

# Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes

**Guillaume A. Rousselet**

Centre de Recherche Cerveau et Cognition,  
CNRS-UPS UMR 5549, Toulouse, France



**Marc J.-M. Macé**

Centre de Recherche Cerveau et Cognition,  
CNRS-UPS UMR 5549, Toulouse, France



**Michèle Fabre-Thorpe**

Centre de Recherche Cerveau et Cognition,  
CNRS-UPS UMR 5549, Toulouse, France



Object categorization can be extremely fast. But among all objects, human faces might hold a special status that could depend on a specialized module. Visual processing could thus be faster for faces than for any other kind of object. Moreover, because face processing might rely on facial configuration, it could be more disrupted by stimulus inversion. Here we report two experiments that compared the rapid categorization of human faces and animals or animal faces in the context of upright and inverted natural scenes. In Experiment 1, the natural scenes contained human faces and animals in a full range of scales from close-up to far views. In Experiment 2, targets were restricted to close-ups of human faces and animal faces. Both experiments revealed the remarkable object processing efficiency of our visual system and further showed (1) virtually no advantage for faces over animals; (2) very little performance impairment with inversion; and (3) greater sensitivity of faces to inversion. These results are interpreted within the framework of a unique system for object processing in the ventral pathway. In this system, evidence would accumulate very quickly and efficiently to categorize visual objects, without involving a face module or a mental rotation mechanism. It is further suggested that rapid object categorization in natural scenes might not rely on high-level features but rather on features of intermediate complexity.

**Keywords:** rapid visual categorization, human performance, natural scenes, human faces, animals and animal faces, inversion effect, mental rotation, configural processing

## Introduction

Recent biologically plausible models of object visual processing have emphasized that much of the computation underlying scene categorization might rely on essentially parallel feed-forward mechanisms (Riesenhuber & Poggio, 2000; Thorpe & Imbert, 1989; VanRullen, Gautrais, Delorme, & Thorpe, 1998; Wallis & Rolls, 1997). These suggestions are supported by the finding that in humans, a differential brain activity develops between target and distractor trials from 150 ms in various categorization tasks using natural images (Thorpe, Fize, & Marlot, 1996; Rousselet, Fabre-Thorpe, & Thorpe, 2002). This processing time seems to correspond to an optimum, because it cannot be speeded up even with highly familiar natural images (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001). Moreover, when considering the number of processing steps between the retina and the high-level visual cortical areas of the ventral pathway, this 150-ms delay challenges most models of visual processing because it appears compatible only with a first feed-forward wave of information processing (Thorpe & Fabre-Thorpe, 2001). Thus, this delay appears as the minimal processing time from which discriminability between two categories of stimuli can

develop. However, even if the human visual system is able to extract a great deal of information in under 150 ms, visual perception does not end up after a first pass through the visual system that might not even allow access to a conscious representation (Dehaene & Naccache, 2001; Thorpe, Gegenfurtner, Fabre-Thorpe, & Bulthoff, 2001); in many cases, reaching a decision will require more time consuming detailed analysis.

In parallel, growing evidence suggests that faces may have a special computational status (Farah, Wilson, Drain, & Tanaka, 1998; Kanwisher, 2000; but see Tarr & Gauthier, 2000) that would allow them to be processed more efficiently and even faster than any other class of objects. However, the precise speed of face processing remains a controversial question. Indeed, very rapid categorization of isolated and relatively homogenous face stimuli has been reported in the literature, with brain activity onsets appearing as early as 50-80 ms poststimulus (George, Jemel, Fiori, & Renault, 1997; Mouchetant-Rostaing, Giard, Bentin, Aguera, & Pernier, 2000a, 2000b; Seeck et al., 1997). These findings have been disputed as other groups have reported early face processing in the 100-130-ms latency range (Debruille, Guillem, & Renault, 1998; Halgren, Raij, Marinkovic, Jousmaki, & Hari, 2000; Halit, de Haan, & Johnson,

2000; Irier & Taylor, 2002; Linkenkaer-Hansen et al., 1998; Pizzagalli, Regard, & Lehmann, 1999; Schendan, Ganis, & Kutas, 1998; Yamamoto & Kashikura, 1999; Liu, Harris, & Kanwisher, 2002) or even later in the 150-200-ms latency range (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Carmel & Bentin, 2002; Eimer, 2000; Jeffreys, 1996; Rossion et al., 2000; Taylor, Edmonds, McCarthy, & Allison, 2001).

However, the vast majority of experiments with faces used isolated, homogeneous, and well-centered stimuli. Such a bias in stimulus sets could explain early face selective brain activity that could be due either to a higher predictability of the expected stimuli that would speed up processing (Delorme, Rousselet, Macé, & Fabre-Thorpe, 2003) or to the bottom-up extraction of low-level physical properties from a set of homogenous stimuli (VanRullen & Thorpe, 2001b). Thus, the data obtained with isolated face stimuli may not necessarily apply to real-world situations. For instance, it is known from single-unit recordings in monkeys that the responses of neurons tuned to faces and other object categories are affected by the presence of other competing objects, and by the presence of a background (Chelazzi, Duncan, Miller, & Desimone 1998; Trappenberg, Rolls, & Stringer, 2002). Thus, it is interesting to investigate the functioning of the biological visual system in more realistic situations when faces are presented in the context of natural scenes. In order to obtain such a “realistic” estimate of face processing speed, we used a rapid go/no-go categorization task with briefly presented (20 ms) photographs of real-world scenes in which subjects had to react when the photograph contained a human face. Such a go/no-go design involves the simplest motor output possible, allowing subjects to respond as fast as they could with the minimal motor constraints. For comparison with another class of targets, subjects alternated between this face categorization task and an animal categorization task used in a series of earlier studies from our group.

The second issue we wanted to address concerned the characteristics of the object representations activated during rapid categorization tasks. These early representations could be specific to canonical presentations of the stimuli used in the tasks. Alternatively, they might rely on relatively view invariant representations. One way to address this issue is to analyze how processing is affected with inverted pictures. Indeed, face processing has been shown to be more sensitive to inversion than other object categories (Bentin et al., 1996; Rossion et al., 2000; Yin, 1969). This pattern of results has been taken as evidence that face perception relies on specific mechanisms dedicated to the processing of the configural information present in upright faces (Maurer, Le Grand, & Mondloch, 2002). To explain the additional time necessary to process inverted pictures, some models of object recognition postulate the existence of a normalization stage at which an object orientation must be aligned with a memory

template before matching can take place (see review in Tarr & Bülthoff, 1998; Ullman, 1996). Such normalization stage might be associated with a time consuming mental rotation of misaligned objects (Jolicoeur, 1988; Tarr & Pinker, 1989; Vannucci & Viggiano, 2000). Here we wanted to assess whether this inversion effect would affect the rapid categorization of human faces or animals presented in the context of natural scenes. To address this last issue, half of the pictures (faces, animals, and other natural scenes), whether targets or distractors, were presented upside-down.

Behavioral performance was analyzed in subjects alternating between rapid categorization of human faces and of animals presented randomly, upright or inverted, in the context of natural scenes. The processing speed and the magnitude of the inversion effect were compared for human faces and animals in two experiments, in which the main difference was in the presentation scale of the targets.

## Experiment 1

The first experiment was designed to compare directly the animal task used by our group in several previous experiments to a homologue human face task. In both tasks, target images were photographs of real-world scenes in which human faces or animals were shown at different scales, orientations, and positions (Figure 1). Because “face” stimuli did not contain isolated items, but faces in the context of human bodies embedded in natural scenes, we will refer in the remaining of the text to “human” pictures and “contextual face task.”

## Methods

### Participants

The 24 adult volunteers in this study (12 women and 12 men; mean age 31 years, ranging from 19 to 53 years; 5 left-handed) gave their informed written consent. All participants had normal or corrected-to-normal vision.

### Experimental procedure

Subjects were seated in a dimly lit room at 100 cm from a computer screen (resolution, 800 x 600; vertical refresh rate, 75 Hz) piloted from a PC computer. To start a block of trials, they had to place their finger on a response pad for 1 s. A trial was organized as follows: a fixation cross (0.1° of visual angle) appeared for 300-900 ms and was immediately followed by the stimulus presented during two frames (i.e., about 23 ms in the center of the screen). Participants had to lift their finger as quickly and as accurately as possible (go response) each time a target was presented and to withhold their response (no-go response) when the photographs did not contain a target. Responses were detected using infrared

