

Limits of Event-related Potential Differences in Tracking Object Processing Speed

Guillaume A. Rousselet, Marc J.-M. Macé, Simon J. Thorpe,
and Michèle Fabre-Thorpe

Abstract

■ We report results from two experiments in which subjects had to categorize briefly presented upright or inverted natural scenes. In the first experiment, subjects decided whether images contained animals or human faces presented at different scales. Behavioral results showed virtually identical processing speed between the two categories and very limited effects of inversion. One type of event-related potential (ERP) comparison, potentially capturing low-level physical differences, showed large effects with onsets at about 150 msec in the animal task. However, in the human face task, those differences started as early as 100 msec. In the second experiment, subjects responded to close-up views of animal faces or human faces in an attempt to limit physical differences between image

sets. This manipulation almost completely eliminated small differences before 100 msec in both tasks. But again, despite very similar behavioral performances and short reaction times in both tasks, human faces were associated with earlier ERP differences compared with animal faces. Finally, in both experiments, as an alternative way to determine processing speed, we compared the ERP with the same images when seen as targets and nontargets in different tasks. Surprisingly, all task-dependent ERP differences had relatively long latencies. We conclude that task-dependent ERP differences fail to capture object processing speed, at least for some categories like faces. We discuss models of object processing that might explain our results, as well as alternative approaches. ■

INTRODUCTION

If most researchers agree that we are fast and efficient at extracting information from visual stimuli, exactly how fast and how efficient remain very debated questions (e.g., Johnson & Olshausen, 2005; Rousselet, Macé, & Fabre-Thorpe, 2003; Riesenhuber & Poggio, 2002; Thorpe & Fabre-Thorpe, 2001). At the behavioral level, a series of experiments from our group has documented the extraordinary capacity of human observers to categorize briefly presented natural scenes, typically with mean accuracy of approximately 94% and median reaction time (RT) of less than 400 msec (e.g., Rousselet, Fabre-Thorpe, & Thorpe, 2002; Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; VanRullen & Thorpe, 2001a; Thorpe, Fize, & Marlot, 1996).

However, any behavioral RT measurement includes not only the time required for sensory processing but also the time to initiate and execute the motor response. One approach, which has been used successfully in normal human subjects to tackle more precisely the sensory processing time, involves event-related potential (ERP) recording. By analyzing the averaged waveforms associated with correct responses to images containing targets

and subtracting the averaged waveforms associated with correct responses to nontarget images, one obtains a difference waveform that can be used to determine the moment when cerebral responses to targets and nontargets start to differ. The time at which the difference waveform starts to diverge significantly from baseline can provide an estimate of the time necessary to discriminate targets from nontargets. Earlier studies using an animal categorization task found that the difference between target and nontarget ERPs becomes statistically significant at about 150 msec after the onset of the images (Rousselet et al., 2002; Fabre-Thorpe et al., 2001; Thorpe et al., 1996).

However, interpreting these differential responses is not without difficulties. In some conditions, one can obtain significant differences in the ERPs of two classes of images that could simply be because of low-level differences in the physical properties of the pictures and not because of recognition per se. For example, a set of images darker than another one could easily produce differences in the neural responses in areas such as V1 and lead to ERP differences with remarkably short latencies. One way to tackle only task-related effects is to change the target status of the images so that one can compare the ERP responses to exactly the same set of images processed either as targets or as nontargets. In such a case, the same physical images are compared,

thus, any differential signals cannot be because of low-level features. This approach using task-dependent ERP differences was first developed to study the effects of attention in the auditory system (e.g., Hillyard, Hink, Schwent, & Picton, 1973) and then in the visual system using relatively simple stimuli (e.g., Hillyard & Munte, 1984). VanRullen and Thorpe (2001b) extended this approach to natural scenes and showed that although very early differential effects were abolished by such a manipulation, the differential signal that developed from 150 msec was preserved. This remarkable processing speed might correspond to an optimum since even extensive training with the same set of images cannot speed up the earliest behavioral responses or shorten the latency of the differential signal at 150 msec (Fabre-Thorpe et al., 2001). Furthermore, a very simple task requiring the recognition of a single well-learned image among distractors can only be performed about 30–40 msec faster than the categorization of natural scenes (Delorme, Rousselet, Macé, & Fabre-Thorpe, 2004).

Typical experiments from our group have requested subjects to categorize animals, means of transport, food, and others in natural scenes. But other categories might potentially be processed even faster. For instance, it has been suggested that human faces, because of a special computational status (e.g., Kanwisher, 2000; Farah, Wilson, Drain, & Tanaka, 1998), might be processed more efficiently and even faster than any other class of objects. However, the exact speed at which faces and other objects can be processed remains a controversial question, as onsets of brain activity differences have been reported as early as 50–80 msec poststimulus (Mouchetant-Rostaing, Giard, Bentin, Aguera, & Pernier, 2000; Mouchetant-Rostaing, Giard, Delpuech, Echallier, & Pernier, 2000; George, Jemel, Fiori, & Renault, 1997; Seeck et al., 1997) in the 100- to 130-msec latency range (Itier & Taylor, 2004; Halgren, Raji, Marinkovic, Jousmaki, & Hari, 2000; Halit, de Haan, & Johnson, 2000; Pizzagalli, REGARD, & Lehmann, 1999; Yamamoto & Kashikura, 1999; Debruille, Guillem, & Renault, 1998; Linkenkaer-Hansen et al., 1998; Schendan, Ganis, & Kutas, 1998), in the 150- to 200-msec latency range (Rossion, Joyce, Cottrell, & Tarr, 2003; Carmel & Bentin, 2002; Taylor, Edmonds, McCarthy, & Allison, 2001; Eimer, 2000a, 2000b; Bentin, Allison, Puce, Perez, & McCarthy, 1996; Jeffreys, 1996), or even later, after 200 msec (Johnson & Olshausen, 2003, 2005). It is important to note that all those experiments used different kinds of stimuli, tasks, and conditions of presentation.

Here we explore the possibility that faces might be processed even faster than other categories of objects in the context of natural scenes. Our goal was to determine face processing speed using a methodology that has been used extensively and with reliable results in previous experiments from our group. To obtain an estimate of face processing speed, we used a rapid go/no-go categorization task with briefly presented (26 msec)

photographs of real-world scenes in which subjects had to react only when the photograph contained a human face. This sort of go/no-go design involves the simplest motor output possible, allowing subjects to respond as fast as they could with the minimal motor constraints (Bacon-Macé, Kirchner, Fabre-Thorpe, & Thorpe, in press). For comparison with another class of targets, subjects alternated between this face categorization task and an animal categorization task used in previous experiments. In the context of this speeded task, we also investigated the robustness of visual processing to stimulus inversion, a manipulation known to disrupt more extensively human face processing compared with other object categories (Rossion & Gauthier, 2002; Bentin et al., 1996; Yin, 1969).

Two experiments were performed, with the main difference being the size and the variability of the stimuli. In the first experiment, subjects had to decide whether the image contained either an animal or a human face presented in the context of a human body. The animals and humans could be at different sizes and positions within a natural scene. In the second experiment, subjects were only presented with close-up views of faces and had to respond either to animal faces or to human faces, but in this case, the images were all relatively close-up views of just the head region. As previously reported, behavioral performance was exceptionally good (Table 1), despite the wide range of stimuli used (Rousselet et al., 2003). Inverting images had remarkably little effect on performance. One important result was the extreme similarity of accuracy and RT between human and animal faces, suggesting no particular advantage for human faces.

In the present article, we provide a systematic analysis of ERP data recorded while subjects categorized human and animals to help determine the time course of the underlying processes. Here, we do not focus on the N170 ERP component (Rousselet, Macé, & Fabre-Thorpe, 2004) but rather track differences related to the categorization process at every time point and electrode. Our predictions were very straightforward: Given the similarity of the behavioral results obtained with human faces and animals in the two experiments, we expected to find almost identical patterns of ERP differences for the different categories tested, exactly as VanRullen and Thorpe (2001b) found when comparing animal and vehicle categorization. The actual results were far from those expected. ERP differential activities that might contain physical differences were weaker but appeared earlier for human compared with animal stimuli. In addition, particularly in the case of face stimuli, the task-dependent ERP differences (that did not contain low-level differences) were surprisingly of relatively long latencies. This result, which contrasts strongly with the remarkably accurate behavioral responses of the subjects and their very short behavioral RTs, implies that task-dependent ERP differences cannot

Table 1. Summary of the Behavioral Data for the “All-scales” and “Close-up-views” Experiments Reported by Rousselet et al., 2003

	<i>Human Task</i>		<i>Animal Task</i>	
	<i>Upright Scenes</i>	<i>Inverted Scenes</i>	<i>Upright Scenes</i>	<i>Inverted Scenes</i>
<i>“All-scales” experiment</i>				
Mean accuracy (%)	96.4 (1.7)	94.7 (2.3)	96.3 (2.0)	94.8 (2.3)
Median RT (msec)	368 (43)	391 (50)	371 (42)	380 (44)
Minimal RT (msec)	260	260	260	260
<i>“Close-up-views” experiment</i>				
Mean accuracy (%)	97.7 (1.8)	97.2 (1.7)	97.9 (1.3)	96.9 (1.6)
Median RT (msec)	371 (31)	382 (26)	384 (37)	391 (34)
Minimal RT (msec)	260	280	270	270

SD is indicated in parentheses. The minimal RT corresponds to the first time bin where correct hits significantly outnumber false alarms (see the work of Rousselet et al., 2003, for details).

be used systematically to infer object processing speed, at least for some categories such as faces.

METHODS

Forty-eight subjects volunteered in two studies and gave their written informed consent. All had normal or corrected-to-normal vision. Nine subjects participated in both experiments.

Task Setup

Subjects sat in a dimly lit room at 100 cm from a computer screen (resolution: 800 × 600; vertical refresh rate: 75 Hz). To start a block of trials, subjects had to place a finger on a response pad for 1 sec, then a fixation cross (0.1° of visual angle) appeared for 300–900 msec and was followed by the stimulus presented in the middle of the screen for two frames (i.e., ~26 msec) (Figure 1A). Participants had to lift their finger as quickly and as accurately as possible (go response) each time a target was presented. Responses were detected using infrared LEDs. Subjects had 1000 msec to respond, after which their response was considered as a no-go response. A 300-msec black screen followed this maximum response time delay, before the fixation point was presented again for a variable duration, resulting in a random 1600- to 2200-msec intertrial interval. When the photographs contained no target, subjects had to keep their finger on the pad for at least 1000 msec (no-go response).

