unable to ascertain the underlying source of the differences. Motor processes associated with the response were ruled out as possible sources of the target-related differences (Johnson & Olshausen, 2003), but it could not be determined whether the signals were due to facilitated sensory processing (i.e., enhanced neural responses in the course of visual processing due to a match with a top-down target template) or to postsensory process-

ing (e.g., a correlate of target status on a given trial independent of task or modality). Because these signals were identified as an *upper bound* for the time required to perform object recognition, the distinction is quite relevant to computational models of object recognition. Some models propose that recognition may be accomplished in a purely feedforward manner (Fukushima, 1980; Mel, 1997; Riesenhuber & Poggio, 1999; VanRullen & Thorpe, 2001a; VanRullen & Thorpe, 2002), whereas others stipulate that it relies on the circulation of activity in recurrent circuits (top-down feedback) (Mumford, 1994; Ullman, 1996; Lewicki & Sejnowski, 1997; Rao & Ballard, 1999; Lee & Mumford, 2003). If the target-related differences correspond to facilitated visual processing, object recognition need not begin sooner than 150 ms after presentation, but if they correspond to higher level decision processes, object recognition must in some cases be reasonably *complete* by then, reducing the amount of time the visual system has available to devote to feedback and recurrent processing in the recognition stream.

Although attempts to use the first ERP difference between targets and nontargets as a proxy for the time of object recognition are fairly recent, target-related ERP signals themselves are nothing new. The discovery of one such signal, the P300, in the mid 1960s (Sutton, Braren, Zubin, & John, 1965; Picton, 1992) led to a fruitful period of work in the area in subsequent decades. P300s, which can be elicited in the auditory, visual, or even somatosensory and olfactory modalities (Soltani & Knight, 2000), are typically evoked by inserting rare target stimuli, which are to be counted or responded to, within a sequence of frequent, identical nontarget stimuli, which are to be ignored. A target-related subcomponent of the P300, also known as the P3b (Squires, Squires, & Hillyard, 1975), manifests as a large (often 20  $\mu$ V) centroparietal positivity on target trials with a peak latency that lies between 300-600 ms after presentation of the target stimulus. The amplitude of the P3b is variable, and difficult or degraded stimuli are known to decrease the amplitude of the signal (Verleger, 1988). Although there is disagreement on the issue (Donchin & Coles, 1988), several reports suggest that the latency of the P3b is correlated on a trial-by-trial basis with the subject's reaction time (e.g., Ritter, Simson, & Vaughan, Jr., 1972; Kutas, McCarthy, & Donchin, 1977). The P3b is generally accepted to be a postsensory signal; some theorists believe that the P3b is related to stimulus evaluation (Verleger, 1988), whereas others, partly based on the fact that reaction times often precede the peak latency, have suggested that the P3b is related to updating working memory (Donchin & Coles, 1988).

Is there reason to believe that the target-related differences found in recent object recognition experiments are related to the long-studied P300? Although the scalp topography and time course of target-related signals found in cued-target experiments bear initial resemblance to P300 activity (Johnson & Olshausen, 2003), the same can be said

of other, slightly earlier components in single-category experiments (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Thorpe, Fize, & Marlot, 1996) that appear to arise from sensory processing. Because the methods used to evoke cued-target differences and the techniques used to analyze them differ slightly from those for the traditional P300, it is important to more fully characterize these signals before concluding that they are, in effect, postsensory processing rather than a reflection of advanced visual processing.

In this work, we introduce the critical manipulation of reversing the cued-target experiment, which forces the target decision to be made in response to a test word and removes it temporally from the object processing, allowing us to show that target minus nontarget differences are not due to object-level processing. Further experiments assess the sensitivity of these target-related differences to changes in the difficulty as well as the visual and recognition demands of the task. The results allow us to assert more confidently that these neural signatures of object recognition are another reflection of P300 activity. Because they are postsensory rather than a reflection of object recognition itself, their onset marks only a loose upper bound for the time required to perform sufficient visual processing to decide on the identity of an object.

## Methods

### Participants

A total of 36 adult subjects (15 males and 21 females, aged 18 to 31 years, average 20.4 years) participated in the three experiments reported in this study, 12 subjects in each experiment. All participants had normal or corrected-to-normal spatial vision. One subject in [Experiment 1](#) self-reported a "slight" colorblindness, but exhibited normal behavioral performance. All participants gave informed consent and the UC Davis Human Subjects Institutional Review Board approved all studies.

### Stimuli

Sample images used in the three experiments are shown in [Figure 1](#). All images consisted of a digitally centered cutout photograph of an object ([Hemera Photo-Objects](#)), a solid colored background, and nine ([Experiment 1](#)) or eight ([Experiments 2 and 3](#)) solid colored ovals. There were three types of images, intact (full object in front of ovals), occluded (object behind ovals), and deleted (partial object in front of ovals). In the occluded and deleted images, 60% of the pixels from the original cutout objects were either covered by ovals (occluded) or removed in the shape of an oval (deleted). Images were created by computer; click [here](#) for a full description of the image creation process. Full size samples for [Experiment 1](#) and [Experiments 2 and 3](#) can be found on the web.

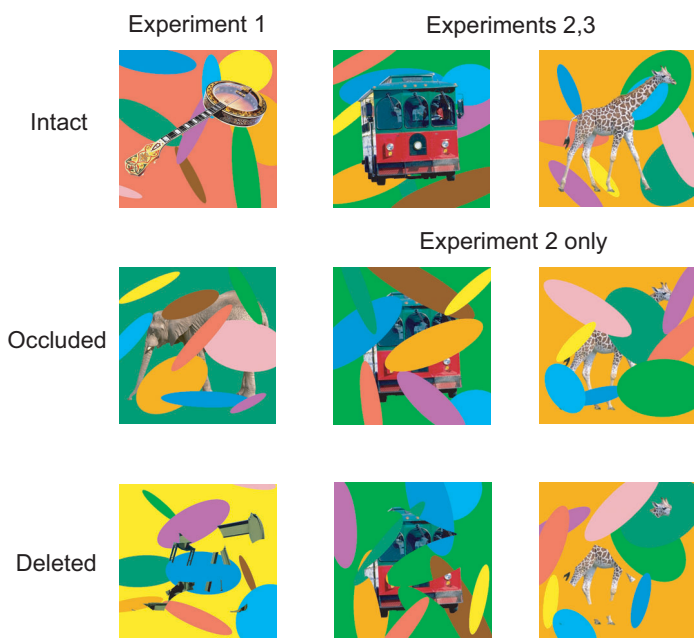


Figure 1. Sample images. All images subtended 15 x 15° of visual angle. In occluded and deleted images, the central object was missing 60% of pixel information. In Experiment 1, each individual object was assigned to one of the three types (intact, occluded, or deleted). In Experiments 2 and 3, each object was created in all three types. The missing pixels were in the same locations in the occluded and deleted images. Each subject saw each object only once. Images from Experiment 1 and images from Experiments 2 and 3 differed slightly in technical detail. Hyperlinks lead to web pages with full-sized examples and image creation details.

### Experimental procedures

Two main types of tasks were used in this study. In the Forward task, subjects were presented a target cue in the form of a word, which would remain onscreen until they pressed a button on a gamepad to initiate the trial. Following a short, randomized delay, the subject was presented a brief image and asked to respond yes/no, as quickly as possible, whether the object in the image matched the previous target cue. In the Reverse tasks, subjects were presented with a target cue in the form of an image, then asked to respond yes/no, as quickly as possible, whether the subsequent word matched the previous target cue.

All images were centrally presented on a CRT monitor. Viewing distance was 75 cm and all images subtended 15 x 15° of visual angle. Image presentation was controlled by a PC running the Presentation software (NeuroBehavioral Systems). Packaged versions of the experiments for Presentation are available via FTP.