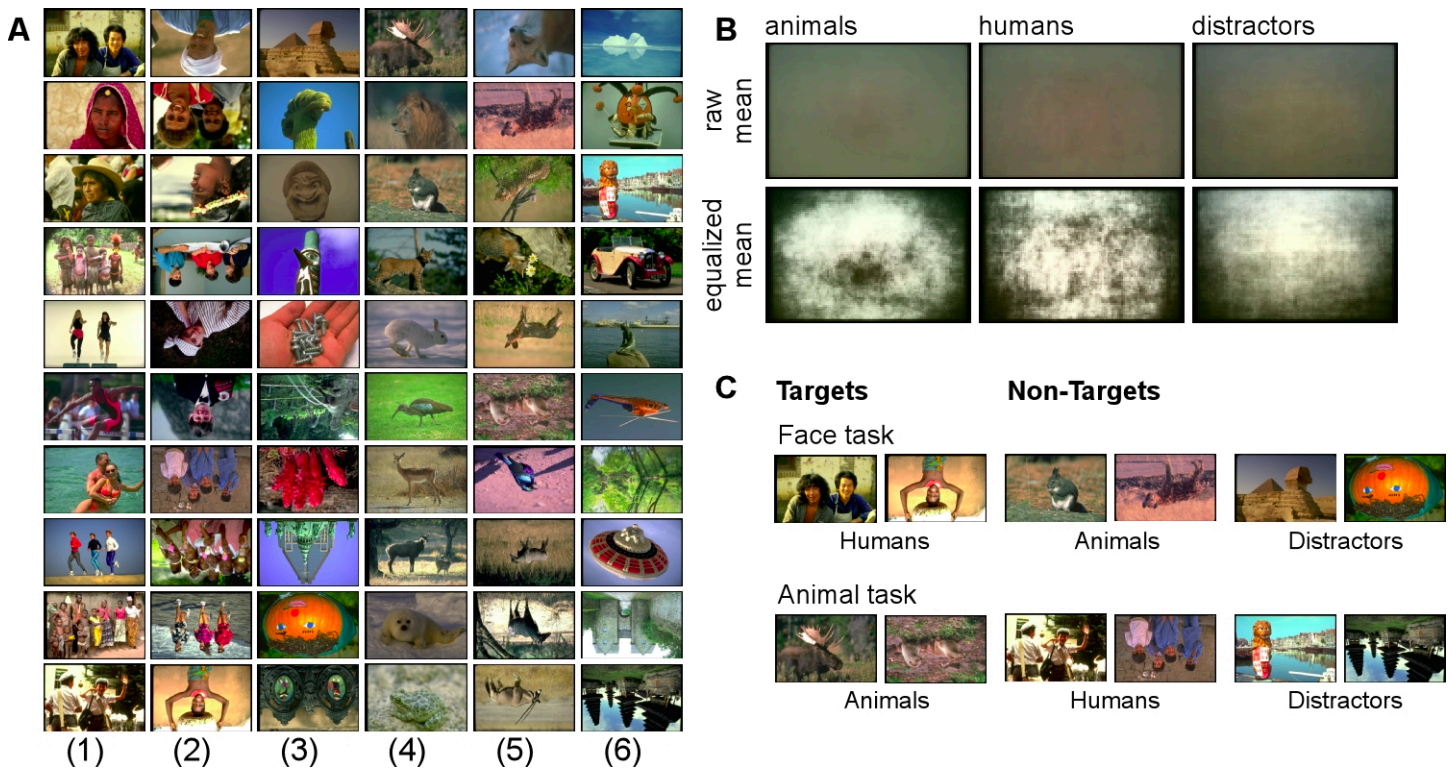


Figure 1. Tasks and stimuli. **A**. Examples of pictures used in Experiment 1. The 10 upright and inverted target pictures never missed by the subjects and associated with the fastest reaction time are presented for the face categorization task (columns 1 and 2, respectively) and for the animal categorization task (columns 4 and 5). Some examples of upright and inverted distractors that did not contain humans nor animals ("neutral" distractors) and on which subjects made no error are also illustrated in the upper and lower parts of column 3 for the face task and of column 6 for the animal task. **B**. Pixel-by-pixel average picture (raw mean) for each stimulus category (distractors refer to the neutral distractors) with equalized version computed using a commercial graphic software. The raw mean images were virtually uniform gray fields. The equalized images were obtained using the equalize function in a commercial graphic software. For each color channel and the luminance channel, the function attributes a "black" value to the darkest pixel and a "white" value to the brightest one. It then redistributes regularly the intermediate pixel values of the distribution between these two extremes. **C**. Tasks. While performing one of the two tasks, half of the non-targets were targets of the other task, and the other half were neutral distractors. Note the variety of stimuli used in this experiment.

diodes. Subjects were given 1000 ms to respond; longer reaction times were considered no-go responses. This maximum response time delay was followed by a 300-ms black screen, before the fixation point of the next trial was presented again for a variable duration, resulting in a random 1600-2200-ms intertrial interval.

An experimental session included 16 blocks of 96 trials. In 8 blocks, the target was an animal and in the remaining 8 blocks, the target was a human face. In each block, target and non-target trials were equally likely. Among the 48 non-targets, 24 contained targets of the other categorization task. Thus, when performing the face categorization task on a 96-trial block, 48 pictures contained at least one face, 24 non-target scenes contained animals, the last 24 non-targets "neutral distractors" being other types of natural scenes (see stimuli). Moreover, half of the targets and half of each of the non-target subsets were presented upright while the other half was presented inverted ( $180^\circ$  rotation). Each image was seen only once by a given subject, with one orientation (upright or inverted)

and one status (target or non-target), but the design was counterbalanced so that across all 24 subjects (1) each image ("neutral" distractor, animal or face image) was seen 12 times both in upright and inverted positions, and (2) each animal or face image was seen 16 times as a target and 8 times as a non-target. Half of the subjects started with the animal categorization task, the other half with the human face categorization task and conditions alternated by blocks of two. Subjects had two training blocks of 48 images before starting the test session. Training pictures were not repeated during testing.

Performance was evaluated by determining the percentage of correct trials and the latency at which subjects triggered their finger movement response, computed between stimulus onset and finger lift. An ANOVA was run on reaction times (RT) and rates of correct responses with category (animals vs. humans) and orientation (upright vs. inverted) as within-subject factors. A Greenhouse-Geisser correction for nonsphericity was applied.

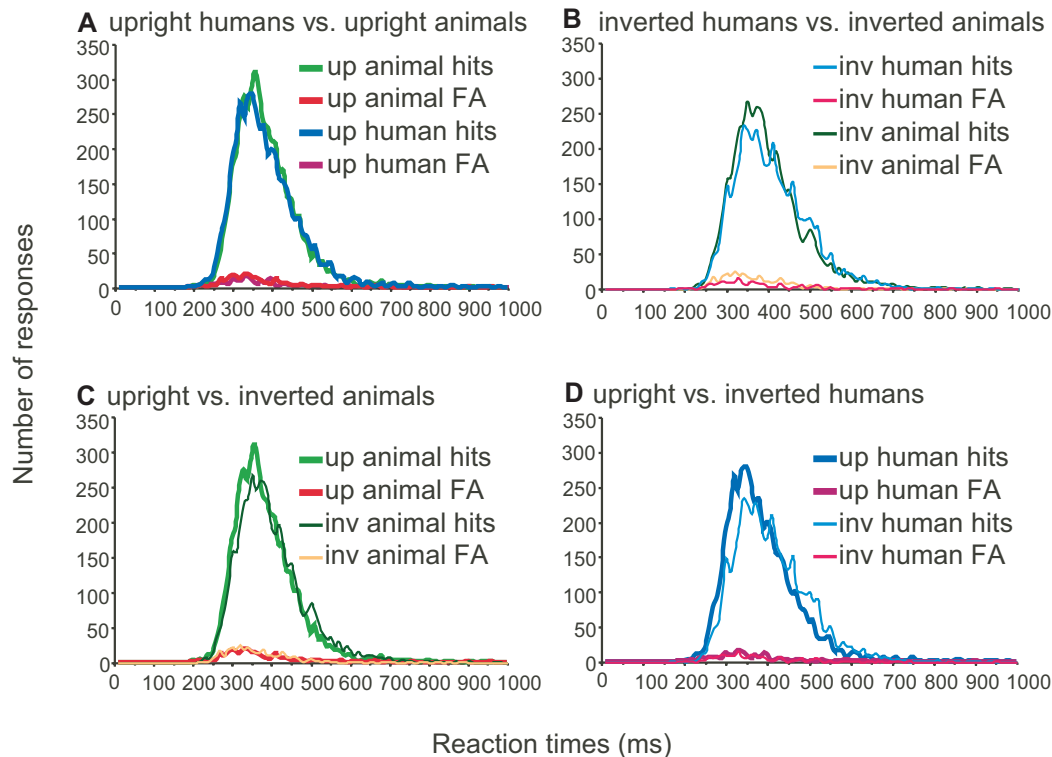


Figure 2. Reaction time (RT) distributions on correct and incorrect go-responses. RT distributions are presented with the number of responses expressed over time, with 10-ms time bins. Overall, no effect of the categorization task is seen on the early part of the RT distributions. Whether upright or inverted, responses to faces followed virtually the same time course as responses to animals (A and B). Inversion slightly disrupted the processing time course of both target-categories (C and D), an effect that was slightly more pronounced for faces.

### Stimuli

We used photographs of natural scenes taken from a large commercial CD-ROM library (Corel Stock Photo Library, see Figure 1). From this database, we selected 576 images that contained human faces, 576 images that contained animals, and 384 images that contained neither human faces nor animals. They were all horizontal photographs (768 by 512 pixels, sustaining a visual angle of about  $19.9^\circ \times 13.5^\circ$ ) and chosen to be as varied as possible. Animals included mammals, birds, fish, and reptiles. Human faces were presented in real-world situations with views ranging from whole bodies at different scales to face close-ups and including Caucasian and non-Caucasian people. There was also a wide range of non-target images that included outdoor and indoor scenes, natural landscapes (mountains, fields, forests, beaches, etc.), street scenes, pictures of food, fruits, vegetables, plants, buildings, tools, and other man-made objects, as well as some trickier distractors (e.g., dolls, sculptures, and statues, and a few non-target images containing humans for which the faces were not visible).