In experiment 1 (targets at all scales), a session included 16 blocks of 96 trials, and subjects alternated between two categorization tasks. In 8 of the blocks, the targets were animals, and in the 8 other blocks, the targets were human faces. Half of the subjects started

with 2 blocks of animal categorization, the other half with 2 blocks of human face categorization, and conditions alternated every 2 blocks. In experiment 2 (close-up views of targets), there were 8 blocks. In the first 4 blocks, the targets were animal faces, and in the other 4 blocks, the targets were human faces (starting blocks counterbalanced across subjects). For both experiments, in each block, target and nontarget trials were equally likely. Among the 48 nontargets, 24 were targets of the other categorization task (Figure 1B). Thus, when performing a human face categorization task, on a 96-trial block, 48 pictures contained human faces, 24 nontarget scenes contained animals, with the last 24 nontargets being other types of natural scenes. Moreover, half of the images in each subset were presented upright, whereas the other half were presented inverted (rotation 180°). A given subject saw each image only once, with one orientation (upright or inverted) and one status (target or nontarget). All images were processed in all conditions across subjects. Subjects had two training blocks of 48 images before starting the test session. Training pictures were not used during the test.

Stimuli

We used photographs of natural scenes taken from a large commercial picture library (Corel Stock Photo Library). They were all horizontal photographs (768 × 512 pixels, sustaining about 20° × 13.5° of visual angle) and chosen to be as varied as possible (Figure 1B). Animals included essentially mammals, but birds, fish, and reptiles were also included. Human faces were presented in real-world situations, with views ranging from whole bodies at different scales to face close-ups and including white and nonwhite people. There was also a very wide range of nontarget images that included

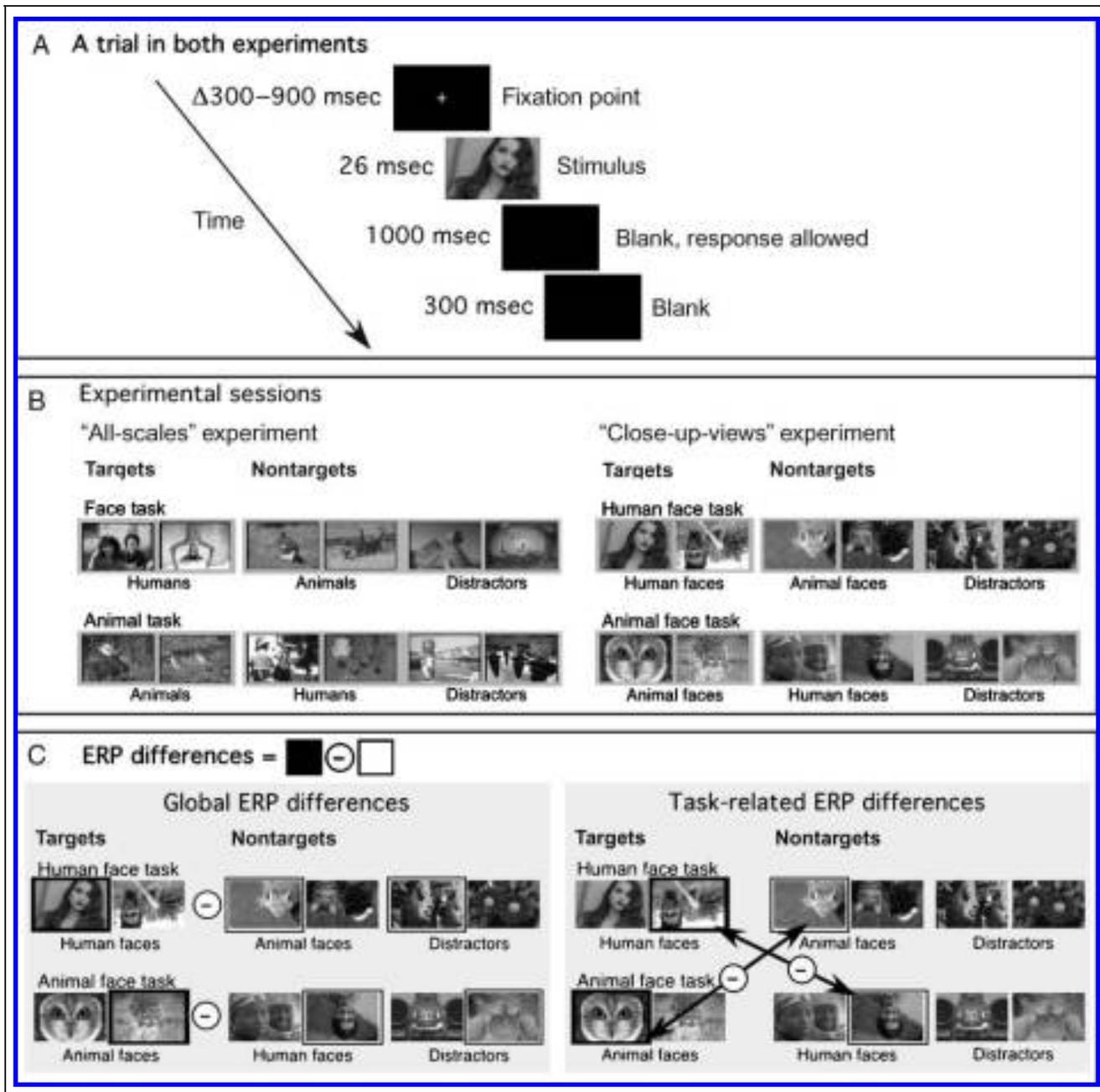


Figure 1. Experimental design. (A) Organization of a trial in both experiments. A trial started with the presentation of a fixation point for a random delay ranging from 300 to 900 msec. Then, a stimulus was presented for about 26 msec, followed by a black screen for 1300 msec. Subjects could provide a response during the first 1000 msec of the black screen presentation. (B) Experimental sessions in the "all-scales" and "close-up-views" experiments. In both experiments subjects performed two tasks. In each task, 50% of nontarget stimuli were targets of the other task. (C) Global and task-related ERP differences. Global ERP differences are illustrated for upright human faces and inverted animal faces. In each case, ERP recorded in response to all nontarget trials at one orientation (upright or inverted) were subtracted from ERP recorded in response to all target trials with the same orientation. Task-related ERP differences are illustrated for inverted human faces and upright animal faces. Task-related ERP differences were computed by subtracting ERP recorded in response to nontarget trials of one category and one orientation in one task from ERP recorded in response to target trials of the same category and the same orientation in the other task.

outdoor and indoor scenes, natural landscapes (mountains, fields, forests, beaches, etc.), street scenes, pictures of food, fruits, vegetables, plants, buildings, tools, and other manmade objects, as well as some tricky distracters (e.g., dolls, sculptures, statues, as well as nontarget images containing humans for which the faces

were not visible). In experiment 2, only close-up views of target objects were used, and a special attempt was made to use many tricky nontargets and "blob" objects appearing in positions similar to human and animal faces. Subjects had no a priori information about the presence, the size, the position, or the number of

targets in an image. Brief presentation prevented ocular exploration, and trial unique presentation prevented learning.

Electroencephalographic Recording and Analysis

A SynAmps amplifier system (Neuroscan, El Paso, TX) was used to record brain electrical activity with 32 electrodes mounted in an elastic cap (Oxford Instruments, Oxford, UK) in accordance with the 10-20 system with the addition of extra occipital electrodes from the 10-10 system (FP1/2, F3/4, F7/8, Fz, C3/4, Cz, T7/8, Pz, P3/4, PO3/4, POz, TP78, T5/T6, PO7/8, O1/2, Oz, Iz, PO9/10, and O9/10). The ground electrode was placed along the midline, ahead of Fz, and impedance was systematically kept below 5 k Ω . Signals were digitized at a sampling rate of 1000 Hz and low-pass filtered at 40 Hz before analysis. Potentials were on-line referenced on electrode Cz and re-referenced off-line by subtracting the average of all electrodes from each individual electrode signal. Baseline correction was performed using the 100 msec of prestimulus activity. Two artifact rejections were applied over the [−100 msec; 400 msec] time period: first, on frontal electrodes with a criterion of [−80; 80 μ V] to reject trials with eye movements; second, on parietal electrodes with a criterion of [−40; 40 μ V] to remove trials with excessive activity in the α range. Only correct trials were averaged.

Differences between pairs of conditions were assessed using a percentile method with 999 permutations (Wilcox, 2005). For a given comparison and for each subject, the ERPs for the two conditions were randomly shuffled, which is equivalent to assigning the “condition 1” and “condition 2” labels randomly to the two conditions while keeping electrodes and time points together. The two “fake” ERPs were then averaged across subjects, and the difference was stored. This procedure, repeated 999 times, provided a confidence interval around the mean difference under the null hypothesis that the two conditions were actually sampled from the same population. Permutation procedures do not assume a specific data distribution and provide more power than parametric statistics such as paired *t* tests. Differences reported in this article are significant at $p < .01$, corrected for multiple comparisons by dividing α by the number of electrodes (i.e., 0.01/32). This Bonferroni correction is more conservative than some other available methods (Wilcox, 2005).

As stated in the Introduction, the design of the experiments was similar to that of VanRullen and Thorpe (2001b) and allowed us to compute two types of differential signal, with or without physical differences between targets and nontargets (Figure 1C). The differential activity classically computed, and reported in our previous publications, is called “global differential activ-

ity” throughout this article. It corresponds to the subtraction of the signal recorded on nontargets from the signal recorded on targets in a given categorization task. These images are different, and hence, the global differential activity reflects both stimulus- and task-related differences. In addition, we computed a task-related differential activity that corresponds also to a target minus nontarget subtraction, but with the same images processed as target or nontargets in two different categorization tasks, thus eliminating any low-level differences between target and distractor image sets.

RESULTS

“All-scales” Experiment: First Estimation of Processing Speed

In the first experiment, subjects ($n = 24$, 12 women and 12 men, mean age: 31 years) performed remarkably well. A detailed analysis of the behavioral results has been published separately (Rousselet et al., 2003). Upright humans and animals were processed on average as efficiently (96.4% and 96.3%, respectively) and at the same speed (median RT: 368 and 371 msec, respectively). Furthermore, inversion produced a global decrease of accuracy that was very similar for both humans and animals (<2%). Inverted pictures led to significantly longer RT than upright pictures, and the inversion effect was reliably more pronounced for humans (23 msec on median RT) than for animals (9 msec on median RT).

In a first attempt to compare the processing speed of the different object categories, target ERPs in one task and for one orientation were compared with nontarget ERPs in the same task and the same orientation (global differential activity, Figure 1C, left). Early and large differences were found over posterior and frontal electrodes in both hemispheres for the two categorization tasks, with the strongest differential effects at lateral occipito-temporal sites. The peak latency of the global differential activity was very different across the four conditions, with up to 100-msec delay between upright human and inverted animal differential activity peaks (Figure 2, left).