### Experiment 1: Forward experiment

As in previous cued-target experiments (Johnson & Olshausen, 2003), subjects were presented with a word cue on each trial, followed by a test image (Figure 2a). The word cue remained on screen until the subjects pressed a button to initiate a trial. Upon pressing the button, the cue was replaced by a central fixation point for 300-900 ms (randomized by computer). The test image would then appear for 150 ms, followed by a 1700-ms fixation period before the next cue appeared. To avoid EEG artifact, subjects were instructed not to blink from the time they initiated the trial until the appearance of the next cue word. Cue words were formulated for entry level categorization (e.g., cat, chair; for discussion on levels of categorization, see Rosch, Mervis,

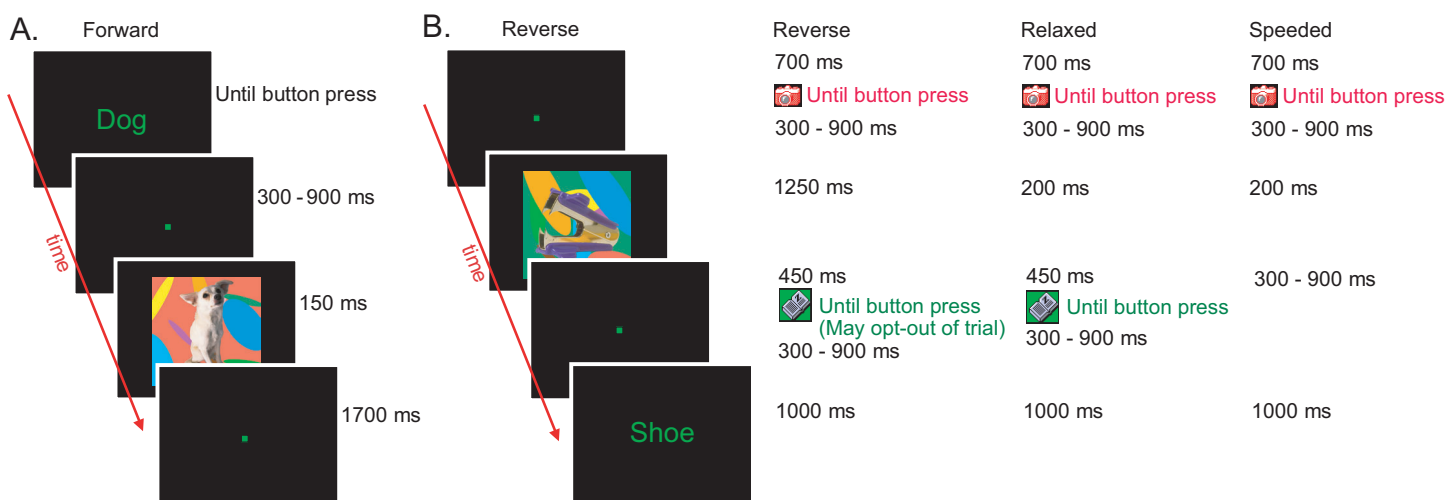


Figure 2. Schematic of task. In the Forward task, participants were given a new target category before each image by means of a word presented on screen. In the Reverse task, participants were given a new target category before each word by means of an image presented on screen. There were three types of Reverse tasks with different timing details, which are presented at right. Small fixation icons indicating the subsequent type of presentation (image and word) were presented in the Reverse task between fixation points and are schematized in the text sequence.

Gray, Johnson, & Boyes-Braem, 1976; Jolicoeur, Gluck, & Kosslyn, 1984). Each object from the original image set was randomly assigned to be intact, occluded, or deleted at the time of image creation. During the experiment, cue words were changed so that the 900 individual images were counterbalanced across subjects for target/nontarget status. Over the course of the experiment, the subject saw 150 images from each target/image condition (e.g., occluded nontarget) in a randomized order. Trial presentation was broken into 9 runs of 100 images each.

## Experiment 2: Reverse experiment

In [Experiment 2](#) ([Figure 2b](#)), the order of images and words was reversed so the images became the target cues and the words became the test stimuli. In this experiment, subjects separately initiated both the cue and test stimuli. At the beginning of a trial, an icon indicating an upcoming cue image was presented, and remained on screen until the subject pressed a button. The icon then disappeared and after a 300-900-ms fixation period, the cue image appeared for 1250 ms. Following another 450-ms fixation period, an icon indicating the upcoming test word appeared. This icon also remained on screen until the subject pressed one of two buttons. If the subjects felt that they could not identify the object, they were instructed to press one button to “opt out” of the trial; the test word was skipped and the icon indicating the next cue item was presented. Otherwise, the subjects pressed a second button to initiate the test phase. Following a 300-900-ms fixation period, the test word appeared on screen for 1000 ms, and after a 700-ms fixation period, the next cue icon appeared. To avoid EEG artifact, subjects were instructed to blink only when a cue or test icon was on screen. Each object from the original image set was made into intact, occluded, and deleted images. In the occluded and deleted versions, the same object pixels were present, only the locations and relative depth of the ovals were changed. Each of the 900 objects was seen once per subject, and objects were counterbalanced for appearance in intact, occluded, and deleted versions across subjects. Over the course of the experiment, the subject saw 150 images from each target/cue condition (e.g., deleted cue, target) in a randomized order. Trial presentation was broken into 9 runs of 100 images each, and test words were formulated for entry-level categorization.

## Experiment 3: Relaxed/Speeded experiment

In [Experiment 3](#) ([Figure 2b](#)), subjects performed three related tasks. One task was an entry-level Relaxed paradigm. This paradigm was Reversed and was designed to be similar to [Experiment 2](#) with three exceptions: (1) Only 250 total images were used, all intact; (2) the duration of the cue image was only 200 ms; and (3) subjects were not given the opportunity to opt out of trials. The second task was a superordinate-level Relaxed paradigm, which differed only in that the test words were presented exclusively at a superordinate level of categorization. Separate blocks for cate-

gorization at the entry and superordinate levels were used to ensure that the subject could prepare for the expected target word with minimal interference from the other level of categorization. Subjects were explicitly asked to have the target test word in mind for both Relaxed paradigms before initiating the test portion of the trial. The third task was a Speeded version of the Reverse scenario, with test words at both the entry and superordinate levels of categorization. In the Speeded paradigm, the test words followed the 500 cue images by 300-900 ms without any intervention by the subjects, and the entry- and superordinate-level test words were intermixed so the subjects did not know at which level the test word would appear. The images were created in the same manner as those in [Experiment 2](#). In all three tasks, each image was counterbalanced independently across subjects for target/nontarget status and entry/superordinate-test level, so each subject viewed 125 instances of each condition in both the Relaxed and Speeded experiments. The Speeded experiment consisted of 5 runs of 100 trials each, and both halves of the Relaxed experiment consisted of 2 runs of 100 trials and one run of 50 trials. Because there was a greater possibility of confusion and a smaller number (25) of possible test words, subjects were informed of all possible superordinate categories before [Experiment 3](#).

Word lists for [Experiment 1](#), [Experiment 2](#), and [Experiment 3](#) are available on the web.

## EEG recording and data analysis

Subjects were fitted with a 19-channel electrode cap (Electro-Cap International, Eaton, OH) and were prepared for EEG recording according to standard techniques. Recorded channels (FP1, FP2, F7, F3, FZ, F4, F8, T7, C3, CZ, C4, T8, P7, P3, PZ, P4, P8, O1, and O2) were selected from the International 10-20 set of electrode positions (American Electroencephalographic Society, 1994). In addition to the cap electrodes, facial electrodes were attached to record horizontal and vertical electrooculogram (EOG). All recordings were referenced to the right mastoid, and were not re-referenced offline. Subjects performed the experiment in a darkened, sound-dampened, electrically shielded booth. EEG signals were amplified ([SA Instrumentation, San Diego](#)) with a high-pass cutoff of 100 Hz and a low-pass cutoff of 0.01 Hz, then sent through an analog-to-digital converter before being recorded at 256 samples/s on a PC running Digitize (Arthur Jones, LBNL).