Subjects had no a priori information about the presence, the size, the position, or the number of targets in an image. Unique presentation of images prevented

learning, and brief presentations prevented exploratory eye movements.

### Results

In this section we will address three different aspects of processing: (1) processing of upright stimuli, comparing task performance for upright humans and upright animals; (2) processing of inverted stimuli, comparing inverted humans and inverted animals; and (3) effects of inversion on processing, comparing upright and inverted stimuli.

Overall, subjects were very accurate on both tasks, scoring 95.6% in the human task and 95.5% in the animal task (n.s.d.) and very fast (mean RT of 393 ms vs. 388 ms, respectively, n.s.d.). ANOVA tests performed on the overall results revealed that subjects categorized human targets with a lower accuracy than animal targets (95.7% vs. 98.3%, respectively;  $F = 16$ ,  $p = .001$ ), whereas they correctly ignored a higher proportion of distractors in the contextual face task than in the animal task (95.3% vs. 92.8%, respectively;  $F = 20.8$ ,  $p < .0001$ ). There was no main effect of category on mean and median RT. However, both measures presented a significant interaction between the category and orientation factors (both:  $F = 18.0$ ,  $p < .0001$ ). These main effects are

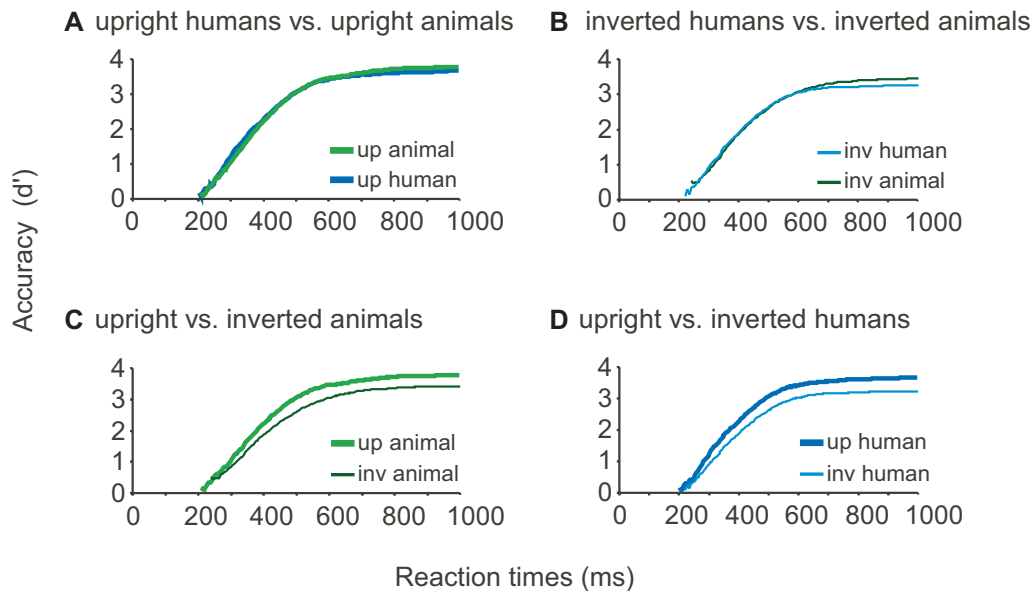


Figure 3. Time course of performance. Average performance accuracy (in  $d'$  units) is plotted as a function of processing time with 10-ms time bins. Cumulative numbers of responses were used. The  $d'$  was calculated from the formula  $d' = z_n - z_s$ , where  $z_n$  is chosen such that the area of the normal distribution above that value is equal to the false-alarm rate, and where  $z_s$  is chosen to match the hit rate. Note that the  $d'$  calculated here is not presumed to represent the actual distributions of signal and noise that underlie performance in the response time task. By taking into account the hit and false alarm rates in a single value at each time point, this time course of performance gives an estimation of the processing dynamics for the entire subject population. The plateau values correspond to the  $d'$  calculated from the overall accuracy results. Confirming results from Figure 2, performance time course functions were virtually identical for contextual human face and animal categories, independent of the orientation (i.e., upright or inverted). The inversion effect was very similar in both cases with a slightly earlier onset for human pictures.

explored in details in the two next sections using post hoc ANOVA, paired  $t$  tests, and Wilcoxon tests.

### **Contextual faces versus animals: upright stimuli**

Here only the trials (over 9,200) performed in each task with upright scenes are considered. Mean accuracy was virtually identical in the two tasks (96.4% and 96.3% for faces and animals) (Figure 2A and Figure 3A).

Accuracy, however, was biased differently in each of them. Subjects categorized upright human targets with a lower accuracy than upright animal targets (humans = 97.5%, animals = 98.7%, Wilcoxon test,  $z = -2.3$ ,  $p = .02$ ), whereas no significant effect was present at the level of upright distractors (humans = 95.3%, animals = 93.9%, n.s.d.).

Regarding processing speed, upright contextual faces were not categorized faster than upright animals. First, this was shown by the RT distributions of correct go-responses in both tasks (Figure 2A). Second, there was no task effect on either mean (382 ms in both conditions) or median RT (368 ms for faces and 371 for animals) (Figure 2A and Figure 3A). Thus, on average, animals and faces were processed at the same speed according to mean and median RT. Given the problems associated with using only mean RT values to evaluate processing speed (Perrett, Oram, & Ashbridge, 1998;

McElree & Carrasco, 1999), we used two more appropriate values: the time course of performance (Figure 3) and the minimal RT. The analysis of these two factors confirmed that contextual faces and animals were categorized at the same speed within natural images. Comparing the time course performances of each task (Figure 3A) clearly shows that early responses were produced at similar latencies regardless of the task and that performances follow time courses that are virtually undistinguishable. The minimal behavioral processing time was evaluated by determining the latency at which correct go-responses started to significantly outnumber incorrect go-responses ( $\chi^2$ ,  $p < .001$ ) using a noncumulated RT histogram with 10-ms time bins (Figure 2). These early responses cannot be considered as anticipations because if behavior was random on target and distractor trials (which are equally likely), hits and false alarms should have the same probability. The latency at which go-responses are statistically biased toward hits gives an indication of the minimal processing time required to trigger a motor response in the task while eliminating any bias due to anticipations. The analyses were performed either on the overall data (set by pulling together all trials from all subjects) or for each subject separately. No significant differences between the contextual face and the animal categorization tasks were found. The minimal processing

Table 1. Average Results From Experiment 1

	Contextual human face task		Animal Task	
	Upright scenes	Inverted scenes	Upright scenes	Inverted scenes
<b>Accuracy (%)</b>				
Mean	96.4 (1.7) [92.2-99.2]	94.7 (2.3) [88.3-98.2]	96.3 (2.0) [91.1-99.2]	94.8 (2.3) [89.8-98.4]
Correct go	97.5 (2.6) [90.1-100]	93.9 (4.9) [78.7-99.5]	98.7 (1.3) [95.3-100]	97.9 (1.4) [95.3-100]
Correct nogo (tD)	94.5 (5.9)	94.9 (5.0)	94.7 (4.1)	92.8 (4.2)
Correct nogo (nD)	96.1 (2.3)	95.8 (2.0)	93.1 (4.2)	90.5 (4.6)
<b>RT (ms)</b>				
Mean	382 (43) [317-468]	405 (49) [338-500]	382 (41) [312-465]	395 (43) [324-486]
Median	368 (43) [309-457]	391 (50) [317-484]	371 (42) [305-460]	380 (44) [298-470]
<b>Minimal RT (ms)</b>				
Overall data	260	260	260	260
Individual data	329 (43) [250-370]	353 (50) [270-430]	333 (35) [260-380]	348 (41) [270-460]

(tD) and (nD) refers respectively to the distractors that were used as targets in the other task or to the neutral distractors used in both tasks. SD is indicated in brackets. Range of individual responses (min and max) is indicated in square brackets.

time was 260 ms with the overall data set (for both faces and animals) and 329 ms (contextual faces) versus 333 ms (animals) for individual data. These results do not support any processing speed advantage for human faces.

### **Contextual faces versus animals: inverted stimuli**

The comparison of performance did not show any difference between the processing of contextual human faces and animals when presented in an upright orientation. In our protocol, half of the stimuli were also presented upside down and the present section compares the processing of inverted contextual faces and inverted animals to investigate whether the similarity found with upright stimuli extends to inverted ones. As in the preceding section, the comparison is carried out on over 9,200 trials for each condition.

Mean accuracy was virtually identical for inverted faces (94.7%) and inverted animals (94.8%) (Figure 2B and Figure 3B). Accuracy showed the same biases than with upright stimuli, with a higher accuracy (97.9% vs. 93.9%; Wilcoxon test,  $z = -4.1$ ,  $p < .0001$ ) on inverted animal targets than on inverted contextual faces. Moreover, the higher accuracy on inverted distractors observed in the contextual face task (95.4%) when compared to the animal task (91.7%) was highly significant (Wilcoxon test,  $z = -3.9$ ,  $p < .0001$ ).