The main differences between ERPs to upright target animals and nontargets had significant onsets around 150 msec, a result that constitutes a direct replication of previous studies performed in our laboratory (Fabre-Thorpe et al., 2001; VanRullen & Thorpe, 2001b; Thorpe et al., 1996) (Figures 3A and 4A). However, differential effects when faces were targets started even earlier, with significant effects as early as 100–130 msec (Figures 3C and 4B).

The weak effects of image inversion at the behavioral level were confirmed by ERP results. The differential activity for inverted animals started virtually at the same latency (\approx 150 msec) but developed with a shallower slope and reached lower amplitude than for upright animals

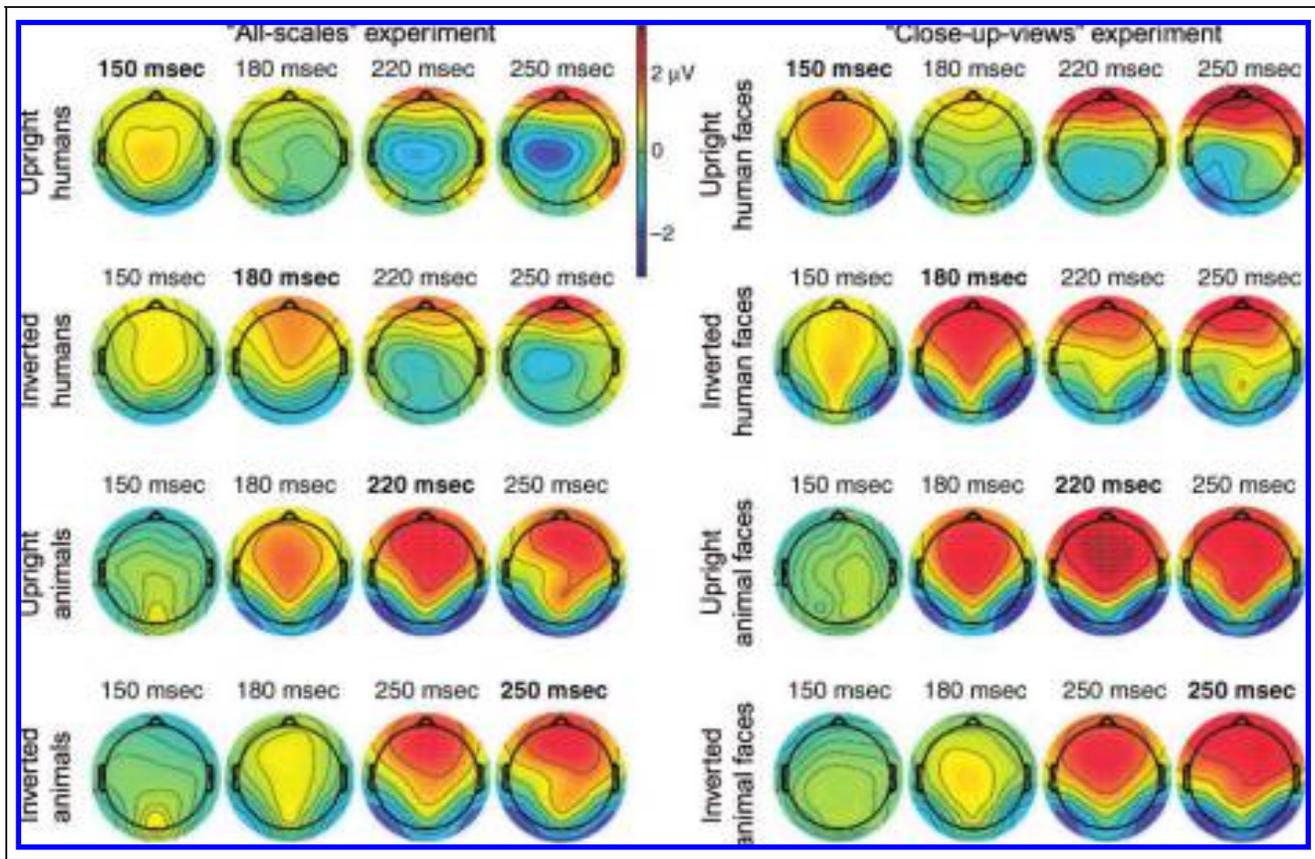


Figure 2. Two-dimensional linear interpolation maps of the global differential activities in each experiment and for each condition. The maps represent the signal recorded at the peak latencies (in **bold**) of the grand average occipital differential activity in the four conditions (upright humans = 150 msec, inverted humans = 180 msec, upright animals = 220 msec, inverted animals = 250 msec). Nose is pointing upward, and the black circle represents the line going from the nasion to the inion.

(Figure 3E). The differential activity onset for faces was slightly delayed by inversion (Figure 3F).

“Close-up-views” Experiment: Reducing Physical Differences

In addition to the main ERP differences for animals and humans, small but reliable ERP differences were found before 100 msec (Figure 4A and B). They appeared as early as 70–90 msec for upright and inverted animals and 50–60 msec for upright and inverted faces. We suspected that these very early differences might be because of uncontrolled low-level differences between the sets of target and nontarget images, as previously suggested for the categorization of natural images (Johnson & Olshausen, 2003, 2005; VanRullen & Thorpe, 2001b).

In a first attempt to reduce physical differences between target images, a new experiment was designed in which subjects ($n = 24$, 12 women and 12 men; mean age: 30 years; 9 of which participated in the first study) were required to categorize human or animal faces in pictures depicting close-up views of these targets (see Methods). At the behavioral level, the use of close-up views led to excellent performance levels, with slightly higher accuracy

and slightly longer RTs for both categories compared with the first experiment (Rousselle et al., 2003).

In the second experiment, the selection of close-up views of animal and human faces had several consequences at the ERP level. The key finding was that the very early differences recorded in the “all-scales” experiment (Figure 4B) for humans were not found anymore (Figure 4F). They were also reduced with animal faces with the main exception of the occipital midline electrode Oz (Figure 4A vs. E) that still showed very early differential activities associated with both upright and inverted animal faces. However, the large lateral occipito-temporal differential activities reported in the “all-scales” experiment were still present and even reached higher amplitudes in this second experiment (Figure 5). These ERP differences were larger than those reported for animals and vehicles by VanRullen and Thorpe (2001b). These differences appeared approximately at the same time as in the previous experiment reaching statistical significance at about 150 and 130 msec for upright animal and human faces, respectively. They had also similar posterior frontal distributions and latency peaks (Figure 2, right). Inversion had a very limited effect on these onset latencies (Figure 4E and F and Figure 5E and F). On the other hand, whereas