Raw data were normalized, artifact rejected, and analyzed using Matlab software developed in-house. The trial-averaged EEG waveform – known as the event-related potential (ERP) – was computed separately for target and nontarget stimuli. ERP waveforms were combined into grand averages over all subjects. All ERP waveforms (as well as ERP images; see below) were computed using only trials on which the subject correctly identified the test stimulus as either a target or a nontarget, and all were time-locked to the presentation of the test stimulus. Software for the display of scalp topographies was developed by

Scott Makeig (SCCN, UC San Diego). Before averaging, all data were artifact rejected on a trial-by-trial basis for eye-blink and on a channel-by-channel basis for drift, blocking, and excessive alpha wave using automated in-house software. Data collected from 200 ms before presentation to 600 ms after presentation were checked for artifact. A trial was rejected for blink if the mean slope of the EEG exceeded a threshold of  $1 \mu\text{V}/\text{ms}$  over the 60 ms preceding and following the maximum value for a given trial. A channel was rejected for drift if the mean slope of the trial (excluding any blinks) exceeded  $70 \mu\text{V}/\text{s}$ . A channel was rejected for alpha if the power in the 8-12-Hz range exceeded 30% of total power for either the first half or the second half of the data. A channel was rejected for blocking if its value changed less than  $2 \mu\text{V}$  in a 40-ms interval.

To create the ERPimages, individual correct-response EEG trials (the first 700 ms following the presentation of the test stimulus) were assigned to 11.7-ms (three samples) wide bins on the basis of reaction time (RT). An average EEG was calculated for each bin. Bins with RTs between 300 and 700 ms were retained and were sorted by RT. Each averaged bin was then re-expanded in the vertical direction so its height was proportional to the number of EEG trials assigned to the bin. To create the difference ERPimages, the averaged nontarget bin was subtracted from the averaged target bin to create one difference wave at each RT. These difference waves were then sorted by RT and re-expanded vertically. In this case the expansion was proportional to the lesser of the number of EEG trials, target or nontarget, assigned to the original RT bins before subtraction. By making the vertical height proportional to the lesser of the two, RT bins that are noisier due to a small number of trials are given less weight in the plot. All ERPimages were then smoothed vertically with a Gaussian filter having a SD of one fiftieth the height (number of expanded trials) of the plot.

## Results

### Forward and reverse tasks

To determine if the target/nontarget difference (as reported in Johnson & Olshausen, 2003) was related to facilitated visual processing of the objects, we compared the results of our Forward and Reverse experiments, where the target decision was linked to the processing of the image or the word, respectively.

The grand average ERP waveforms for intact images in the Forward experiment (electrode CZ) are plotted in Figure 3a. As in previous cued-target paradigms, targets and nontargets begin to differ less than 200 ms after presentation. Figure 3c plots the target minus nontarget difference for the Forward task in blue, overlaid with results from two previous cued-target experiments (in black and gray), published in Johnson and Olshausen, 2003. The only major difference between the three tasks was the type of image used; all experiments used natural objects as test images, but the object could be in either a natural setting or one of two artificial backgrounds. In all three tasks the target minus nontarget difference is quite similar. In the Forward task described here, targets and nontargets first differed by our statistical criteria at 184 ms after presentation (two-sample  $t$  test for difference of means, 10 consecutive samples with  $p < .01$ , compared with 187/171 for the previous experiments), although the ERP waveforms suggest that the difference may onset 20-30 ms earlier. These data suggest that the background manipulations made to the images in this experiment do not strongly affect the target minus nontarget differences seen in previous results.

Figure 3b shows the target and nontarget ERPs for the Reverse experiment, time-locked to the presentation of the

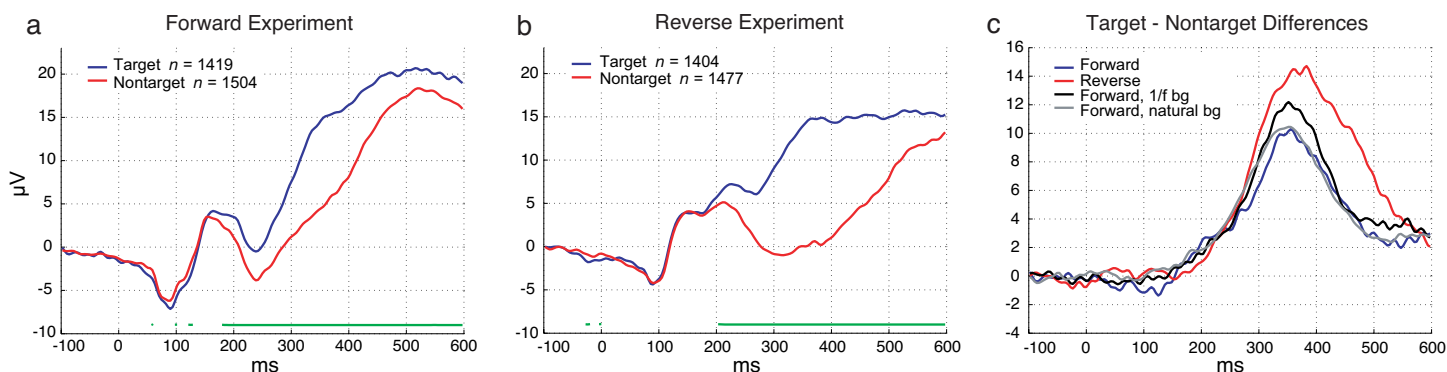


Figure 3. Differences in cued-target paradigms are not the result of scene-based processing. ERPs for the Forward experiment are locked to presentation of intact test image; ERPs for the Reverse experiment are locked to presentation of test word. Target/nontarget status of image or word is determined by preceding word cue (Forward) or intact image cue (Reverse). All potentials are calculated at electrode CZ. In (a) and (b), the green bar below indicates time points where waveforms differ significantly ( $p < .05$ ). (a). Forward experiment. (b). Reverse experiment. (c). Difference waves, target minus nontarget. Blue trace = Forward task. Red trace = Reverse task. Black trace (Forward, 1/f background) and gray trace (Forward, natural background) are taken for comparison from tasks previously published in Johnson and Olshausen, 2003. Differences are based on target status, not selective facilitation of scene processing for target scenes.

test word following an intact cue image. Although these ERPs were elicited by words rather than images, the target minus nontarget difference signal (Figure 3c, red) is quite similar to those in the image-based experiments. The difference reaches statistical significance slightly later (211 ms) and peaks about 4  $\mu\text{V}$  higher than that in the Forward experiment, but the overall time course as well as the scalp topography of the signal is the same as that in the Forward case (see Figure 6). This similarity across widely different test images would be expected of a postsensory signal that is related to the target status of the trial but is unlikely to arise from sensory processing itself.

## Effects of image difficulty

The target minus nontarget difference that we have shown bears some resemblance to the classically defined P300 in its time course and scalp topography. Another hallmark of the P300 is that its amplitude is reduced when the eliciting stimulus is degraded (Verleger, 1988). To determine whether the signal we see follows the same general pattern as the P300, we performed both the Forward and Reverse experiments with images that varied in their perceptual difficulty.

To vary image difficulty, we presented three types of images, interleaved, in both the Forward and Reverse tasks

Experiment	Target	Target	Nontarget	Nontarget
	%	RT	%	RT
Forward, Intact	92.7	453	96.0	479
Forward, Occluded	82.6	511	95.6	519
Forward, Deleted	78.4	513	95.2	518
Reverse, Intact	93.4	523	98.1	596
Reverse, Occluded	88.7	546	95.5	609
Reverse, Deleted	84.6	550	95.5	625
Speeded, Entry	93.9	514	96.9	584
Speeded, Super	88.7	545	94.3	589
Relaxed, Entry	94.7	456	97.9	529
Relaxed, Super	92.6	439	96.8	519

Table 1. Accuracy and reaction time (RT) separated by target and nontarget. The Forward experiment is separated into intact, occluded, and deleted test images. The Reverse experiment is separated into intact, occluded, and deleted cue images. Speeded and Relaxed experiments are separated into entry-level and superordinate-level categorizations.