Figure 4 illustrates the higher number of errors performed on inverted distractors in the animal task both when compared to the set of upright stimuli in the animal task and when compared to the set of inverted distractors processed in the contextual face task. The figure also illustrates that, regardless of their orientation, neutral distractors induce a higher number of false alarms in the animal categorization task. Again this is true when compared to the other set of distractors in the animal task, or when compared to the performance on neutral distractors in the contextual face task.

When considering the average categorization speed, inverted faces were categorized about 10 ms slower than inverted animals. This was true (both paired  $t$  test  $p < .006$ ) for both mean RT (405 ms and 395 ms, respectively, for contextual faces and animals) and median RT (391 ms and 380 ms, respectively) (Figure 2B and Figure 3B). However, this processing speed difference failed to reach statistical significance for the minimal processing time (as defined in the preceding section). Minimal RT was 260 ms, regardless of the kind of targets to categorize, when calculated on the overall data set. Mean minimal RT calculated on all individual subject data was 348 ms for animals and 353 ms for faces. The RT distributions and the performance time course functions for each task also show a good overlap of early responses regardless of the task. Differences are observed later (around mean RT or for late responses).

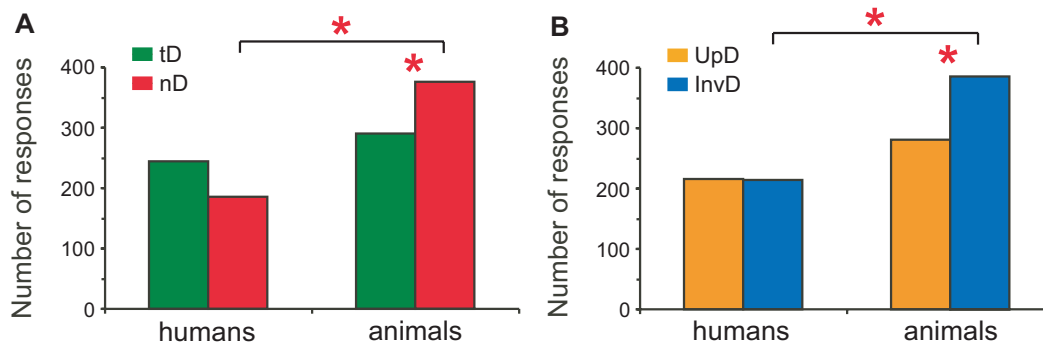


Figure 4. Analysis of incorrect go-responses made toward distractors in the “contextual human face” task and in the “animal” task. The data indicate a different processing of the distractors depending on the task performed by the subject. Statistically significant differences between two conditions are illustrated with an asterisk. A. Comparison of incorrect go-responses triggered by neutral distractors (nD in red) and by distractors that were targets in the other categorization task (tD in green). Independent of picture orientation, the responses on distractors showed a significant bias (interaction between task and type of distractor factors,  $F = .0$ ,  $p = .002$ ). More errors were made on neutral distractors in the animal task compared to the human face task ( $F = 36.9$ ,  $p = .0001$ ). Within the animal task, neutral distractors induced more errors than human faces (tD) ( $F = 6.8$ ,  $p = .016$ ). B. Comparison of incorrect go-responses triggered by upright (UpD in orange) and inverted (InvD in blue) distractors. An interaction between task and orientation factors ( $F = 7.0$ ,  $p = .014$ ) showed that more errors were made on inverted distractors in the animal task ( $F = 18.7$ ,  $p = .0001$ ), whereas no difference was seen in the contextual human face task (n.s.d.). Inverted distractors were also better categorized in the human face task than in the animal task ( $F = 37.5$ ,  $p = .0001$ ).

### Contextual faces versus animals: the inversion effect

In this section, we focus more specifically on the presence and the strength of the inversion effect as a function of the target category.

Inversion produced a very weak decrease of global accuracy (<2%) that was very similar for both animals and human faces (orientation effect:  $F = 37.1$ ,  $p < .0001$ ; no interaction between task and orientation factors) (Figure 2C and 2D and Figure 3C and 3D). The percentage of correct go-responses decreased significantly with inversion for both animals (98.7% vs. 97.9%, Wilcoxon test,  $z = -2.7$ ,  $p = .006$ ) and contextual faces (97.5% vs. 93.9%,  $z = -4.1$ ,  $p < .0001$ ). Statistically, this was shown by a main orientation effect ( $F = 27.6$ ,  $p < .0001$ ) that was stronger for faces (interaction between orientation and task factors:  $F = 19.7$ ,  $p < .0001$ ). In parallel with the slight decrease of global accuracy, inverted pictures were also categorized on average with significantly longer RT (Figure 2C and 2D and Figure 3C and 3D) than upright pictures (mean RT:  $F = 140.7$ ,  $p < .0001$ ; median RT:  $F = 72.9$ ,  $p < .0001$ ). This held true for both categories but with an inversion effect on speed that was reliably more pronounced for faces (+23 ms on both mean and median RT, both paired  $t$  test:  $p < .0001$ ) than for animals (+13 ms on mean RT,  $p < .0001$ ; +9 ms on median RT,  $p = .001$ ). Although the global reaction time increase appears robust with both kinds of inverted targets at the level of mean and median RT, it is far from being as obvious when considering the minimal processing time. When determined on the overall data, no effect was seen

regardless of the categorization task. At the individual level, however, there was a small inversion effect for both categories with a nonsignificant tendency to be more pronounced for faces (+24 ms,  $p < .0001$ ) than for animals (+15 ms,  $p = .004$ ). The time course of performance showed that the stimulus inversion did not simply shift the curve toward longer latencies but rather decreased the slope of the functions that originate at similar early latencies.

## Discussion

Overall, subjects were able to respond both very accurately and rapidly in the two tasks. This level of performance is impressive given the extreme variability of the photographs used in this experiment. It can be taken as the hallmark of the sophistication of the fast mechanisms implemented in the ventral pathway of the human brain (Riesenhuber & Poggio, 2000; Thorpe & Imbert, 1989; VanRullen et al., 1998; Wallis & Rolls, 1997). If this conclusion had already been reached from results of earlier studies, here we extend these findings by showing that (1) the fast coarse categorization of objects in natural scenes is very weakly affected by inversion; (2) contextual human faces cannot be processed faster or more efficiently than another relevant visual category such as animals; and (3) the inversion effect, although very weak in both tasks, is slightly more pronounced for faces.

The fact that animals are processed with the same speed and accuracy as contextual human faces when both types of targets are presented at different scales, in varied

number and position in the image, argues against a hardwired face mechanism that would be more efficient than other non-face object mechanisms (Tarr & Cheng, 2003). Because it has been shown previously that animals could not be processed faster than another relevant, nonbiological category, such as means of transport (VanRullen & Thorpe, 2001a, 2001b), contextual faces cannot be said to benefit from specific temporal advantages, at least in our task. We do not want to argue that this kind of rapid categorization process would apply to any object category; instead, it might depend on a certain level of expertise (that needs to be determined) beyond which the categorization of any behaviorally relevant object could rely on such fast processes.

Although we found evidence that inversion of natural scenes did produce reliable effects on performance, with responses delayed (13 ms vs. 23 ms for animal and faces) and accuracy impaired for inverted pictures (1% vs. 3.5% for animal and faces), it is important to note that these effects were both very weak (although slightly more pronounced for faces). With such temporal constraints, very little time would be available to implement a mental rotation mechanism during the time course of the categorization process. On the other hand, the speed of recognition of an object might depend on the rate of accumulation of activity from object selective neurons (Perrett et al., 1998; Ashbridge, Perrett, Oram, & Jellema, 2000). Neurons in higher-level occipito-temporal visual areas respond to complex stimuli such as animals and faces. At the level of neuronal populations, the strength of the population response is correlated to the number of activated neurons. Now, we can hold the very plausible assumption that the population response must reach a given constant threshold activation level (Hanes & Schall, 1996) in order for a behavioral response to be triggered. Through experience, more neurons, each one more selectively tuned, respond to animals, human faces, and body parts in the upright position compared to inverted positions. Groups of neurons responding to upright and inverted objects would start to respond at about the same latency but responses would accumulate more slowly in the case of inverted stimuli, leading to an increase in response latency. This hypothesis is supported by the time course of performance (Figure 3) that originated at similar latencies but increased with different slopes depending on whether the stimuli were presented upright or upside down. It follows that, on average, it takes slightly more time to reach the threshold for inverted stimuli, and therefore to categorize them.

If the processing of upright faces and animals followed the same behavioral temporal course, what is special in faces that led to differences in the processing of inverted stimuli? The inversion effect is usually taken as evidence that face processing relies preferentially on configural mechanisms distinct from part-based mechanisms thought to be more important in the

processing of other objects (e.g., see review in Itier & Taylor, 2002; Rossion & Gauthier, 2002). When faces appear in their typical upright orientation, configural information is extracted. This extraction is disrupted by inversion, except for objects whose discrimination relies on characteristic features that are not affected by inversion. However, following Perrett's hypothesis, the fact that faces were more sensitive to inversion than animals can be explained by a face population selectivity more strictly linked to the canonical upright view through experience (see support for such a view in Rossion & Gauthier, 2002; Tarr & Gauthier, 2000). Accordingly, neurons would fire less efficiently in response to inverted than upright faces, leading to a smaller accumulation of activity for inverted faces compared to inverted animals (because the latter might be represented by a cell population less strictly tuned to the upright orientation). As a consequence, the stronger inversion effect for faces often explained by the specificity of face processing (Farah et al., 1995, 1998) can be alternatively explained by the rate of accumulation of selective neural activity.