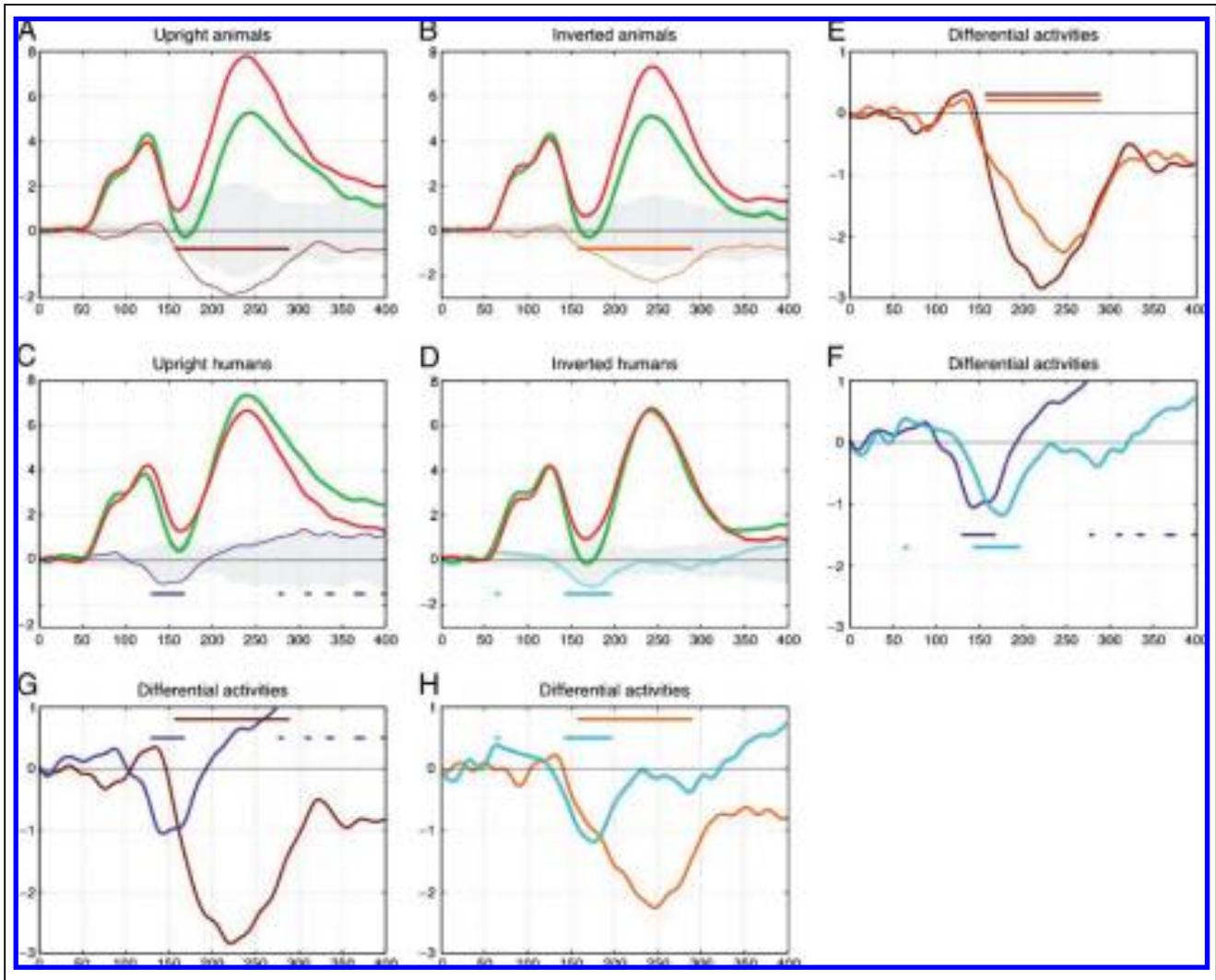


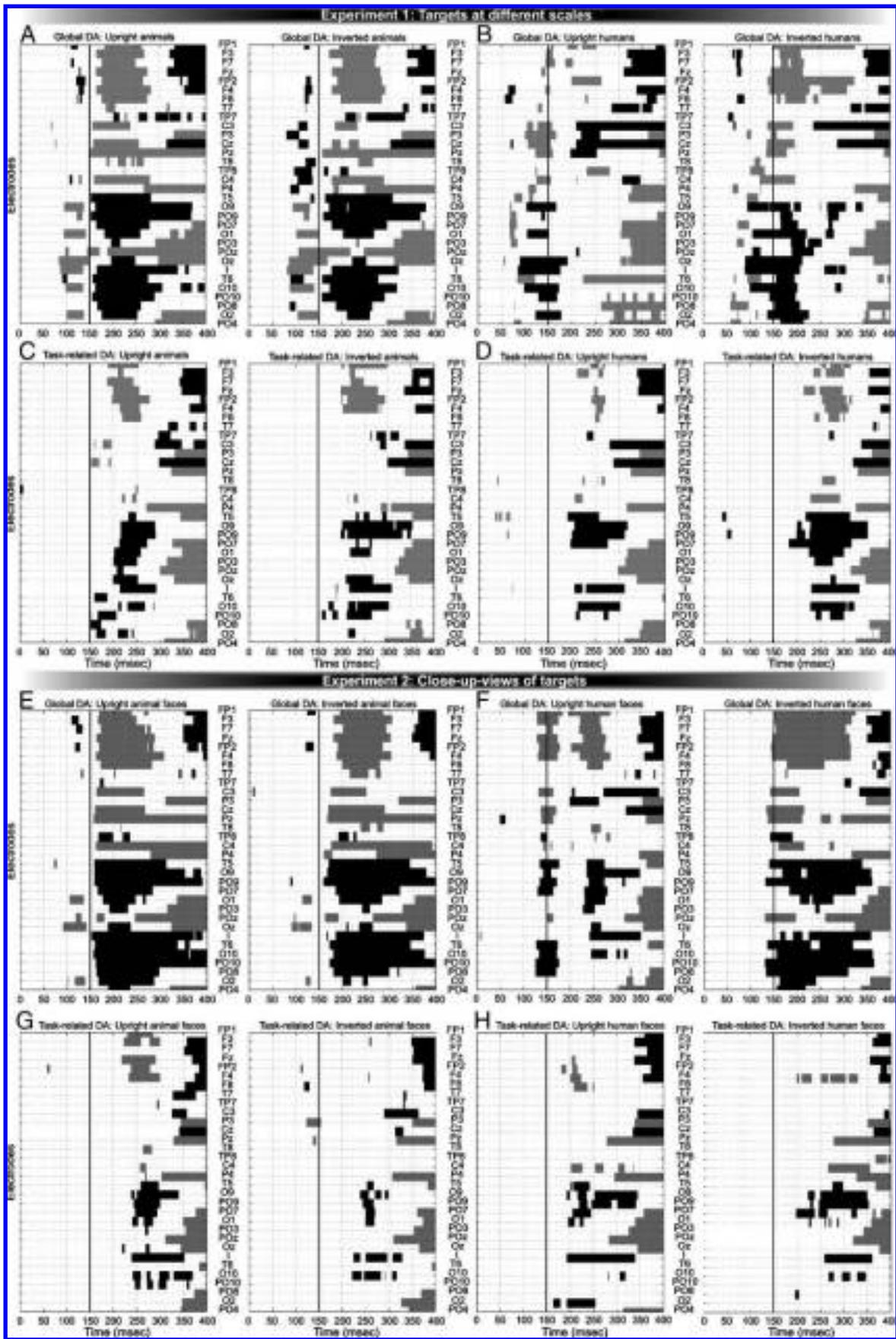
Figure 3. Global differential activities in the “all-scales” experiment. Each graph represents the grand average signal (in microvolts) recorded from the right posterior electrode PO10. This electrode was chosen for illustration because it showed the largest differential effects in amplitude. Detailed analyses for all electrodes can be found in Figure 4. For each target category, the ERPs are presented for upright (A and C) and inverted (B and D) stimuli. Target ERPs (in green) were computed from trials in which the indicated category was seen as target. Nontarget ERPs (in red) were computed from trials in which nontarget pictures were seen with the same orientation as the targets. They included neutral nontargets and pictures from the target category of the other task. The global differential activities were computed by subtracting nontarget ERPs from target ERPs separately for each category and each orientation (thin lines in A, B, C, and D). Upright animals (dark brown in A, E, and G), inverted animals (light brown in B, E, and H), upright humans (dark blue in C, F, and G), and inverted humans (light blue in D, F, and H). The confidence interval of the difference between the two conditions is shown for each differential activity by the gray area (A–D). It corresponds to a simulation of the difference that would be obtained by chance if the compared conditions (target ERP and distractor ERP) were sampled from the same population. When the difference between two conditions falls outside the confidence interval, the difference is significant at $p < .01$ and is indicated by the colored horizontal lines (A–D). Differential activities illustrated in A–D are also compared, with larger scales in E–H. The effect of inversion on global differential activities is shown for animals (E) and humans (F). The category effects (animals vs. humans) are shown for upright stimuli (G) and inverted stimuli (H).

animal faces inversion resulted in reduced amplitude of the differential activity, as well as a delayed peak and a reduced slope (Figure 5E), inverted human faces led to larger differences than upright ones (Figure 5F).

Estimating Processing Speed Independently of Visual Attributes

So far, we have described the results from the global differential activity, which typically reflects both task-

related and physical differences. The processing speed of the different categories was also estimated independently of their visual attributes by computing the task-related differential activity (Figure 1C, right), in which the ERP associated with a given image set seen at a given orientation and processed as targets was compared with the ERP associated with the same image set with the same orientation but seen as nontargets. This manipulation controlled for physical differences, because across subjects, the same pictures were seen as targets and



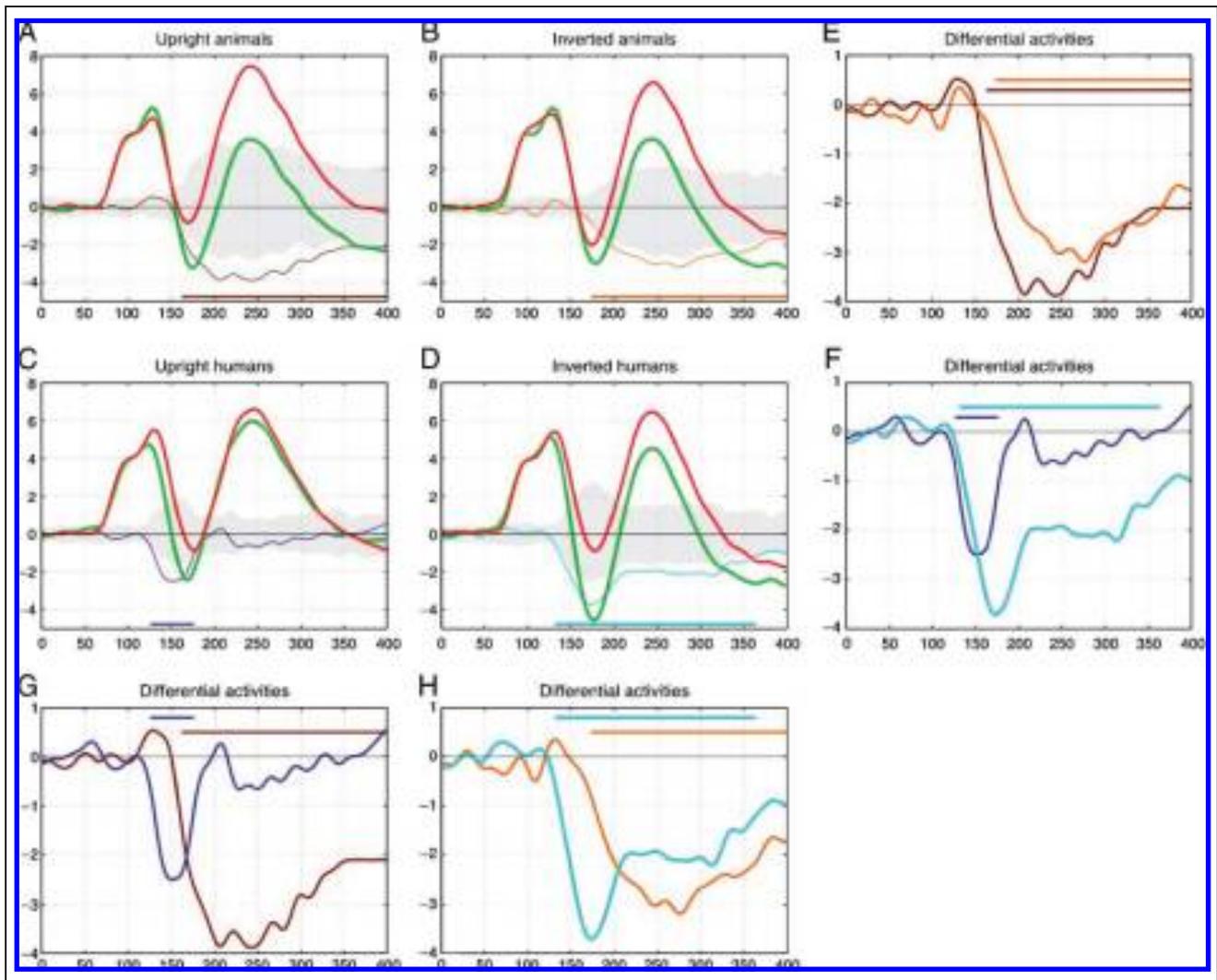


Figure 5. Global differential activities in the “close-up-views” experiment (same caption as Figure 3).

nontargets. The only differences that remained in the electrophysiological signal were because of task status and should thus give us an estimate of the time required to access task-related categorical information independently of physical differences. The results across all electrodes are depicted in Figure 4C and D for the “all-scales” experiment and Figure 4G and H for the “close-up-views” experiment. Task-related differential activities had very small amplitude compared with global differential activities (compare Figures 3 and 5 with Figures 6 and 7). With targets at all scales, the animal task was again found to affect ERP at around 150 msec (Figures 4C and 6E). This latency was almost unaffected

by inversion. On the contrary, the human face task did not induce a task-related differential activity before about 180 msec for upright pictures (Figures 4D and 6F). Task status had a slightly earlier effect on inverted human ERP at electrode PO7. With close-up views, the effects of task status on ERPs to both human and animal faces developed surprisingly late (Figure 4G and Figure 7), especially in regards to the high behavioral performance. With the exception of electrode O2, which showed an isolated “early” difference at about 160 msec for upright human faces, all other electrodes across all categories presented their earliest effects after 200 msec.