(in order of increasing difficulty of recognition): intact images, occluded images, and deleted images. Accuracy and RT data for these images are found in Table 1. (A fuller assessment of the behavioral differences between the three types of images can be found in Johnson & Olshausen, 2005.) Figure 4a plots the difference waves for the three image conditions at electrodes FZ and CZ for the Forward

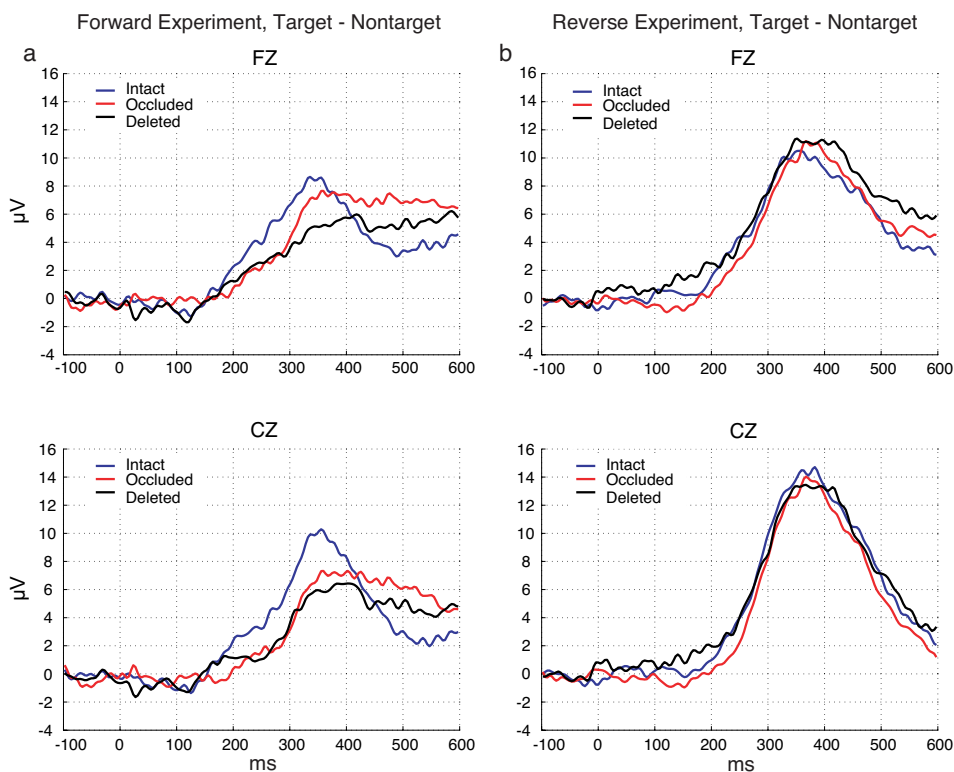


Figure 4. Target minus nontarget differences are sensitive to image difficulty. Shown are target minus nontarget difference waves at electrodes FZ and CZ for the Forward and Reverse experiments. (a). Forward experiment: target minus nontarget differences for intact test images (blue), occluded test images (red), and deleted test images (black). The amplitude of the difference is modulated by the difficulty of the test image. (b). Reverse experiment: target minus nontarget differences for test words. Cue images were either intact (blue), occluded (red), or deleted (black). Subjects viewed the test word only if they were confident they had identified the cue image. The amplitude of the difference is not modulated in this case.

experiment, where the images were acting as the test stimuli. There is a clear decrease in the amplitude of the target minus nontarget difference as the difficulty of the test image increases. The onset of the target minus nontarget difference is also delayed for occluded and deleted images relative to intact images (first time point to reach statistical criterion for FZ/CZ: intact images = 191/184 ms; occluded images = 223/285 ms; deleted images = 219/270 ms).

This decrease in amplitude, however, is not strictly related to the uncertainty in object identity brought about by the increase in image difficulty. The Reverse experiment also used intact, occluded, and deleted images as target cues – the test words were uniform regardless of the condition of the target cue. The target minus nontarget ERP differences time-locked to the presentation of the test word in the Reverse experiment are shown in Figure 4b. In this case there is no clear change in the amplitude of the signal across different types of cue image. The onset time of the target minus nontarget difference is more variable than is seen in other tasks (first time point to reach statistical criterion for FZ/CZ: intact images = 199/211 ms; occluded images = 227/242 ms; and deleted images = 160/164 ms).

The design of the Reverse experiment allowed subjects to opt out of the test phase of any trial where they did not feel they accurately identified the cue object. Even so, accuracy decreased and RTs increased as the type of cue image became more difficult. Because the test words did not vary in difficulty, the behavioral deficit for occluded and deleted cue images relative to intact ones must presumably be the result of uncertainty in what the correct target is, despite the subject's opportunity to abort the trial. The peak amplitude of the target minus nontarget difference decreases with difficulty of the test image in the Forward experiment, but does not decrease with cue difficulty in the Reverse experiment. The major differences arising from the type of cue image in the Reverse experiment occur before the peak

of activation, from the time of onset (where deleted images express the earliest onset) to about 250 ms after presentation, when all three target minus nontarget differences appear to come together.

## Effects of task difficulty

We made two further manipulations of the task to assess the effects of difficulty on the target minus nontarget signal: increasing the semantic difficulty of the task by changing the level of categorization of the target word, and increasing the overall difficulty of the task by increasing task demands.

Images are thought to first be semantically categorized at what is termed the entry level (Rosch et al., 1976; Jolicoeur et al., 1984), which is generally an object-name level (e.g., cat, chair). Abstraction of an entry level object to a superordinate or category level (e.g., animal, furniture) is thought to require additional nonperceptual processing. In addition to increases in RT and decreases in accuracy, superordinate categorization causes a marked decrease (about 5  $\mu$ V) in the target minus nontarget signal in a Forward cued-target task (Johnson & Olshausen, 2003). Does the same relationship hold true in a Reversed cued-target task?

We tested this question in both our Relaxed and Speeded tasks. ERPs from the Relaxed task (electrode CZ) are shown in Figure 5a, and target minus nontarget differences are shown in Figure 5c, the entry-level task in blue and the superordinate-level task in green. Although the amplitude difference between the entry and superordinate levels is not as pronounced as in our previous Forward experiment, the entry level task has a maximal amplitude about 1.5  $\mu$ V greater than the superordinate-level task.

The difference between entry- and superordinate-level categorization was also compared in the Speeded task. In this case, task demands were increased by two simultaneous manipulations. First, the entry- and superordinate-level

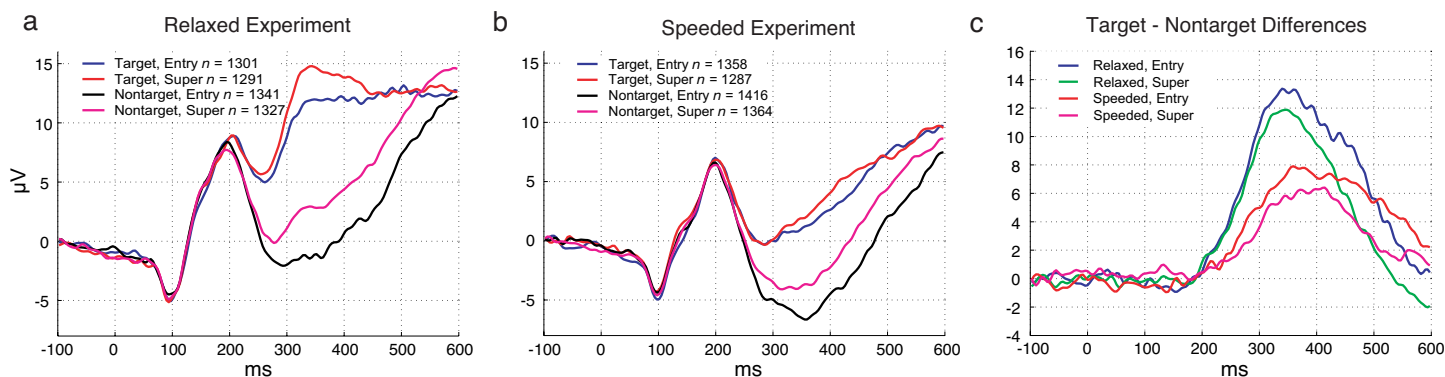


Figure 5. Target minus nontarget differences are sensitive to task demands. All potentials are calculated at electrode CZ. (a). ERPs for the Relaxed experiment. Entry-level and superordinate-level categorization are performed in separate blocks, explicitly specified. Blue = target at entry level. Red = target at superordinate level. Black = nontarget at entry level. Magenta = nontarget at superordinate level. (b). ERPs for Speeded experiment. Entry-level and superordinate-level categorizations are randomly mixed within the same block. Blue = target at entry level. Red = target at superordinate level. Black = nontarget at entry level. Magenta = nontarget at superordinate level. (c). Target minus nontarget differences for both experiments. Blue = Relaxed at entry level. Green = Relaxed at superordinate level. Red = Speeded at entry level. Magenta = Speeded at superordinate level.

categorizations were intermixed, so the subject was not aware at what level the target word would be until it appeared. Second, the target word was presented much more rapidly, 300-900 ms after the offset of the cue image without subject initiation. Figure 5b plots the ERPs from this experiment, and the target minus nontarget differences are shown in Figure 5c, entry level in red and superordinate level in magenta. As in the Relaxed task, we see a greater signal amplitude of about 1.5  $\mu$ V between the peak values of the two difference waves.