However, it remains possible that different strategies or brain mechanisms were used in the two tasks. Inversion had different effects on each category: when looking for animals, subjects made a high number of incorrect responses on inverted distractors, whereas when looking for human faces, they tended to miss more inverted targets. This could be the consequence of a greater similarity between animals and distractors than between faces and distractors, and the use of more specific representations to perform the face task than the animal task. This hypothesis is supported by the fact that more errors on neutral distractors and on inverted distractors were performed during the animal task than during the face task.

Finally, animals were slightly more easily detected in natural scenes than faces, which might indicate that the two sets of images were not equated in difficulty and might potentially have masked a processing speed advantage in favor of faces. Furthermore, this discrepancy might also potentially explain the very weak inversion effect found for faces. To test these alternative explanations and further characterize the processing of faces in natural scenes, we designed a second experiment.

## Experiment 2

Experiment 2 was designed to compare the rapid categorization of faces and animals with more homogenous sets of images. In Experiment 2, subjects were only presented with close-up views of human and animal heads and were required to categorize human faces and animal faces. Human and animal faces were chosen to be as varied as possible but always in the context of natural scenes; furthermore, neutral distractor pictures (that did not contain animal or human faces)



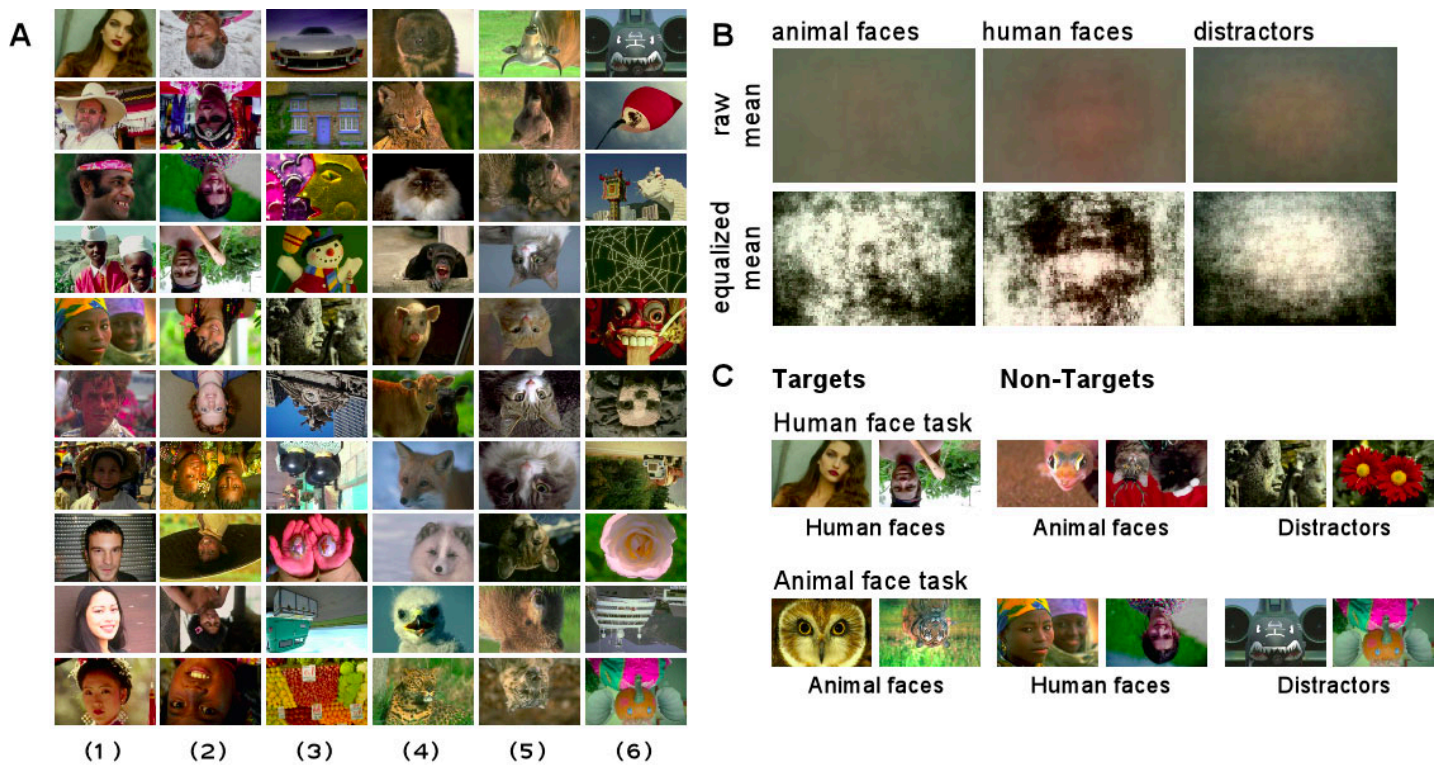


Figure 5. Picture examples and experimental design. Nomenclature as in Figure 1.

were chosen to include “tricks,” such as dolls, statues, flowers, and other headlike “blobs.”

## Methods

Except where otherwise mentioned, methods were identical to those used in Experiment 1.

### Participants

The 24 human participants (12 women and 12 men, mean age 30 years, ranging from 19 to 51 years, 3 left handed) who volunteered in this study gave their informed written consent. Nine of them had participated in the first experiment. All participants had normal or corrected-to-normal vision.

### Experimental procedure

An experimental session included 8 blocks of 96 trials. Subjects performed two categorization tasks: in 4 blocks the target was an animal face and in the 4 other blocks the target was a human face. In each block, target and non-target trials were equally likely. Among the 48 non-targets, 24 were targets in the other categorization task. Thus, when performing a human face categorization task on a 96 trial block, 48 pictures contained at least one human face, 24 non-target scenes contained animal faces, the last 24 non-targets being neutral distractors (i.e., other types of natural scenes and “trick” stimuli) (see [Stimuli](#) and [Figure 5](#)). Half of the targets and half of each non-target subset were presented upright while the other half

was presented inverted. The design was counterbalanced so that in the overall group of subject, each image was seen in upright and inverted positions and processed as a target and as a non-target. Half of the subjects started with the animal face categorization, the other half with the human face categorization. Subjects had one training block before starting each of the two test sessions. Training pictures were not used during testing.

### Stimuli

A total of 768 photographs were selected from the Corel Stock Photo Library; 288 contained human faces, 288 additional images contained animal faces, and the last 192 photographs contained neither human nor animal faces ([Figure 5](#)). They were all horizontal photographs (768 by 512 pixels, sustaining about 19.9° by 13.5° of visual angle) and chosen to be as varied as possible. Faces were always highly visible with views ranging from close-up to views showing the most upper part of the body. Animals included mammals, birds, fish, and reptiles. They did not include arthropods and were chosen so that a face configuration could always be seen (eyes, mouth, and nose). Human faces were presented in real-world situations and included humans from all over the world. There was also a very wide range of non-target images that included outdoor and indoor scenes, natural landscapes, street scenes, pictures of food, fruits, vegetables, plants, flowers, buildings, tools, and other man-made objects, as well as many “tricky” distractors, such as dolls, sculptures, and statues. A particular attempt