Figure 4. Latencies of the differential activities (DA) at all electrodes and in all conditions. Along the vertical axes, electrodes have been sorted from anterior (top) to posterior (bottom) sites. (A–D) “All-scales” experiment. (A and B) Latencies of the global differential activities computed by subtracting the ERP associated with all nontargets from the targets for animals (A) and humans (B), either upright or inverted. (C and D) Latencies of the task-related differential activities, when physical differences were removed. (E–H) “Close-up-views” experiment. Nonsignificant differences are represented in white. Significant differences ($p < .01$) for which the differential activity was negative (positive) are represented in black (gray).

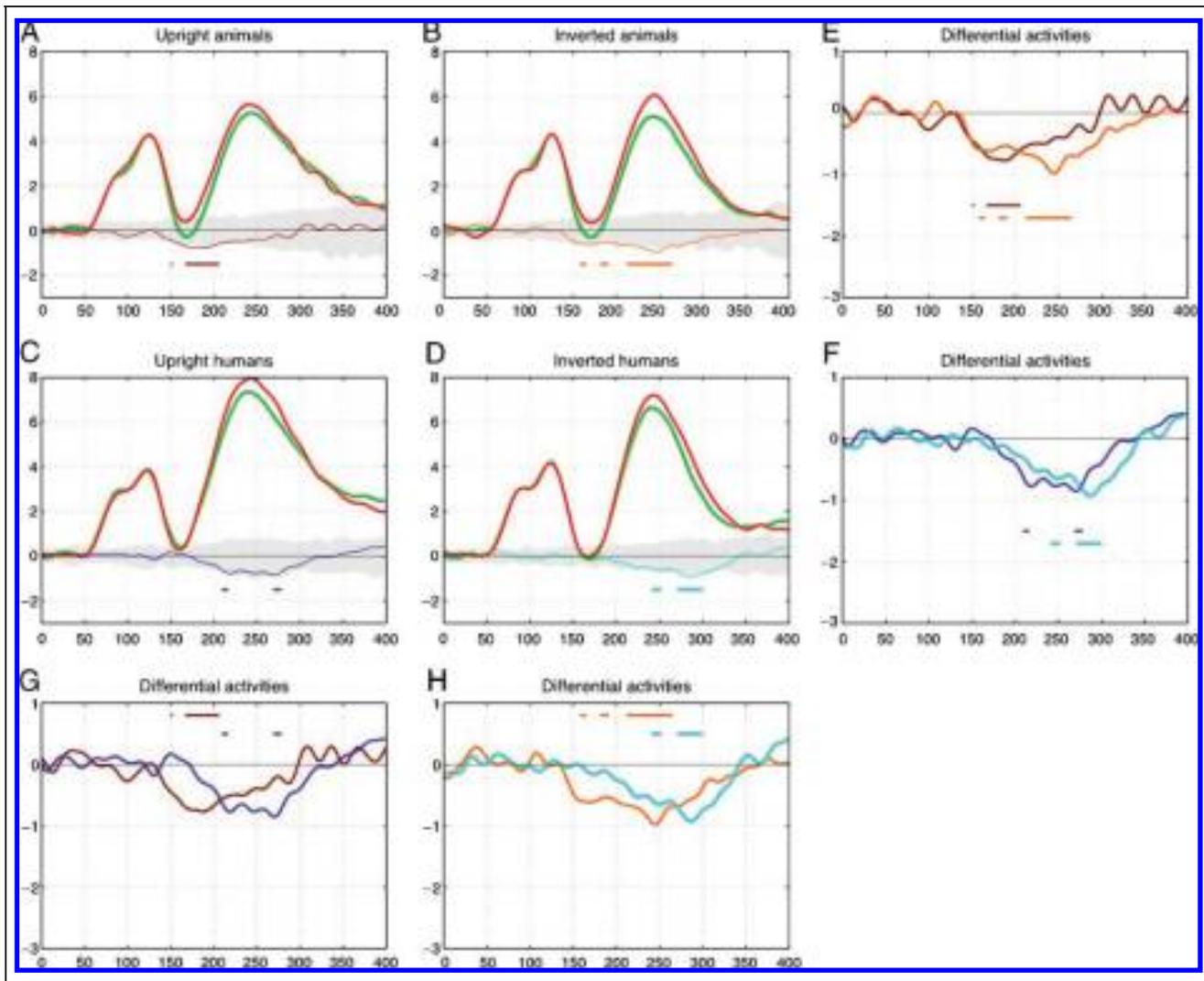


Figure 6. Task-related differential activities showing the effects of task status independently of physical differences at electrode PO10 in the “all-scales” experiment. These task-related differential activities (in microvolts) were computed by subtracting the ERP associated with images of a given category when processed as a nontarget (in red) from the ERP associated with the same image category when processed as target (in green). Other details are described in the caption of Figure 3.

DISCUSSION

In this article, we analyzed the cerebral activity associated with the processing of upright and inverted animal and human faces presented in natural images. We investigated how ERP differences can help tackle the question of the speed of underlying processes. We used a go/no-go behavioral paradigm, coupled with EEG recordings, that had been used successfully in previous experiments to determine the temporal constraints on the processing of other categories such as animals and means of transport (e.g., Macé, Thorpe, & Fabre-Thorpe, 2005; Rousselet et al., 2002; VanRullen & Thorpe, 2001a, 2001b). Two experiments were performed, one in which humans and animals appeared at various scales and positions and another one with only close-up views of human and animal faces. With very similar behavioral results obtained for humans and animals targets in both

experiments (Rousselet et al., 2003), we worked under the hypothesis that humans and human faces cannot be processed faster than other stimuli such as animals in natural scenes. For instance, similar behavioral results for the categorization of animals and vehicles were associated with similar ERP differences (VanRullen & Thorpe, 2001a, 2001b). Accordingly, we were expecting very similar time courses of ERP differences for the different animal and human target categories used in the present experiments.

Global Differential Activities

Despite similar behavioral results, the processing of humans or close-up views of human faces resulted in global differential activities that were weaker than for animals and animal faces but that developed with much

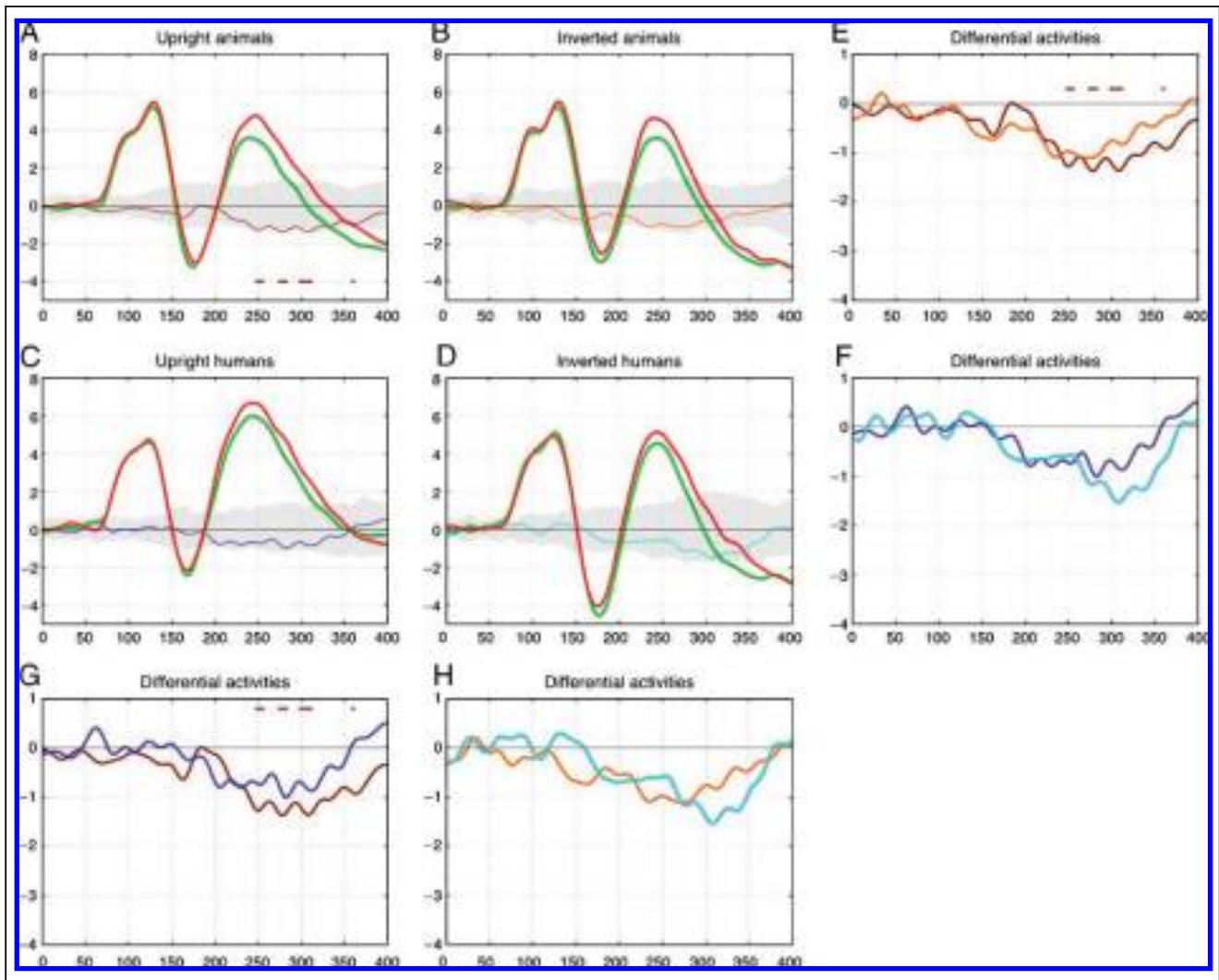


Figure 7. Task-related differential activities showing the effects of task status independently of physical differences at electrode PO10 in the “close-up-views” experiment (same caption as Figure 6).

shorter latencies (100–130 msec). These results seem to imply that humans and human faces can be processed faster than animals that would require at least an additional 20–50 msec. Furthermore, because of the very weak inversion effects on processing efficiency, this ability would rely on relatively view invariant representations. With animal targets, the differential activities recorded with inverted and upright stimuli had similar onsets but with a much shallower slope when associated with inverted stimuli. These results are compatible with a model of accumulation of evidence in which a neuronal population starts to fire at the same latency whatever the stimulus orientation but accumulates activity faster for prototypic (here upright) stimuli (Perrett, Oram, & Ashbridge, 1998). On the other hand, inverting human targets results either in a delay of the differential activity with very little effect on its amplitude (“all-scales” targets) or in a strong increase in amplitude without any delay (close-up views of human faces), an effect totally absent with inverted animal faces. This strong

effect of inversion on associated ERPs might be the hallmark of human face processing as often reported, although the origin of this phenomenon is still poorly understood (Itier, Latinus, & Taylor, 2006; Rousselet, Macé, et al., 2004; Rossion & Gauthier, 2002).