The 1.5- $\mu$ V entry/superordinate-level decrement seen in both the Relaxed and Speeded tasks is small compared to the 5  $\mu$ V seen in a previous Forward cued-target task. Note that in the previous Forward experiment, the superordinate categorization of the test image had to be performed on-line (i.e., at the same time that the image was being evaluated for target status). Under the Reverse paradigm, the superordinate categorization could be done immediately after presentation of the cue image and thus might be complete before the presentation of the test word and subsequent evaluation of target status. The difference that remains between entry and superordinate cases in the Reverse experiment presumably consists only of the uncertainty that one has correctly categorized the cue image, whereas the Forward case contains the additional effort involved in the superordinate categorization itself as well as any uncertainty.

Finally, we compared the target minus nontarget differences at the same semantic level of categorization, varying difficulty by changing the task demands. The Relaxed experiment allowed subjects to determine, at their leisure, the expected target word at the correct level of categorization. The Speeded experiment afforded subjects very little time to prepare for the upcoming test word, and forced them to entertain the possibility that the test word could appear at either of two levels of categorization. The behavioral data in Table 1 show that the Speeded task is in fact more difficult than the Relaxed task. Although entry-level accuracy was not statistically different between the Relaxed and Speeded tasks, accuracy at the superordinate level was significantly better in the Relaxed task than the Speeded task (targets,  $p < .0005$ ; nontargets,  $p = .001$ ,  $z$  test for two independent proportions). Similarly, the Speeded task resulted in a large increase in reaction times, between 60-100 ms, over the RTs for the Relaxed task regardless of level of categorization or target status (all four comparisons  $p < 10^{-19}$ , two-sample  $t$  test for difference of means). In the ERPs, Figure 5c shows that the peak amplitude of our target minus nontarget signal at electrode CZ is reduced by about 5.5  $\mu$ V in the more difficult Speeded task, regardless of the level of semantic categorization, suggesting that general increases in task demands cause a reduction in the amplitude of the target minus nontarget difference. The onsets of statistically significant differences in the Relaxed task are also

earlier than those in the Speeded task, according to the same criteria used above (onset time, entry-level/superordinate-level: Relaxed task = 210/203 ms; Speeded task = 242/258 ms).

## Scalp topography

It is important to note that there is a great deal of variability across several of our target conditions in the quality of the classically measured P300 (as defined as the first positive peak in the ERP on central/parietal electrodes following the N200). In the Forward experiment (Figure 3a), the P300 latency for targets would be defined at about 500 ms, and a P300 latency for nontargets would also be found only slightly later. In the Reverse experiment (Figure 3b), the target peak occurs before 400 ms, but there does not appear to be a nontarget peak before 600 ms have elapsed. In the Relaxed experiment (Figure 5a), strong P300s are evident with a peak around 350 ms for both entry and superordinate levels of categorization. There is a hint of an intermediate peak in the superordinate-level nontargets at the same time, but otherwise little evidence for a nontarget P300 before 600 ms. In the Speeded experiment (Figure 5b), no P300 peak is evident for either targets or nontargets. Despite this variability, it appears that the target minus nontarget differences in each case are fairly uniform, with only amplitude differences of consequence.

In Figure 6 we demonstrate this further by showing a grand average scalp topography of all 10 target minus nontarget differences recorded at the time point of maximum differential amplitude (labeled above each plot). Each plot is scaled to its own maximum value, which falls between 340-402 ms after presentation in all cases. The result is a series of plots that are remarkably similar. Each column in Figure 6 represents a different group of 12 experimental subjects. Despite drastic changes in the occlusion condition of the target image (Figure 6a, Forward experiment), replacement of target images with target words (Figure 6b and 6c Reverse paradigms), and changes in the semantic level of categorization and task demands (Figure 6c, Relaxed and Speeded experiments), there is very little variability in the scalp topography of the target minus nontarget difference between the various conditions. What variability can be seen is most notable in the left-right skew of the center of differential activity. In the Forward group and the Relaxed and Speeded group, the centroid of activity is shifted to lie right of the sagittal line, whereas in the Reverse group, the centroid is more or less centered. Because the recordings are referenced to the right mastoid, they may have a slightly different overall scalp distribution than if the recordings were re-referenced to the common average, but this would not affect their similarity to one another. Altogether, these remarkably similar scalp topographies suggest that the target minus nontarget difference, although subject to amplitude and latency variability, is created by the same neural generators independent of the eliciting task.

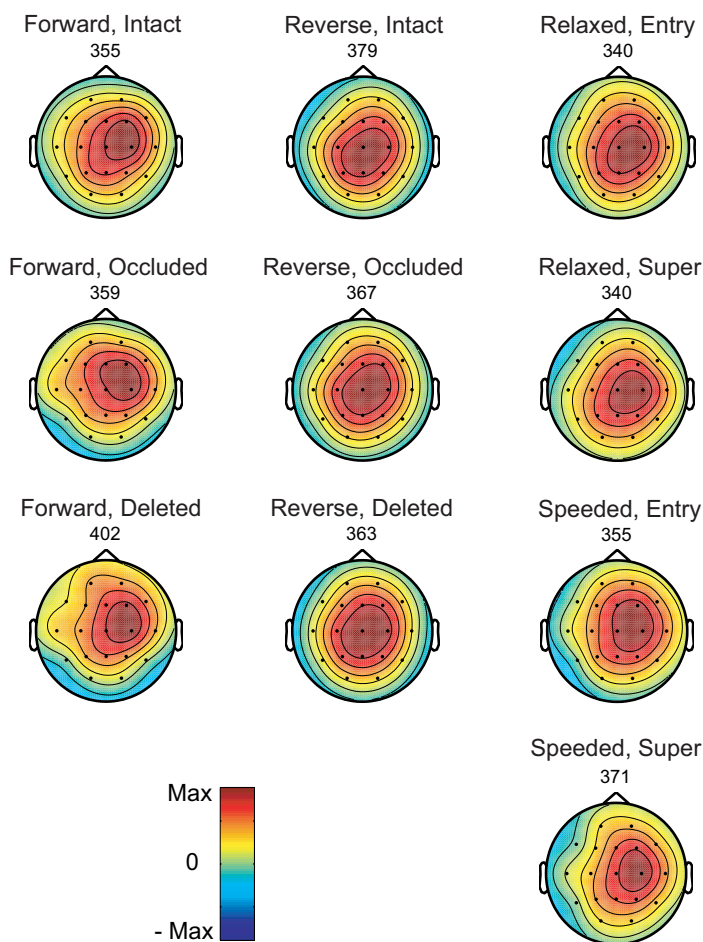


Figure 6. Target minus nontarget difference topographies. All topographies are shown at the time of maximum amplitude on electrode CZ. The number above the figure indicates actual time point in ms after presentation. Each plot is scaled to the maximum voltage value for that sample. Each column (a, b, and c) consists of data from a different group of 12 subjects. (a). Forward experiment. (b). Reverse experiment. (c). Relaxed/Speeded experiment (also a Reverse paradigm). Difference topographies across subject groups and task manipulations are quite similar to each other and to known P300 topographies.