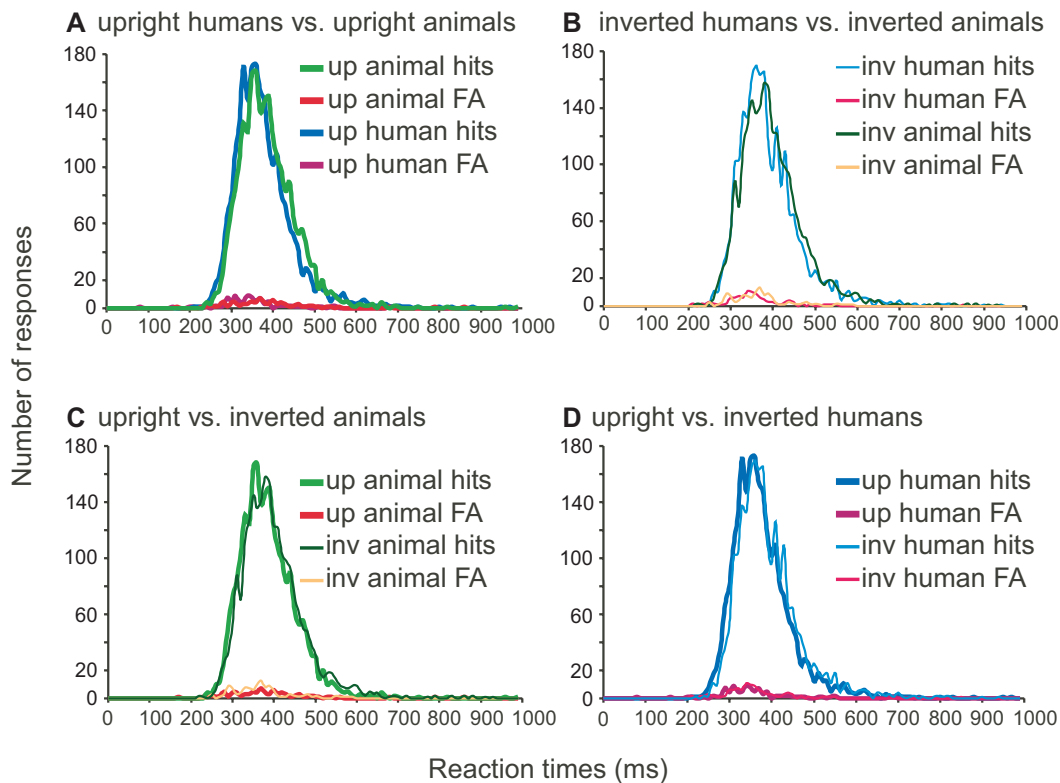


Figure 6. Reaction times (RT) distributions on correct and incorrect go-responses. (See caption Figure 2.) Overall, no effect on processing speed is seen on the early part of the RT distributions except in D, where the hits on upright human faces start to diverge early from the hits on inverted faces. Whether upright or inverted, responses to human faces followed virtually the same time course as responses to animal faces (A and B). Inversion slightly disrupted the processing time course of both target-categories (C and D), an effect that was slightly more pronounced for faces.

was made for most distractors to have one or more headlike “blobs” positioned centrally or laterally in the picture, as were human and animal faces.

Subjects had no a priori information about the presence, the size, the position, or number of targets in an image, and to prevent learning, each image was seen only once in one orientation (upright or inverted), either as a target or as a non-target, by each subject.

## Results

In Experiment 2, despite the greater target/distractor similarity compared to Experiment 1, the use of close-up views led to excellent performances both in terms of accuracy and speed. ANOVA tests performed on the overall results showed no category effect on global accuracy (97.4% for both human and animal faces), target accuracy (99.3% for both) or distractor accuracy (95.5% for both). However, median RT were shorter in response to human faces (377 ms) than to animal faces (387 ms) ( $F = 4.6, p = .043$ ), a main effect that was not significant for mean RT (humans: 389 ms; animals: 397 ms). The next two sections will present a detailed analysis of these global results using post hoc ANOVA, paired  $t$  tests, and Wilcoxon tests. The first section will compare the

processing of upright human faces to the processing of upright animal faces. The second section will concentrate on inverted stimuli. The third section will present specifically the differences between upright and inverted stimuli on the processing of human faces and animal faces.

### *Human faces versus animal faces: upright stimuli*

Mean accuracy was virtually identical for both kinds of upright pictures with 97.7% in the human face task versus 97.9% in the animal face task (Figure 6A and Figure 7A). Targets were better categorized than non-targets (99.5% vs. 96%, respectively,  $F = 37.7, p < .0001$ ), with similar proportions of go-responses for upright humans (99.6%) and upright animals (99.5%). Contrary to Experiment 1, subjects tended, on average, to respond about 10-ms faster for human than for animal faces (Figure 6A and Figure 7A). This slight advantage reached significance for median RT (371 ms vs. 384 ms, paired  $t$  test:  $p = .031$ ) but not for mean RT (382 ms vs. 392 ms, n.s.d.). This effect is relatively clear on the RT distribution for intermediate and long latency responses. On the other hand, although it is barely visible on the initial part of the RT distribution of Figure 6A or at the

onset of the performance time course functions of Figure 7A, the 10-ms global advantage in favor of human compared to animal pictures was also observed with the minimal processing time computed on cumulated population data (260 ms vs. 270 ms, respectively). The same tendency in favor of human pictures was seen for individual minimal processing time in both tasks, but it did not reach significance (327 ms vs. 338 ms, n.s.d.).

### **Human faces versus animal faces: inverted stimuli**

No statistical difference could be seen between the accuracy scores computed for each task. Indeed, subjects again reached very similar performances (Figure 6B and Figure 7B) scoring 97.2% with inverted human faces and 96.9% with animal faces. Correct go responses were triggered in similar proportion in both tasks (99.0% vs. 99.2%).

The overall mean RT showed a 6-ms lag between human face (396 ms) and animal face processing (402 ms) that did not reach significance. This lag reached 8 ms when calculated on the overall median RT between human faces (median RT: 382 ms) and animal faces (median RT: 391 ms), an effect that did not reach significance either.

When it was calculated on the overall population data, the earliest responses were found earlier for animal faces (270 ms) than for human faces (280 ms). A pattern that was not consistent when individual data were considered as mean individual data showed a nonsignificant advantage for inverted animal faces (345 ms) versus human faces (335 ms) (Figure 6B and Figure 7B).

As in the first experiment, the incorrect go responses produced on distractors were analyzed (Figure 8) and outlined different biases depending on the task performed by the subject. As in Experiment 1, subjects made fewer errors on neutral distractors in the human face task than in the animal face task, regardless of their orientation (Figure 8A), but a bias was found within the human face task for the two different subsets of distractors: subjects made more errors on pictures that contained animals than on neutral distractors. Finally, Figure 8B shows the same bias as that already seen in Experiment 1, with more errors on inverted stimuli in the animal task.

### **Human faces versus animal faces: the inversion effect**

As in Experiment 1, inversion had a reliable but weak effect on performance. Inversion decreased global

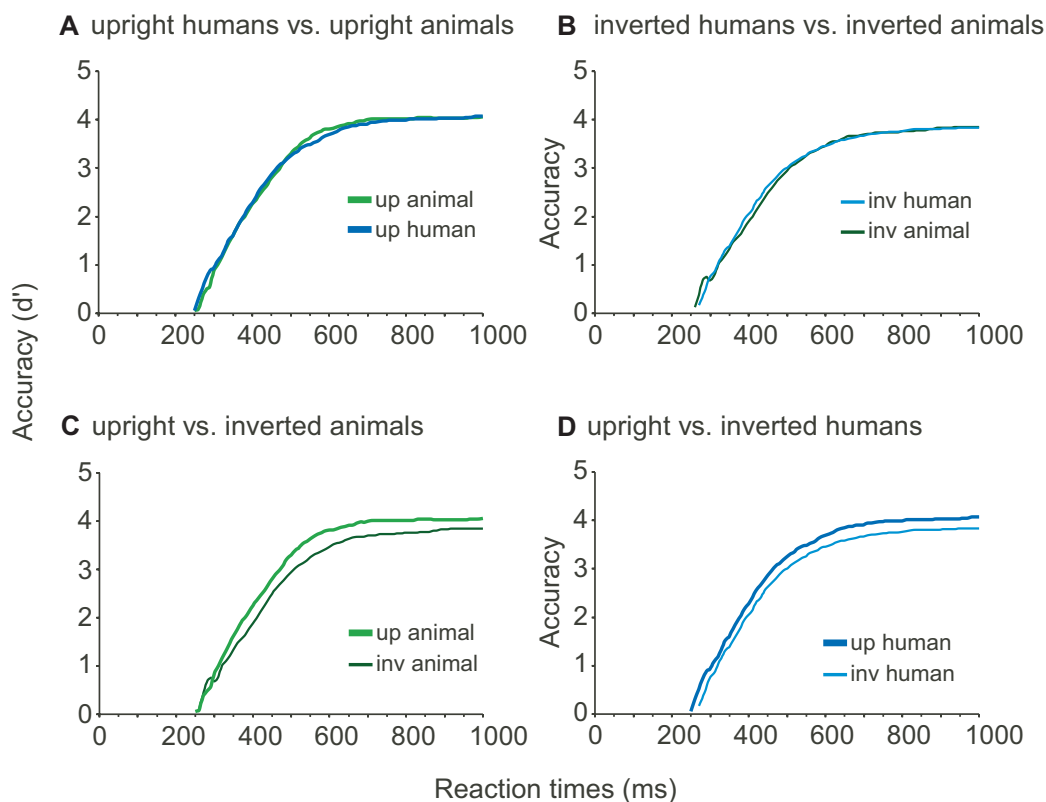


Figure 7. Performance time course. (See caption Figure 3.) A and B show that human and animal faces follow the same type of processing course. C and D show the slight decrease of accuracy in both tasks and the temporal cost associated with inverted stimuli. The temporal cost is seen from the very beginning with human faces whereas the  $d'$  curves for upright and inverted animal faces, initially superimposed, diverge later on.