What underlying processes could give rise to this differential activation? It is useful to distinguish different potential causes, each characterizing activity at a particular level in the visual system. With two different sets of targets and nontarget images, differential activity could be linked, at one extreme, with the processing of very low-level stimulus characteristics. For example, a statistically significant difference between the two image sets might arise in the proportion of horizontal edges that can be processed in V1. At the other extreme, in the late stages of visual processing, neurons can be highly selective to objects (e.g., Tanaka, 1996; Perrett, Hietanen, Oram, & Benson, 1992). If one was to measure the average response of this population of neurons in response to two different image categories, there could also be a

strong differential response as shown by Vogels (1999) for tree and nontree pictures. In the first case, the differential activity can hardly be attributed to target categorization, but in the second case, the difference would have considerable significance for the task, because it would reflect the activity of selective populations of cells that could be involved in recognition and categorization.

There are some arguments to support the second hypothesis. The frontal differential activity has an opposite polarity to the one observed at lateral occipito-temporal sites. This bipolar arrangement can be seen clearly in Figure 2 of the present study (see also Rousselet et al., 2002; VanRullen & Thorpe, 2001b). The close similarity between the onset times of the occipital and frontal differential responses as well as the results from source analysis and functional magnetic resonance imaging studies are consistent with the idea that a considerable proportion of the differential response at frontal sites results from the activity of two dipoles located in the occipito-temporal extrastriate cortex (Fize, Fabre-Thorpe, Richard, Doyon, & Thorpe, 2005; Delorme et al., 2004; Fize, Boulanouar, Chatel, Ranjeva, Fabre-Thorpe, & Thorpe, 2000).

It is worth noticing that the activity of prefrontal areas could play a role in the late part of the differential effects (Rousselet, Thorpe, & Fabre-Thorpe, 2004). It is also interesting to note that in both experiments the global differential activity had larger amplitudes at electrodes located more medial than the temporal electrodes at which the N170, often associated with face processing, was reported to be maximal in the same experiments (Rousselet, Macé, et al., 2004). This result could imply the existence of at least partially different cortical sources for the two signals, a hypothesis that deserves to be further explored.

Task-related Differential Activity

Whereas global differential activity might potentially reflect a mixture of low- and high-level stimulus properties, task-related differential activity avoids this bias and highlights the additional processing devoted to an image because of its task pertinence.

The methodology used by VanRullen and Thorpe (2001b) and used again here provides one way of tackling this issue. By switching between two different target categories, the same images can be presented either as targets or nontargets. If a difference still exists under these conditions, it is clear that no simple low-level difference between the images could explain the effects because the two image sets are physically identical. The differential responses computed in this experiment showed clear onset differences depending on experimental conditions. In the standard “animal/non-animal” task in which animals are presented at all scales, clear differential effects emerged at about 150 msec for both upright and inverted stimuli. This result replicates

the data reported by VanRullen and Thorpe and generalizes the finding to inverted animals. In this case, information related to the animal category must have started to be encoded around 150 msec or even before as discussed later. However, because the task-related differences have a small amplitude compared with the global differences (both in the present study and in the study of VanRullen & Thorpe, 2001b), it is clear that global differences also include the effect of physical differences between image sets.

In contrast to the results with animals (or vehicles as reported by VanRullen & Thorpe 2001b), the onset latency of the differential activity for humans shown at all scales did not become significant until about 180 msec after stimulus onset. Furthermore, when shown as close-up views, human faces induced a differential effect that started even later, with the earliest significant effects appearing not until over 200 msec. With close-up views, this effect extended to animal faces with differential onsets in the same range as human faces. Other studies have reported an absence of task-related effect on early face-sensitive ERP components (N170: Carmel & Bentin, 2002; Séverac-Cauquil, Edmonds, & Taylor, 2000; N200: Puce, Allison, & McCarthy, 1999), although a modulation of the N170 can be observed when demanding visual tasks require directing attention away from face stimuli (Eimer, 2000a, 2000b). The absence of task effect before 180–200 msec in the present study might thus result from the “relative simplicity” of the tasks used. However, independent of the reasons for an absence of early task effect, our results raise the question of the relationship between onsets of differential activities and behavioral RTs.

If onset latencies of differential activity were good predictors of behavioral RT, one could expect subjects to show about 50 msec faster RTs for human targets based on global differential activities. On the other hand, one could predict they would show at least 30–50 msec longer RTs based on task-related differential activities. But behavioral results have clearly shown that accuracy, mean RTs, and minimal RTs were very similar for animal and human targets (Table 1; Rousselet et al., 2003). Why is there no link between onset of the differential activity and behavioral RTs?

No Correlation between Differential Activity Onsets and Behavioral Reaction Times

With close-up views, subjects can perform the challenging visual task without showing clear signs of task-related activity in ERP recordings. Thus, it is clear that the onset of task-related ERP differences sets an upper limit on the time required to extract a certain type of visual information. This result might be explained by a “by-default” processing that is deeper for some categories of targets, such as human faces, for instance. The inferotemporal cortex (IT) contains large populations of “face-selective” cells. Let us suppose that these neurons

have responses that are relatively “hard-wired,” in that they will respond to the presence of a face whether the face is a pertinent task stimulus or not. In such a case, switching the status of face from target to distractor will have little or no effect on the magnitude of the neuronal response and no “task-related” differential activity would be observed, and yet, neurons could still perfectly signal whether or not the scene contains a human face. The possibility that responses to faces are particularly robust is reinforced by the recent demonstration that in regions of the monkey temporal lobe with strong functional magnetic resonance imaging responses to faces, virtually all the cells within those regions are highly selective for faces (Tsao, Freiwald, Tootell, & Livingstone, 2006). Furthermore, in a visual categorization task, there is some evidence that the decision mechanisms are located outside the visual processing pathway itself, such as in prefrontal cortex (Freedman et al., 2001, 2003). There is also some evidence that IT cells build their specificity with training and tend to respond to specific elements of visual stimuli that are pertinent to task performance (Sigala & Logothetis, 2002). Our suggestion is that, because of expertise, special target categories such as human faces are “by-default” processed further than other object categories. In such case, there will be no task modulation of responsiveness within the visual system itself, with the result that no task-related differential effect would be visible. The “depth of processing” for faces might be developmentally driven by their intrinsic social interest (Rousselet, Macé, et al., 2004; Carmel & Bentin, 2002; Séverac-Cauquil et al., 2000; Puce et al., 1999; Rossion et al., 1999; Young, Hellawell, & De Haan, 1988). This hypothesis could be tested within an expertise framework (Tanaka & Curran, 2001), because it predicts that with intensive training with one category of stimuli, task-related differential activities should decrease in amplitude. With close-up views of faces, the by-default processing would be extended to animal faces because of strong basic similarities and explain the long latencies and small amplitude of the task-related differential activity on animal face targets. On the other hand, the by-default processing would not be as deep for the various images of animals seen at different scales resulting in the 150-msec-onset task-related differential activity that is seen in the “all-scales” experiment.

Similarly to the task-related differences, results using the global ERP differences were at odds with the behavioral data. Indeed, global differential activities for human targets in both experiments appeared with shorter latencies and peaked with smaller amplitude than for animals. Together, such results should imply faster behavioral RTs on human targets. There are different ways to interpret these results. First, one might consider the existence of a floor effect. With a categorization task, it might not be possible to trigger the go/no-go hand movement any faster than with animals. This is in keep-

ing with results showing that intensive training with a given set of images fails to shorten the latency of the earliest behavioral responses (Fabre-Thorpe et al., 2001). A real shift of the earliest go responses toward shorter latencies was only observed in a modified task in which subjects had to respond to a single totally predictable target (Delorme et al., 2004). Even with this very simple task, the latency gain barely reached 30 msec! To avoid this floor effect, subjects might be asked to respond with a faster effector. For instance, in a recent experiment, subjects had to perform a forced-choice animal categorization task using ocular movements (Kirchner & Thorpe, 2006). They were very fast with earliest responses triggered 120–130 msec after stimulus onset. Using a similar task, a new pilot experiment suggests an RT advantage of about 20–30 msec for human faces over animal faces (Kirchner, Crouzet, & Thorpe, personal communication, 2006). This last result is in keeping with a recent study showing that IT cells start to fire on average 20 msec earlier for human faces compared with animal faces (Kiani, Esteky, & Tanaka, 2005). Thus, it seems that the difference between animal and human faces that was almost invisible in our go/no-go behavioral results (Rousselet et al., 2003) may be visible in the latency of the (a) global differential activity, (b) single-unit responses in IT, and (c) saccadic choice responses. If this interpretation is correct, then it implies that behavioral responses in the go/no-go task reflect, at least in part, something different than the response latencies of IT cells (see further discussion on this topic in Bacon-Macé et al., in press).

An alternative way to interpret our data is to address the question of why there should be a correlation between visual ERP differences and behavioral RTs at all. In two recent ERP studies, Johnson and Olshausen (2003, 2005) provided compelling evidence, using single-trial analyses, that the ERP differential activity starting around 150 msec (corresponding to the global differences reported in this article) was not correlated with RTs. Based on this result, they concluded that these ERP differences were not related to decision making in the context of the task and almost certainly reflect physical differences between image sets. Although their conclusion might seem perfectly legitimate, it relies on the assumption that neuronal activity related to high-level categorization should necessarily be reflected in ERP differential activity, and therefore, one should find a correlation between behavioral RT and latencies of ERP differences. However, this assumption is called into question by a recent study showing that the response latencies of neurons in the anterior inferotemporal cortex are effectively completely independent from behavioral RTs (DiCarlo & Maunsell, 2005). This does not imply that object-selective neurons in high-level areas of the ventral visual stream are not involved in the categorization process itself. But they might not be involved in the explicit decisional process, which might take place in

the prefrontal cortex (Freedman et al., 2003). Frontal areas also seem to be the main source of RT variability during rapid object categorization (Gerson, Parra, & Sajda, 2005). Following this idea, a link between RT and onset latency of ERP effects should not be regarded as the necessary hallmark for high-level visual processing leading to categorization.