## Reaction time dependence

To investigate the reaction time dependence of the P300 and the target minus nontarget difference, we utilized the ERPImage (Jung et al., 1999; Makeig et al., 2002) and the difference ERPImage (Johnson & Olshausen, 2003). The ERPImage shows the EEG waveforms from all trials sorted by reaction time and smoothed vertically, allowing the simultaneous identification of stimulus-locked and RT-dependent events. Using the ERPImage, estimates of the onset of any ERP component relative to the RT can be made on a trial-by-trial basis rather than on an experiment-wide basis.

In Johnson and Olshausen, 2003, we used the ERPImage to report the results of a forward cued-target task, and showed that the onset of target minus nontarget differences varies with the reaction time on a given trial. To determine whether this reaction time dependence also holds in the Reverse experiment, we plot in Figure 7 the ERPImages

and difference ERPImages for the intact image cue case of the Reverse experiment (channels FZ and CZ). Target trial ERPImages are shown in Figure 7a. The black curve shows the trial-by-trial RT. For both electrodes, P300 activity can be seen as positive-voltage activity that is RT-dependent, onsetting slightly before 300 ms on the fastest trials (RTs around 300 ms) and onsetting between 400–500 ms on the slowest trials (RTs around 700 ms). Nontarget ERPImages are shown in Figure 7b, and distinctly lack the P300 component seen in the target case while exhibiting a later, RT-dependent component. This later component onsets before the RT on CZ, but after the RT on FZ. Difference ERPImages shown in Figure 7c demonstrate that the target minus nontarget differences are also RT-dependent. Figure 7d shows ERPs for the two channels split by reaction time – fast trials (300–450 ms RT) in blue and slow trials (450–600 ms RT) in red – demonstrating the delay of both onset and peak of the difference on slow RT trials in a standard ERP format.

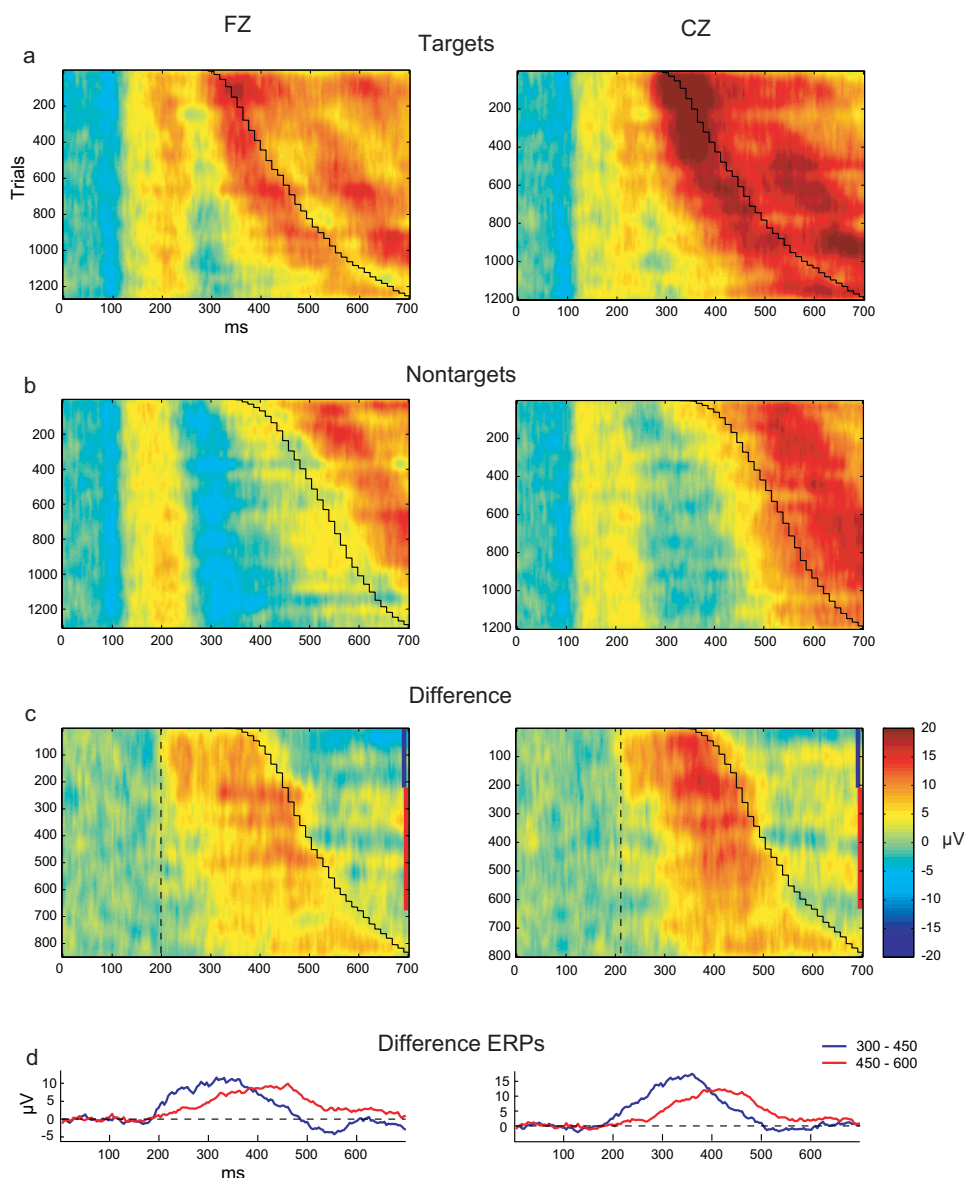


Figure 7. ERP images show RT-dependence on a trial-by-trial basis. The black curve indicates reaction time on each trial. All data are taken from the Reverse experiment, intact cue images. (a). Target ERP images. Onset of P300 activity is RT-dependent. (b). Nontarget ERP images. (c). Target minus nontarget ERP images. The vertical dashed line shows first time of significance (see Figure 3). Vertical solid lines (blue = 300 – 450 ms, red = 450 – 600 ms) identify trials used in fast-RT and slow-RT ERPs in (d). Onset of target minus nontarget differences is RT-dependent. (d). Difference ERPs, created for fast- and slow-RT trials. Difference ERPs also demonstrate RT-dependence of target minus nontarget signals.

The P300 seen in the target ERP images appears to peak after the RT itself on the fastest trials but peaks before the RT on trials with reaction times longer than about 350 ms. In contrast, the target minus nontarget difference clearly precedes the RT for all trials. Importantly, as can be seen in the difference ERP image and to a lesser extent in the raw ERPs (Figures 3 and 5), the target minus nontarget difference does not begin at the time of onset of the P300. Rather, the initial positivity that constitutes the difference between target trials and nontarget trials encroaches upon two earlier, largely stimulus-locked peaks that can be seen more clearly in the nontarget ERP images.

## Discussion

In this work we have attempted to determine what information processing stages underlie the target minus nontarget signal seen in cued-target tasks. This study produces four main findings regarding this signal: (1) that it persists across conditions where the target match is made to either photographic objects or to words, suggesting that it is post-sensory in nature; (2) that it shows a decrease in amplitude under several conditions that increase difficulty: degradation of the target stimulus (but not degradation of the cue

stimulus), change in the semantic level of categorization of the task, and generally increased task demands; (3) that across all of these manipulations, the scalp topography of the difference remains remarkably stable, suggesting that the same neural areas support the ERP differences between targets and nontargets in all cases; and (4) that for word targets the onset of the difference is related to the reaction time, as previously shown for object targets. Despite having an onset as fast as 200 ms, these observations lead us to two main conclusions: that the signal is postsensory in nature and that it is a component of the classically defined P300.