Table 2. Summary of Results From Experiment 2

	Human face task		Animal face task	
	Upright stimuli	Inverted stimuli	Upright stimuli	Inverted stimuli
<b>Accuracy (%)</b>				
Mean	97.7 (1.8) [92.1-100]	97.2 (1.7) [91.0-99.5]	97.9 (1.3) [95.7-100]	96.9 (1.6) [93.2-99.5]
Correct go	99.6 (1.3) [93.6-100]	99.0 (1.2) [95.8-100]	99.5 (0.9) [95.8-100]	99.2 (0.8) [97.9-100]
Correct nogo (tD)	94.6 (6.1)	93.9 (6.4)	96.6 (3.7)	94.5 (4.1)
Correct nogo (nD)	97.0 (3.1)	96.8 (2.5)	95.8 (3.0)	94.9 (4.2)
<b>RT (ms)</b>				
Mean	382 (33) [338-445]	396 (28) [352-444]	392 (35) [328-479]	402 (36) [337-493]
Median	371 (31) [330-428]	382 (26) [338-431]	384 (37) [312-464]	391 (34) [328-468]
<b>Minimal RT (ms)</b>				
Overall data	260	280	270	270
Individual data	327 (27) [290-380]	335 (22) [290-400]	338 (26) [290-410]	345 (31) [270-420]

(tD) and (nD) refers respectively to the distractors that were used as targets in the other task or to the neutral distractors used in both tasks. SD is indicated in brackets. Range of individual responses (min and max) is indicated in square brackets.

accuracy in both tasks (-0.5% in the human face task, -1% in the animal face task,  $F(1,23) = 8.3$ ,  $p = .008$ ) (see [Figure 6C](#) and [6D](#) and [Figure 7C](#) and [7D](#)). This effect was only significantly reliable for animal faces (Wilcoxon test,  $z = -2.5$ ,  $p = .013$ ; human faces: n.s.d.). When considering accuracy on targets and distractors separately, the inversion effect, albeit very small, reached significance only for go-responses on human faces ( $z = -2.1$ ,  $p = .039$ ) and for no-go responses on animal faces ( $z = -2.0$ ,  $p = .042$ ).

Inversion also slightly delayed RT (mean: +14 ms and +10 ms,  $F(1,23) = 58.3$ ,  $p < .0001$ ; median: +11 ms and +7 ms,  $F(1,23) = 34.7$ ,  $p < .0001$ , for human and animal faces, respectively), an effect that was not significantly stronger for human than for animal pictures, as shown by an absence of interaction between task and orientation factors. However, the result concerning minimal RT calculated from the overall population data showed a difference between early processing of human and animal faces. There was no effect of orientation for animal faces (270 ms for upright and inverted stimuli), but the minimal RT was 20 ms shorter with upright faces (260 ms) than inverted faces (280 ms). This small differential effect between the two tasks can be seen in [Figure 7](#) by comparing the initial part of the  $d'$  curves in [Figure 7C](#) and [7D](#). The performance curve with inverted human faces is shifted toward longer latency with the same slope than for upright faces, whereas with animal faces, the earliest responses appear at the same latency, and only the slope of the performance curve is affected when inverting the animal faces. However, this result on the overall data set was not confirmed by the analysis of individual minimal reaction time showing the same inversion effect for human faces (+9 ms) and animal faces (+7 ms) ( $F = 16.5$ ,  $p < .0001$ , no interaction with the category factor).

## Discussion

Experiment 2 tried to provide a more direct comparison of human face versus animal face processing in natural scenes by using more homogenous sets of images. Levels of difficulty in the two tasks were similar regarding target detection accuracy. Despite high feature similarities between targets, and despite our considerable effort to use confusing distractors sharing global features with close-ups of faces, subjects performed remarkably well in these two tasks, in which processing efficiency was virtually identical. The high accuracy level reached in this experiment might be explained by the fact that humans (and faces in particular) constitute a very special object class, automatically categorized and segregated by our visual system, hence producing no interference with other object categories. Indeed, as in Experiment 1, we found evidence that neutral distractors were associated with more errors in the animal task than in the human task, which might imply that there was a higher similarity between neutral distractors and animals than between neutral distractors and humans. However, does this mean that human faces would benefit from computational advantages that would make them easier or faster to detect? We found no clear evidence in favor of this hypothesis. In the present experiment, contrary to the first one, there was a tendency for human faces to be processed on average about 10-ms faster than animal faces, an advantage that was present for both upright and inverted orientations, but appeared only for upright stimuli when considering the earliest behavioral responses. Such a small but reliable effect might be explained at the neuronal population level by a larger number of neurons coding for human faces than for different animal faces, thus slightly reducing the time to threshold decision as previously postulated in the

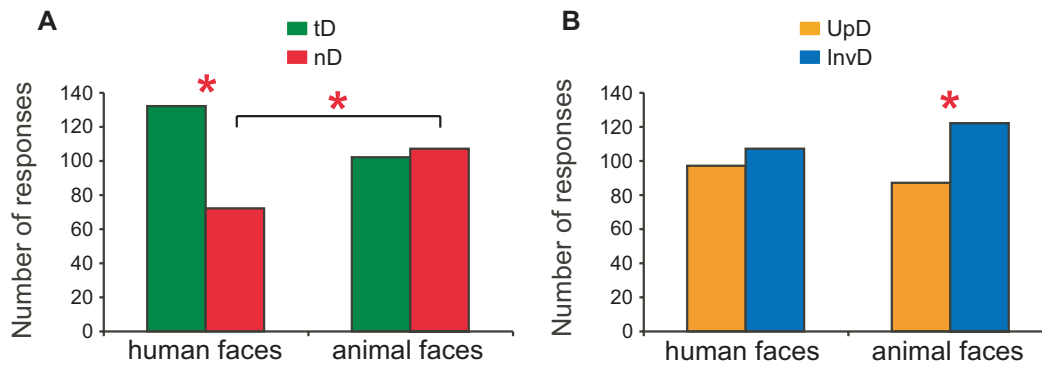


Figure 8. Analysis of incorrect go-responses made on distractors in the human and in the animal face tasks. (See Figure 4 caption for details.) A. Independently of picture orientation, the responses on distractors showed a significant bias (interaction between the task and type of distractor factors,  $F = 4.8$ ,  $p = .04$ ). Neutral distractors were slightly better categorized in the face task than in the animal task (96.9% vs. 95.3%, respectively,  $F = 7.5$ ,  $p = .012$ ). Within the human face task, animal faces (tD) induced more errors than neutral distractors ( $F = 4.5$ ,  $p = .045$ ). B. Furthermore, the orientation of the distractors induced a bias only in the animal task in which more errors were induced by inverted than by upright distractors ( $F = 7.0$ ,  $p = .014$ ).

discussion of Experiment 1. Indeed, a 10-ms difference in processing speed does not fit with the involvement of a different mechanism for the processing of human faces compared to animal faces, but rather point to a quantitative difference in the processing of the two categories rather than to a qualitative difference. The time course of performance in the two tasks strengthens this interpretation. Thus, these results are best explained in a framework in which the ventral pathway is conceived as implementing a unitary mechanism processing all object categories (Tarr & Cheng, 2003). Under such a framework, the speed to categorical decision threshold would depend on the number of neurons tuned to a specific category. According to this working hypothesis, time to threshold would be not surprisingly shorter for an extensively represented category such as human faces compared to another object category such as animal faces. Following this idea, a delay as short as 10 ms can find an explanation at the level of a neuronal population more sensitive to a category than the other, rather than in the involvement of a totally different mechanism. This delay was rather small probably due to the task used in the two experiments used here. Indeed, a superordinate categorization task might rely on coarsely defined diagnostic features (Schyns, 1998; Ullman, Vidal-Naquet, & Sali, 2002). A processing time course similar to the one found here for animals and humans has also been reported for another category like means of transport (VanRullen & Thorpe, 2001a), which suggests that the same level of complexity might be reached in a large range of natural scene categorization tasks. The use of more demanding categorization tasks relying on more specific features might reveal more dramatically an existing bias at the neuronal population level between two categories. If subjects had been asked to realize a gender discrimination task with human and animal faces, the difference between the two categories would certainly have been much larger.

However, even in this condition, the same simple mechanism of accumulation of evidence working at the level of a large neuronal population might be sufficient to explain the results. This kind of experiment will be important in the future to distinguish between different models of organization of the ventral pathway.

A complementary interpretation on the small difference in processing speed between human and animal faces lies in the smaller range of variability between different human faces compared to the large differences between faces of vertebrate animals (birds, monkeys, antelopes, reptiles, etc.). This seemed to be partly the case, given that more structure appeared in the “mean image” for humans than for animals (Figure 5B). It might be that reducing the number of different animal species would have allowed a more specific pre-setting of the neuronal population responding to animals, thus eliminating any differences at all between animal and human faces.