An alternative explanation takes into consideration the crucial role that can be played by top-down influences in a block paradigm as used in the present study (Figure 8). Suppose that to detect with high speed and accuracy any one of a large number of animal forms, some sort of top-down priming of neurons selective for particular animal features was required (Figure 8A). In neurons selective to the pertinent visual features, such top-down priming would induce a stronger response when features are present. This enhanced activation could be forwarded to a later decision stage. The increased response when a target was present in the scene might be visible at the level of the global differential

activity because the amount of neural activity would be increased (e.g., Chelazzi, Duncan, Miller, & Desimone, 1998). However, in this case, changing the target category from “animal” to something else (“means of transport” as in the study of VanRullen & Thorpe, 2001b, or “human” as in the “all-scales” experiment of the present study) would have the effect of modifying the priming effect and revealing a “task-related” differential effect. We could propose a variation of this processing model by introducing an intermediate population of neurons between the high-level feature detectors (IT?) and the decision level (Figure 8B). This population could consist of a relatively small number of neurons, possibly located in the prefrontal cortex (Freedman et al., 2001, 2003) and receiving most of their inputs from IT cells. Following this hypothesis, switching between categorization tasks would not be done by massively shifting top-down activation from one set of high-level feature detectors to another but by switching the activation from a small subset of cells in this intermedi-

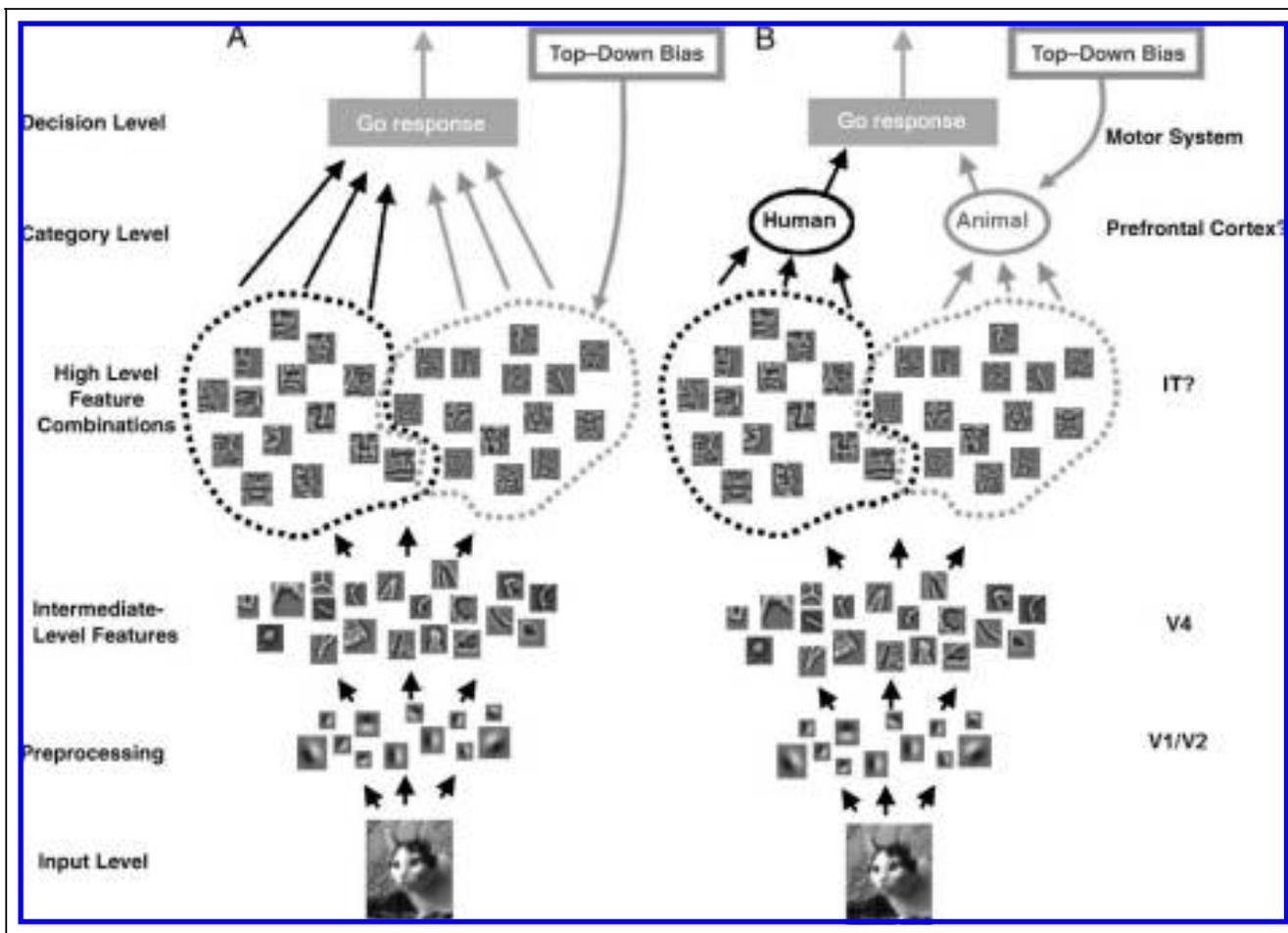


Figure 8. Two different ways in which top-down presetting could be used to switch between tasks with either human or animal faces as targets. In both cases, the basic processing architecture is the same, with preprocessing in V1/V2, the coding of intermediate features in areas such as V4, and the coding of category specific high-level feature combinations in areas such as IT. (A) The top-down biasing directly affects the high-level feature level, causing substantial changes in the population response. (B) In contrast, there is a specific category level (possibly in prefrontal cortex) that would allow top-down biasing to switch between tasks without affecting the activity in high-level visual areas.

ate population to another one. The reduced number of cells involved in this operation could well explain why no early differences could be found in task-related differential activity before 180–200 msec for faces. The extensive experience we have with faces could increase the role of this intermediate processing stage of processing and explain the discrepancies observed between animals and faces.

Note that all these processing strategies would allow the subjects to perform the task reliably, but only when a top–down priming strategy is used would one expect to see changes in the responsiveness of neurons within the visual system as a function of the target class. However, it is interesting to note that with upright faces, the early global differential effects tend to be considerably less ample and less long lasting than for animals, a result that might fit with the idea of a more hard-wired “automatic” face processing.

Object Processing within 100–130 msec

In this context, we would like to argue that the differential activities recorded for humans and animals before 150 msec and as early as 120–130 msec do not necessarily reflect the processing of irrelevant low-level physical differences, but might, in fact, be the signature of the early activation of high-level units coding for diagnostic properties in the image. To illustrate this point, we will refer again to the recent study by Kirchner and Thorpe (2006). In this experiment, subjects had to perform a forced-choice animal categorization task using ocular movements. Two natural scenes were flashed simultaneously on each side of a fixation cross, and subjects had to move their eyes as fast as they could toward the image that contained an animal. The median RT was 230 msec, with 90% of correct responses and the earliest responses seen with latencies as short as 120–130 msec. Such very short behavioral response latencies might suggest that some sort of low-level image properties were used to perform the task. However, Kirchner and Thorpe were unable to find any statistical parameter (e.g., energy, contrast, skewness, kurtosis) in the images that could be used to distinguish reliably between target and nontarget images. It has also been suggested that the earliest evidence for coarse face processing might be found at around 120 msec (Itier & Taylor, 2004; Linkenkaer-Hansen et al., 1998). In keeping with this hypothesis, recent source analysis on ERP data has revealed that the fusiform gyrus, an area of the ventral pathway involved in high-level object recognition, can be activated under 110 msec after stimulus onset (Di Russo, Martinez, Sereno, Pitzalis, & Hillyard, 2002; Martinez et al., 2001). It has also been suggested that such early activities might not be as “early” as generally thought because visual mechanisms in this time window might well be influenced by feedback from prefrontal cortex (Foxy & Simpson, 2002).

This does not mean that object categorization in natural scenes is fully complete in 120–130 msec. Indeed, a significant difference between two ERP waveforms does not imply that the visual task is completed. But, by 120–130 msec after stimulus onset, response of neuronal populations might be already biased, leading to coarse categorization of objects. A similar idea has been proposed in which a coarse processing of the objects in the scene is used to create a saliency map that could orient more detailed visual processing (Macé et al., 2005; Bar, 2004).

At this point, one might consider that “high-level” categorization does not necessarily imply that high-level representations are used to perform the task. For instance, it has been shown that “midlevel” representations can perfectly be used to perform relatively simple classifications, such as the detection of faces or animals in natural scenes, or the categorization of the gist of the scene (Torrallba & Oliva, 2003; Ullman, Vidal-Naquet, & Sali, 2002; Oliva & Torrallba, 2001). Such “midlevel” representations might be used as diagnostic features in our task, allowing subjects to respond for the presence of complex objects (Schyns, 1998). Faces might be differentiable from other nonface objects based on conjunctions of specific orientations, color pigmentation (Vuong, Peissig, Harrison, & Tarr, 2005), and spatial frequency spectrum (Rousselet, Husk, Bennett, & Sekuler, 2005). Because these kinds of “features” might well be processed in areas V4–TEO of the ventral pathway and activated by a feed-forward wave of activation, this strengthens the hypothesis of an early “high-level” process of objects in natural scenes. Indeed, as suggested by Figure 8A, it may be possible to initiate category-specific behavioral responses without explicitly using categorical coding. The solution would be to use top–down biasing to preactivate large numbers of units selective to high-level feature combinations diagnostic for stimuli characteristic of the target category. In this case, a decision mechanism that simply monitors the amount of activation in this population could trigger a category-specific behavioral response without the need for explicit categorization.

Overall, this study reveals potential limitations in using task-related ERP differential activities to understand object processing speed in general. However, alternative strategies exist that involve systematically exploring the stimulus space and the task space. By varying parametrically stimulus properties and task requirements, one might eventually construct a “hypercube” description of how those properties affect behavior and brain responses and, finally, how behavior maps onto brain activity. Parametric studies can be used to track over time correlations between behavioral accuracy and brain response amplitude or between image properties and brain response amplitude (e.g., Philiastides & Sajda, 2006; Tanskanen, Nasanen, Montez, Paallysaho, & Hari, 2005; Jemel et al., 2003; Schyns,

Jentsch, Johnson, Schweinberger, & Gosselin, 2003). Some task modulations can also be specific to some dimensions of the stimuli, for instance, specific frequency bands (Goffaux, Jemel, Jacques, Rossion, & Schyns, 2003) and specific task-related information content (Joyce, Schyns, Gosselin, Cottrell, & Rossion, 2006). Finally, modulations of brain activity, as well as correlations between behavior and brain activity, can be described at finer scales by decomposing signals in time frequencies and working in source space to isolate the time courses of different cortical generators (Makeig, Debener, Onton, & Delorme, 2004; Makeig et al., 2002). We are currently exploring several of those alternative strategies.

Acknowledgments

We acknowledge support from the CNRS and the Integrative and Computational Neuroscience ACl. The first two authors were supported by PhD grants from the French government. G. A. R. also acknowledges support from a CIHR fellowship program. The CNRS Research Ethical Board (COPE) approved this work. We thank Caitlin R. Sternberg and Anne-Sophie Paroissien for their help in running subjects in experiments 1 and 2, respectively. We also thank Nadège M. Bacon-Macé for programming stimulus presentation in experiment 2 and Rufin VanRullen for several brainstorming sessions about these data.