One of the central findings of the present experiments is that under our cued-target paradigm, the large target minus nontarget difference is not a direct correlate of neural activity related to visual processing but rather due to postsensory processing. However, because both viewing an image and reading a word consist of recognizing visual patterns, it could be argued that both processes could be subsumed by object recognition. Although previous results have made clear that in general sensory processing cannot underlie the target minus nontarget difference, sensory processing that is facilitated by a top-down target template (and thus results in higher levels of neural activity in the target case) might. Could not we see the same results from top-down facilitation in the recognition of words as in images? We think there are at least three reasons why this hypothesis is implausible at best.

First, the predictability of targets for the two types of stimuli is quite different. In the Forward case, where the test stimuli are natural objects, any potential top-down prediction of the image must by necessity be vague and distributed among many alternatives. One could never know ahead of time the exact location, orientation, shape, or color of the stimulus. For instance, try to imagine all possible view angles, colors, and styles of exemplars of the category "chair." In this case, top-down facilitation would only be useful at a relatively high level where representations are invariant; at lower levels individual features will be largely uninformative and cannot be retinotopically predicted. On the other hand, in the case of the Reverse experiments, quite a bit of specific visual information about the expected target is known before it appears. The words were always presented in the same color and font, and always centered on the screen, such that a good retinotopic prediction of the target word could conceivably be made. Thus, we would expect any potential top-down templates to be quite different in the two cases. If the target minus nontarget differences that we see were related to facilitation of visual processing exclusively in the target case arising from top-down predictions of that target, we should expect these predictions, for one, to be much more accurate at lower levels of the system in the Reverse experiment, leading to an earlier onset of difference. However, the onset of a statistically significant difference is almost 30 ms later than in the Forward case. Further, the location of some of this target-specific facilitation should be shifted from high-level areas

such as fusiform/parahippocampal gyri to lower level areas of visual cortex (e.g., V1, V4, and LOC). Although the limitations of EEG do not allow strong inference of signal source using scalp topography, a shift in signal source should be accompanied by a *change* in the scalp topography. In this case, however, no major change in topography is evident between the Forward and Reverse paradigms.

Second, if we were to assume that such top-down facilitation is restricted to acting only at the highest stages of recognition, we still run into the problem that the cortical systems responsible for the final stages of visual processing of object stimuli and letter strings seem to be physically distinct (Nobre, Allison, & McCarthy, 1994). Object recognition seems to occur in a bilateral pathway that stretches from the lateral occipital complex ventrally through inferotemporal cortex, including fusiform gyrus and parahippocampal gyrus (Haxby et al., 1991; Ishai, Ungerleider, Martin, & Haxby, 2000; Grill-Spector, 2003), but the recognition of words seems to be highly localized to a left occipitotemporal sulcus region known as the visual word form area (VWFA) (Tarkiainen, Helenius, Hansen, Cornelissen, & Salmelin, 1999). Lesion studies show that normal reading depends critically on the VWFA but that loss of function in VWFA does not affect object naming (Cohen et al., 2003). The EEG signature from VWFA is recorded over left inferior temporal electrodes between 180-200 ms poststimulus (Cohen et al., 2000); it does not match the signals reported here in topography. It also occurs early enough to reliably precede the earliest onset of target minus nontarget activity in the Reverse paradigm, where target-related activity should depend on information from the VWFA.

Finally, the weight of the evidence suggests that we are recording a signal related to the postsensory P300. The P300s themselves, as found in the target and nontarget ERPs, are quite variable: In some cases they have a peak latency similar to the target minus nontarget signal, in some cases their peak latency is over 100 ms later, and in some cases they are not evident at all. Although the target and nontarget ERPs do not always match those of a classic P300, the target minus nontarget difference is present in all cases and is remarkably similar across them. The target minus nontarget signal has a scalp topography that is quite similar to the classic P300, though it is sometimes a bit more frontal. It is also sensitive to some of the same manipulations as the P300, showing decreased amplitude when the elicited image is degraded, or the task is made more difficult. Although it is not evoked or studied in the same fashion as the classic P300, it still bears enough similarity to P300 activity (especially in time course, scalp topography, and RT-dependence) to suggest that it is not completely divorced from P300 activity. Although other, concurrent activity may in some cases render the target minus nontarget signal nearly invisible (e.g., Figure 5b), its large amplitude suggests that it may in general be a major contributor to P300 activity.

All of this evidence leads us to believe that the target minus nontarget activity is a postsensory phenomenon.

Rather than being specifically tied to the parsing of visual information, it appears to be a higher level, decision-related signal evoked by an abstract target match. As such, we propose that the signal should be modality independent, and we expect that cued-target paradigms in, for instance, the auditory domain, would evoke target minus nontarget differences that are similar in topography and in amplitude susceptibility to task difficulty.

One of the apparent unresolved debates in the P300 literature involves a disagreement over the effects of task difficulty on the amplitude of the P300. Although some suggest (e.g., Johnson, Jr., 1988) that the P300 amplitude increases with task difficulty, others suggest (Verleger, 1988) that it decreases. The amplitude effects may be muddy partly because of the multi-component nature of the P300 (Johnson, Jr., & Donchin, 1985), which may conflate competing effects. When looking only at the raw ERP waveform, we find good examples of this sort of ambiguity in our experiments. In the Relaxed experiment (Figure 5a), superordinate categorization elicits a higher amplitude P300 than entry-level categorization, but in the comparison of the Relaxed and Speeded experiments (Figure 5a and 5b), the Speeded task is accompanied by a reduced or eliminated P300. Even within the same group of subjects, two different types of task complexity have opposite effects on P300 amplitude. However, in both cases the target minus nontarget difference, which is certainly a major contributor to the overall waveform, has a reduced amplitude in the more difficult condition (Figure 5c). This suggests that there may be many types of task difficulty that do not have similar effects on P300 amplitudes, semantic difficulty and task load being just two examples. These effects appear to be independent of the target effect (also see, Tueting & Sutton, 1976) that we isolate with our difference method, one advantage of which is that it is definitely robust to task and stimulus-related changes that may obscure classic P300s.

It has been argued that P300 activity cannot reflect the subject's decision on any particular trial because the latency of the P300 peak is often longer than individual trial reaction times (e.g., Donchin & Coles, 1988). Contextual updating, one of the major hypotheses that tries to explain P300 activity, proposes that the P300 is involved in updating the brain's running model of the world and the probabilities of the stimuli that occur in it – a non-decision process that need not occur before the reaction time. But the target minus nontarget difference, which appears to be a subcomponent of the P300, arises quite a bit earlier than the P300 itself, first becoming evident when the overall ERP is dominated by the N200 (and in some cases the P1) component. Although P300 activity continues beyond the reaction time, the target minus nontarget difference itself generally concludes at the time of the reaction (compare Figure 7a and 7c) and, as a subcomponent of the P300, is a decent candidate for being correlated with the subject's decision.

There are several hypotheses as to what function the neural processes underlying the target minus nontarget difference play in the cascade of events leading from the presentation of a test stimulus to the press of a button based on the content of that stimulus. We have previously ruled out motor involvement, and the results here suggest that direct sensory involvement is also out. One possibility is that the ERP differences reflect activation not on target trials but on nontarget trials. Either widespread competitive inhibition on target trials or additional processing of nontargets (to rule them out as targets) might lead to higher activation on nontarget trials. Some fMRI data showing reduced activity in posterior cingulate and parahippocampal and fusiform gyri on target trials in a similar task suggest this hypothesis (Fize et al., 2000).

Another hypothesis is that the target minus nontarget difference reflects task-related working memory. It has been shown that some cells in macaque prefrontal cortex exhibit tonic activity during the delay period of a delayed match to sample (DMS) task (Fuster & Alexander, 1971). These cells abruptly cease firing once a behavioral response has been made in the task. Furthermore, these cells respond more robustly to stimuli that match the target than to stimuli that do not (Miller, Erickson, & Desimone, 1996). The target minus nontarget activity that we find in the cued-target task appears to exhibit both of these qualities, being stronger to targets and offsetting at the time of the subject's reaction. Of course, the standard DMS task involves a physical stimulus match while our cued-target task involves only a conceptual match (the match of the word "train" with a picture of a train), but there is no reason to believe that cells in prefrontal cortex involved in working memory require an explicit physical match. The target minus nontarget activity that we see may very well be a reflection of working memory processes in prefrontal cortex.