As in Experiment 1, another weak but consistent effect was seen with inversion in both tasks. Whereas the accuracy impairment appeared to be of similar magnitude for animal and human faces, the earliest response to inverted human faces could appear with a 20-ms delay when compared with upright human faces. This might be the hallmark of face configural processing, more disrupted by inversion than other object processing routines (Yin, 1969). However, as already developed in the discussion of the first experiment, a more simple explanation, emphasizing experience-induced bias at the neuronal population level, could constitute a viable alternative. According to this model, there is no need to call for the involvement of a mental rotation mechanism or a mechanism specifically dedicated to the processing of upright human faces. One might argue that models of object recognition relying on a time consuming normalization stage between sensory inputs and memory

templates might explain the inversion effects in our two experiments (Tarr & Bülthoff, 1998; Ullman, 1996). However, although we found reliable inversion effects, the maximal increase in processing time was about 20 ms. Thus, if a normalization mechanism (e.g., mental rotation) had to be done at the neuronal level it would have to fit in this demanding 20-ms time window. Instead, it has been suggested that whatever the orientation, neuronal responses start to accumulate at the same latency at the population level (Perrett et al., 1998). Life experience, in which stimuli appear more often in the upright orientation, would bias the population selectivity so that more cells respond to upright than inverted stimuli (Ashbridge et al., 2000). As a consequence, neuronal responses would accumulate faster to reach the categorization threshold in the former rather than in the later case. By integrating both category and orientation biases in this simple mechanism, it is possible to explain the larger orientation effect on processing speed in the human than in animal face task. Our results support this view because we did find a robust inversion effect for animals. Again, this explanation directly supports models of object processing in which there are quantitative rather than qualitative differences between human faces and other object categories. From the point of view emphasized in the first section of this discussion, larger inversion effects for human faces might be found as task requirements become more demanding. Indeed, if the strength of the inversion effect was stronger for human faces than for animal faces in the superordinate categorization task used here, this difference was not extremely important, and might be related to task instructions. A more important disruption of human face processing compared to other objects is found when subjects are asked to perform a recognition task (Diamond & Carey, 1986; Yin, 1969). This effect might be explained by the use of more specific representations that are themselves more specifically tuned to the orientation in which they have been learned. In keeping with this hypothesis, it has been shown that non-face object categories can present the same inversion effect as faces in a recognition task if subjects are experts at distinguishing between individuals of these categories (Diamond & Carey, 1986; Gauthier & Tarr, 1997). It follows that an apparent dichotomy between face and non-face object processing, such as the strength of the inversion effect, is not necessarily the hallmark of an independent face system; alternatively, it could reflect one point along a continuum of dynamically changing computational strategies (Riesenhuber & Poggio, 2002; Tarr & Cheng, 2003; Tarr & Gauthier, 2000).

These two experiments showed that in the context of natural scenes, faces are categorized following a time course very similar to another biological object category such as animals. Because it has been demonstrated that a nonbiological object category such as vehicles could be processed as efficiently as the animal category (VanRullen

& Thorpe, 2001a, 2001b), it might well be that every well known object category could be selected in a “glimpse” by a wave of processing in the ventral pathway (Riesenhuber & Poggio, 2000, 2002; VanRullen et al., 1998). Given the strong temporal constraints in these tasks, with selective responses appearing as early as 260 ms, such a fast coarse categorization process might rely on the activation of neurons selective to visual diagnostic properties by an essentially feed-forward flow of activation. Furthermore, the relatively weak inversion effects found in these experiments indicate that the representations activated to categorize a natural scene are relatively coarse, at least coarser than several high-level properties that have been found to be strongly affected by inversion (Tarr & Bülthoff, 1998). It might thus suggest that this kind of fast visual categorization of complex stimuli do not necessarily rely on similarly complex high-level representations, but might rather be achieved through the detection of diagnostic features of intermediate complexity (Ullman et al., 2002). Further experiments will be necessary to precisely determine the nature of these representations. This pattern of results is overall compatible with models that suggest the existence of a single object processing system whose performance is modulated by expertise, level of recognition, and information availability (Perrett et al., 1998; Schyns, 1998; Tarr & Cheng, 2003). The interplay between these different factors would determine the efficiency of the system, without requiring any face-specific module, or any mental rotation mechanism.

## Acknowledgments

We kindly acknowledge Nadège M. Bacon for her help in programming image presentation in Experiment 2 and Caitlin R. Sternberg and Anne-Sophie Paroissien for their help in testing subjects. We thank Roxane J. Itier and Rufin VanRullen for their valuable comments on an earlier version of the manuscript. This work was supported by the CNRS and the Cognitique grant n°IC2. Financial support was provided to both G.A.R. Rousselet and M.J.-M. Macé by a Ph.D. grant from the French government. Commercial relationships: none.

## References

- Ashbridge, E., Perrett, D. I., Oram, M. W., & Jellema, T. (2000). Effect of image orientation and size on object recognition: Responses of single units in the macaque monkey temporal cortex. *Cognitive Neuropsychology*, *17*, 13-34.
- 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. [PubMed]
- 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. [PubMed]
- 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. [PubMed]
- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*, 1-37. [PubMed]
- Delorme, A., Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2003). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. Manuscript submitted for publication.
- Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*, 107-117. [PubMed]
- Eimer, M. (2000). The face-specific N170 component reflects late stages in the structural encoding of faces. *Neuroreport*, *11*, 2319-2324. [PubMed]
- 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]
- Farah, M. J., Wilson, K. D., Drain, H. M., & Tanaka, J. R. (1995). The inverted face inversion effect in prosopagnosia: Evidence for mandatory, face-specific perceptual mechanisms. *Vision Research*, *35*, 2089-2093. [PubMed]
- Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychology Review*, *105*, 482-498. [PubMed]
- Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Vision Research*, *37*, 1673-1682. [PubMed]
- 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. [PubMed]
- 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. [PubMed]
- Halit, H., de Haan, M., & Johnson, M. H. (2000). Modulation of event-related potentials by prototypical and atypical faces. *Neuroreport*, *11*, 1871-1875. [PubMed]
- Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, *274*, 427-430. [PubMed]
- Itier, R. J., & Taylor, M. J. (2002). Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: A repetition study using ERPs. *Neuroimage*, *15*, 353-372. [PubMed]
- Jeffreys, D. (1996). Evoked potential studies of face and object processing. *Visual Cognition*, *3*, 1-38.
- Jolicoeur, P. (1988). Mental rotation and the identification of disoriented objects. *Canadian Journal of Psychology*, *42*, 461-478. [PubMed]
- Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*, 759-763. [PubMed]
- 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. [PubMed]
- Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, *5*, 910-916. [PubMed]
- Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Science*, *6*, 255-260. [PubMed]
- McElree, B., & Carrasco, M. (1999). The temporal dynamics of visual search: Evidence for parallel processing in feature and conjunction searches. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1517-1539. [PubMed]
- Mouchetant-Rostaing, Y., Giard, M. H., Bentin, S., Aguera, P. E., & Pernier, J. (2000a). Neurophysiological correlates of face gender processing in humans. *European Journal of Neuroscience*, *12*, 303-310. [PubMed]
- Mouchetant-Rostaing, Y., Giard, M. H., Delpuech, C., Echallier, J. F., & Pernier, J. (2000b). Early signs of visual categorization for biological and non-biological stimuli in humans. *Neuroreport*, *11*, 2521-2525. [PubMed]
- 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. [PubMed]

- 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. [PubMed]
- Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience*, *3* (Suppl.), 1199-1204. [PubMed]
- Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology*, *12*, 162-168. [PubMed]
- Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: An electrophysiological account of face-specific processes in the human brain. *Neuroreport*, *11*, 69-74. [PubMed]
- Rossion, B., & Gauthier, I. (2002). How does the brain process upright and inverted faces? *Behavioral and Cognitive Neuroscience Reviews*, *1*, 62-74.
- Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience*, *5*, 629-630. [PubMed]
- 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. [PubMed]
- Schyns, P. G. (1998). Diagnostic recognition: Task constraints, object information, and their interactions. *Cognition*, *67*, 147-179. [PubMed]
- Seeck, M., Michel, C. M., Mainwaring, N., Cosgrove, R., Blume, H., Ives, J., Landis, T., & Schomer, D. L. (1997). Evidence for rapid face recognition from human scalp and intracranial electrodes. *Neuroreport*, *8*, 2749-2754. [PubMed]
- Tarr, M. J., & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition*, *67*, 1-20. [PubMed]
- Tarr, M. J., & Cheng, Y. D. (2003). Learning to see faces and objects. *Trends in Cognitive Sciences*, *7*, 23-30. [PubMed]
- Tarr, M. J., & Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, *3*, 764-769. [PubMed]
- Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, *21*, 233-282. [PubMed]
- 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. [PubMed]
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520-522. [PubMed]
- Thorpe, S., & Imbert, M. (1989). Biological constraints on connectionist models. In R. Pfeifer, Z. Schreter, F. Fogelman-Soulié, & L. Steels (Eds.), *Connectionism in perspective* (pp. 63-92). Amsterdam: Elsevier.
- Thorpe, S. J., & Fabre-Thorpe, M. (2001). Seeking categories in the brain. *Science*, *291*, 260-263. [PubMed]
- Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & Bülthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *European Journal of Neuroscience*, *14*, 869-876. [PubMed]
- Trappenberg, T. P., Rolls, E. T., & Stringer, S. M. (2002). Effective Size of Receptive Fields of Inferior Temporal Visual Cortex in Natural Scenes. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in Neural Information Processing Systems 14*. Cambridge, MA: MIT Press.
- Ullman, S. (1996). High-level vision. Cambridge, MA: MIT Press.
- Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, *5*, 682-687. [PubMed]
- VanRullen, R., Gautrais, J., Delorme, A., & Thorpe, S. (1998). Face processing using one spike per neurone. *Biosystems*, *48*, 229-239. [PubMed]
- 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. [PubMed]
- 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]
- Vannucci, M., & Viggiano, M. P. (2000). Category effects on the processing of plane-rotated objects. *Perception*, *29*, 287-302. [PubMed]
- Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology*, *51*, 167-194. [PubMed]
- Yamamoto, S., & Kashikura, K. (1999). Speed of face recognition in humans: An event-related potentials study. *Neuroreport*, *10*, 3531-3534. [PubMed]
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*, 141-145.