Reprint requests should be sent to Guillaume A. Rousselet, Department of Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, UK, or via e-mail: g.rousselet@psy.gla.ac.uk.

REFERENCES

- Bacon-Macé, N., Kirchner, H., Fabre-Thorpe, M., & Thorpe, S. J. (in press). Effects of task requirements on rapid natural scene processing: From common sensory encoding to distinct decisional mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*.
- Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*, 617–629.
- Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*, 551–565.
- Carmel, D., & Bentin, S. (2002). Domain specificity versus expertise: Factors influencing distinct processing of faces. *Cognition*, *83*, 1–29.
- Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, *80*, 2918–2940.
- Debruille, J. B., Guillem, F., & Renault, B. (1998). ERPs and chronometry of face recognition: Following-up Seeck et al. and George et al. *NeuroReport*, *9*, 3349–3353.
- Delorme, A., Rousselet, G. A., Macé, M. J., & Fabre-Thorpe, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Brain Research, Cognitive Brain Research*, *19*, 103–113.
- DiCarlo, J. J., & Maunsell, J. H. (2005). Using neuronal latency to determine sensory-motor processing pathways in reaction time tasks. *Journal of Neurophysiology*, *93*, 2974–2986.
- Di Russo, F., Martinez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping*, *15*, 95–111.
- Eimer, M. (2000a). Effects of face inversion on the structural encoding and recognition of faces. Evidence from event-related brain potentials. *Brain Research, Cognitive Brain Research*, *10*, 145–158.
- Eimer, M. (2000b). Attentional modulations of event-related brain potentials sensitive to faces. *Cognitive Neuropsychology*, *17*, 103–116.
- 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.
- Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is “special” about face perception? *Psychological Review*, *105*, 482–498.
- 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.
- Fize, D., Fabre-Thorpe, M., Richard, G., Doyon, B., & Thorpe, S. J. (2005). Rapid categorization of foveal and extrafoveal natural images: Associated ERPs and effects of lateralization. *Brain and Cognition*, *59*, 145–158.
- Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from v1 to frontal cortex in humans. A framework for defining “early” visual processing. *Brain Research, Experimental Brain Research*, *142*, 139–150.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, *291*, 312–316.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *Journal of Neuroscience*, *23*, 5235–5246.
- George, N., Jemel, B., Fiori, N., & Renault, B. (1997). Face and shape repetition effects in humans: A spatio-temporal ERP study. *NeuroReport*, *8*, 1417–1423.
- Gerson, A. D., Parra, L. C., & Sajda, P. (2005). Cortical origins of response time variability during rapid discrimination of visual objects. *Neuroimage*, *28*, 342–353.
- Goffaux, V., Jemel, B., Jacques, C., Rossion, B., & Schyns, P. G. (2003). Erp evidence for task modulations on face perceptual processing at different spatial scales. *Cognitive Science*, *27*, 313–325.
- Halgren, E., Raji, T., Marinkovic, K., Jousmaki, V., & Hari, R. (2000). Cognitive response profile of the human fusiform face area as determined by MEG. *Cerebral Cortex*, *10*, 69–81.
- Halit, H., de Haan, M., & Johnson, M. (2000). Modulation of event-related potentials by prototypical and atypical faces. *NeuroReport*, *11*, 1871–1875.
- Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*, 177–180.
- Hillyard, S. A., & Munte, T. F. (1984). Selective attention to color and location: An analysis with event-related brain potentials. *Perception & Psychophysics*, *36*, 185–198.
- Itier, R. J., Latinus, M., & Taylor, M. J. (2006). Face, eye and object early processing: What is the face specificity? *Neuroimage*, *29*, 667–676.
- Itier, R. J., & Taylor, M. J. (2004). N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cerebral Cortex*, *14*, 132–142.
- Jeffreys, D. (1996). Evoked potential studies of face and object processing. *Visual Cognition*, *3*, 1–38.
- Jemel, B., Schuller, A. M., Cheref-Khan, Y., Goffaux, V., Crommelinck, M., & Bruyer, R. (2003). Stepwise emergence of the face-sensitive N170 event-related potential component. *NeuroReport*, *14*, 2035–2039.

- Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision*, 3, 499–512.
- Johnson, J. S., & Olshausen, B. A. (2005). The earliest EEG signatures of object recognition in a cued-target task are postsensory. *Journal of Vision*, 5, 299–312.
- Joyce, C. A., Schyns, P. G., Gosselin, F., Cottrell, G. W., & Rossion, B. (2006). Early selection of diagnostic facial information in the human visual cortex. *Vision Research*, 46, 800–813.
- Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, 3, 759–763.
- Kiani, R., Esteky, H., & Tanaka, K. (2005). Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. *Journal of Neurophysiology*, 94, 1587–1596.
- Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, 46, 1762–1776.
- Linkenkaer-Hansen, K., Palva, J. M., Sams, M., Hietanen, J. K., Aronen, H. J., & Ilmoniemi, R. J. (1998). Face-selective processing in human extrastriate cortex around 120 ms after stimulus onset revealed by magneto- and electroencephalography. *Neuroscience Letters*, 253, 147–150.
- Macé, M. J.-M., Thorpe, S. J., & Fabre-Thorpe, M. (2005). Rapid categorization of achromatic natural scenes: How robust at very low contrasts? *European Journal of Neuroscience*, 21, 2007–2018.
- Makeig, S., Debener, S., Onton, J., & Delorme, A. (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences*, 8, 204–210.
- 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.
- Martinez, A., DiRusso, F., Anllo-Vento, L., Sereno, M. I., Buxton, R. B., & Hillyard, S. A. (2001). Putting spatial attention on the map: Timing and localization of stimulus selection processes in striate and extrastriate visual areas. *Vision Research*, 41, 1437–1457.
- Mouchetant-Rostaing, Y., Giard, M. H., Bentin, S., Aguera, P. E., & Pernier, J. (2000). Neurophysiological correlates of face gender processing in humans. *European Journal of Neuroscience*, 12, 303–310.
- Mouchetant-Rostaing, Y., Giard, M. H., Delpuech, C., Echallier, J. F., & Pernier, J. (2000). Early signs of visual categorization for biological and non-biological stimuli in humans. *NeuroReport*, 11, 2521–2525.
- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42, 145–175.
- Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, 335, 23–30.
- Perrett, D. I., Oram, M. W., & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: An account of generalisation of recognition without mental transformations. *Cognition*, 67, 111–145.
- Philiastides, M. G., & Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex*, 16, 509–518.
- Pizzagalli, D., Regard, M., & Lehmann, D. (1999). Rapid emotional face processing in the human right and left brain hemispheres: An ERP study. *NeuroReport*, 10, 2691–2698.
- Puce, A., Allison, T., & McCarthy, G. (1999). Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials. *Cerebral Cortex*, 9, 445–458.
- Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology*, 12, 162–168.
- Rossion, B., Campanella, S., Gomez, C. M., Delinte, A., Debatisse, D., Liard, L., et al. (1999). Task modulation of brain activity related to familiar and unfamiliar face processing: An ERP study. *Clinical Neurophysiology*, 110, 449–462.
- Rossion, B., & Gauthier, I. (2002). How does the brain process upright and inverted faces? *Behavioral and Cognitive Neuroscience Reviews*, 1, 62–74.
- Rossion, B., Joyce, C. A., Cottrell, G. W., & Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *NeuroImage*, 20, 1609–1624.
- Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience*, 5, 629–630.
- Rousselet, G. A., Husk, J. S., Bennett, P. J., & Sekuler, A. B. (2005). Spatial scaling factors explain eccentricity effects on face ERPs. *Journal of Vision*, 5, 755–763.
- Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision*, 3, 440–455.
- Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2004). Animal and human faces in natural scenes: How specific to human faces is the N170 ERP component? *Journal of Vision*, 4, 13–21.
- Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). Processing of one, two or four natural scenes in humans: The limits of parallelism. *Vision Research*, 44, 877–894.
- Schendan, H. E., Ganis, G., & Kutas, M. (1998). Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms. *Psychophysiology*, 35, 240–251.
- Schyns, P. G. (1998). Diagnostic recognition: Task constraints, object information, and their interactions. *Cognition*, 67, 147–179.
- Schyns, P. G., Jentzsch, I., Johnson, M., Schweinberger, S. R., & Gosselin, F. (2003). A principled method for determining the functionality of brain responses. *NeuroReport*, 14, 1665–1669.
- Seeck, M., Michel, C. M., Mainwaring, N., Cosgrove, R., Blume, H., Ives, J., et al. (1997). Evidence for rapid face recognition from human scalp and intracranial electrodes. *NeuroReport*, 8, 2749–2754.
- Séverac-Cauquil, A., Edmonds, G. E., & Taylor, M. J. (2000). Is the face-sensitive N170 the only ERP not affected by selective attention? *NeuroReport*, 11, 2167–2171.
- Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, 415, 318–320.
- Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science*, 12, 43–47.
- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, 19, 109–139.
- Tanskanen, T., Nasanen, R., Montez, T., Paallysaho, J., & Hari, R. (2005). Face recognition and cortical responses show similar sensitivity to noise spatial frequency. *Cerebral Cortex*, 15, 526–534.
- Taylor, M. J., Edmonds, G. E., McCarthy, G., & Allison, T. (2001). Eyes first! Eye processing develops before face processing in children. *NeuroReport*, 12, 1671–1676.

- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.
- Thorpe, S. J., & Fabre-Thorpe, M. (2001). Seeking categories in the brain. *Science*, *291*, 260–263.
- Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network*, *14*, 391–412.
- Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*, 670–674.
- Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, *5*, 682–687.
- VanRullen, R., & Thorpe, S. J. (2001a). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects. *Perception*, *30*, 655–668.
- VanRullen, R., & Thorpe, S. J. (2001b). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*, 454–461.
- Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys. Part 2: Single-cell study. *European Journal of Neuroscience*, *11*, 1239–1255.
- Vuong, Q. C., Peissig, J. J., Harrison, M. C., & Tarr, M. J. (2005). The role of surface pigmentation for recognition revealed by contrast reversal in faces and greebles. *Vision Research*, *45*, 1213–1223.
- Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd ed.). San Diego, CA: Academic Press.
- Yamamoto, S., & Kashikura, K. (1999). Speed of face recognition in humans: An event-related potentials study. *NeuroReport*, *10*, 3531–3534.
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*, 141–145.
- Young, A. W., Hellowell, D., & De Haan, E. H. (1988). Cross-domain semantic priming in normal subjects and a prosopagnosic patient. *Quarterly Journal of Experimental Psychology, Series A*, *40*, 561–580.