Most importantly, it appears that the target minus nontarget signal seen in these tasks is postsensory. In our previous report on this signal (Johnson & Olshausen, 2003), we were unable to determine its source. The results here suggest that it is not due to either facilitation in low-level visual cortex nor activation in high-level visual cortex, but rather it appears to be postsensory and related to the target decision. The finding that the earliest known neural signature of object recognition is not directly a reflection of visual processing itself has immediate implications for attempts to determine the time course of visual processing.

Target minus nontarget difference signals, which onset in the range of 150-300 ms, place an upper bound on the time required to sufficiently process an object to report its target status, but the precision of that upper bound is inherently dependent on the source of the signal. If the difference we find were a sensory processing signal, we could confidently claim that object processing takes significantly longer on some trials than others. However, we show here that the difference is instead a postsensory decision-related signal – as such it is liable to arise tens to hundreds of milliseconds after sufficient sensory processing for target iden-

tification has occurred. Certainly in the fastest reaction trials, when the target minus nontarget difference may arise in 150 ms, sufficient processing to identify the object must have occurred by about 140 ms, leaving limited time available for recurrent visual processing. The target minus nontarget difference is delayed on long reaction time trials relative to short reaction time trials, but because it is postsensory, we have no evidence for whether the delay on long-RT trials corresponds to longer visual processing or a longer decision process. If most of the RT variability lies in the time of the decision, the highest upper bound may also be about 150 ms; that is, object recognition may be uniformly complete by 150 ms on every trial. If, on the other hand, most of the RT variability lies in the time of completion of object recognition itself, the highest upper bound would be nearer to 300-350 ms, suggesting that the visual processing may require as much as an additional 200 ms on some images, a period which could be used for recurrent and feedback processing. The ability to distinguish between these two alternatives with a method that allows a precise determination of the time of object recognition on a trial-by-trial basis will be an important step in guiding the efforts of those modeling the human visual system.

## Acknowledgments

This work was supported by National Institutes of Health Grant MH57921 (BAO). Thanks to Ken Britten and David Woods for draft comments.

Commercial relationships: none.

Corresponding author: Bruno A. Olshausen.

Email: baolshausen@ucdavis.edu.

Address: UC Davis Center For Neuroscience, 1544 Newton Ct., Davis CA 95616.

## References

- Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I. Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral Cortex*, *9*, 415-430. [PubMed]
- American Electroencephalographic Society (1994). Guideline thirteen: Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, *11*, 111-113. [PubMed]
- Breitmeyer, B. G. (1984). *Visual masking: An integrative approach*. New York: Oxford University Press.
- Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.-A., et al. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, *123*, 291-307. [PubMed]
- Cohen, L., Martinaud, O., Lemer, C., Lehéricy, S., Samson, Y., Obadia, M., et al. (2003). Visual word recognition in the left and right hemispheres: Anatomical and functional correlates of peripheral alexias. *Cerebral Cortex*, *13*, 1313-1333. [PubMed]
- Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*, 357-374.
- 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. [PubMed]
- Fize, D., Boulanouar, K., Chatel, Y., Ranjeva, J. -P., Fabre-Thorpe, M., & Thorpe, S. (2000). Brain areas involved in rapid categorization of natural images: An event-related fMRI study. *NeuroImage*, *11*, 634-643. [PubMed]
- Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*, 193-202. [PubMed]
- Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short term memory. *Science*, *173*, 652-654. [PubMed]
- Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, *13*, 159-166. [PubMed]
- Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience*, *3*, 837-843. [PubMed]
- Haxby, J. V., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., et al. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proceedings of the National Academy of Science U.S.A.*, *88*, 1621-1625. [PubMed][Article]
- Intraub, H. (1999). *Fleeting memories: Cognition of brief visual stimuli* (V. Coltheart, Ed.) (pp. 47-70). Cambridge, MA: MIT Press.
- Ishai, A., Ungerleider, L. G., Martin, A., & Haxby, J. V. (2000). The representation of objects in the human occipital and temporal cortex. *Journal of Cognitive Neuroscience*, *12*(Suppl. 2), 35-51. [PubMed]
- Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision*, *3*(7), 499-512. <http://journalofvision.org/3/7/4/>, doi:10.1167/3.7.4. [PubMed][Article]
- Johnson, J. S., & Olshausen, B. A. (2005). *The recognition of partially visible natural objects in the presence and absence of their occluders*. Manuscript submitted for publication.

- Johnson, R., Jr. (1988). The amplitude of the P300 component of the event-related potential: Review and synthesis. *Advances in Psychophysiology*, 3, 69-137.
- Johnson, R., Jr., & Donchin, E. (1985). Second thoughts: Multiple P300s elicited by a single stimulus. *Psychophysiology*, 22, 182-194. [PubMed]
- Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, 16, 243-275. [PubMed]
- Jung, T. -P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (1999). *Advances in neural information processing systems* 11 (M. S. Kearns, S. A. Solla, & D. A. Cohn, Eds.) (pp. 118-124). Cambridge, MA: MIT Press.
- Keysers, C., Xiao, D. -K., Földiák, P., & Perrett, D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience*, 13, 90-101. [PubMed]
- Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, 197, 792-795. [PubMed]
- Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, 20, 1434-1448. [PubMed]
- Lewicki, M. S., & Sejnowski, T. J. (1997). *Advances in neural information processing systems* 9 (M. C. Mozer, M. I. Jordan, & T. Petsche, Eds.) (pp. 529-535). Cambridge, MA: MIT Press.
- Makeig, S., Westerfield, M., Jung, T. -P., Enghoff, S., Townsend, J., Courchesne, E., et al. (2002). Dynamic brain sources of visual evoked responses. *Science*, 295, 690-694. [PubMed]
- Mel, B. W. (1997). SEEMORE: Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Computation*, 9, 777-804. [PubMed]
- Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience*, 16, 5154-5167. [PubMed]
- Mumford, D. (1994). Large scale neuronal theories of the brain (C. Koch & J. L. Davis, Eds.) (pp. 125-152). Cambridge, MA: MIT Press.
- Nobre, A. C., Allison, T., & McCarthy, G. (1994). Word recognition in the human inferior temporal lobe. *Nature*, 372, 260-263. [PubMed]
- Nowak, L. G., & Bullier, J. (1997). *Cerebral cortex* 12 (K. S. Rockland, J. H. Kaas, & A. Peters, Eds.) (pp. 205-241). New York: Plenum Press.
- Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, 9, 456-479. [PubMed]
- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2, 79-87. [PubMed]
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2, 1019-1025. [PubMed]
- Ritter, W., Simson, R., & Vaughan, H. G., Jr., (1972). Association cortex potentials and reaction time in auditory discrimination. *Electroencephalography and Clinical Neurophysiology*, 33, 547-555. [PubMed]
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-439.
- Soltani, M., & Knight, R. T. (2000). Neural origins of the P300. *Critical Reviews in Neurobiology*, 14, 199-224. [PubMed]
- Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, 38, 387-401. [PubMed]
- Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, 150, 1187-1188. [PubMed]
- Tarkiainen, A., Helenius, P., Hansen, P. C., Cornelissen, P. L., & Salmelin, R. (1999). Dynamics of letter string perception in the human occipitotemporal cortex. *Brain*, 122, 2119-2131. [PubMed]
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381, 520-522. [PubMed]
- Tueting, P., & Sutton, S. (1976). *The responsive brain* (W. C. McCallum & J. R. Knott, Eds.) (pp. 71-75). Bristol: John Wright & Sons.
- Ullman, S. (1996). *High-level vision: Object recognition and visual cognition*. Cambridge, MA: MIT Press.
- VanRullen, R., & Thorpe, S. (2001a). Rate coding vs. temporal order coding: What the retinal ganglion cells tell the visual cortex. *Neural Computation*, 13, 1255-1283.
- 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. [PubMed]
- VanRullen, R., & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Research*, 42, 2593-2615. [PubMed]
- Verleger, R. (1988). Event-related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of P3. *Behavioral and Brain Sciences*, 11, 343-427.