Comparing animal and face processing in the context of natural scenes using a fast categorization task

Guillaume A. Rousselet[a],[1], Marc J.-M. Macé[a],
Michèle Fabre-Thorpe[a],*

[a] Centre de Recherche Cerveau et Cognition (UMR 5549, CNRS-UPS), 133 route de Narbonne,
31062 Toulouse Cedex 4, France

Abstract

Among objects, human faces might be special. The processing underlying the categorization of human faces and animals was compared in two experiments using upright and inverted natural scenes. Humans and animals were first presented in the full range of scales (from close-up to far views); stimuli were then restricted to face close-ups. Data showed (1) virtually no advantage for faces over animals; (2) very little impairment with inversion; and (3) greater sensitivity of faces to inversion. These results support a single object processing system in which evidence accumulates quickly to categorize objects, without requiring a face module or mental rotation mechanisms.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Object processing; Natural scenes; Faces; Animals; Temporal constraint; Inversion effect

1. Introduction

Among objects, the status of human faces might be very special. Growing evidence suggests that the processing of faces relies on a specific "module" distinct from the neuronal network underlying the visual processing of other objects [6]. Faces would be processed faster than any other object and brain activation linked to face categorization has been reported as early as 50–80 ms after stimulus onset [4,7], although brain

---

* Corresponding author. Tel.: +33-5-6217-2807; fax: +33-5-6217-2809.
 E-mail address: michele.fabre-thorpe@cerco.ups-tlse.fr (M. Fabre-Thorpe).
 [1] Present address: Department of Psychology, McMaster University, Hamilton, Ont., Canada.

activation specific to human faces is usually reported later, in the 120–200 ms latency range [1,5]. Moreover, because visual processing is more disrupted for faces than for other objects when presented upside down, face configuration might play an essential role in rapid face recognition [1]. The temporal processing cost associated with inverted faces could also be linked to mental rotation mechanisms [14]. The specificity of face processing is also observed on associated event related potentials (ERP). A negative potential peaking at around 150–170 ms over lateral occipito-temporal electrodes (the N170) appears very sensitive to human faces and to face inversion [1,5,10]. Some authors have suggested that it reflects the structural encoding of faces before recognition takes place [2] and that the inversion effect on the N170 shows the processing disruption of the spatial relationships between face components.

The potential existence of a separate module for face processing is a very important question in terms of computational models of visual perception. However, the idea of a face module is now questioned as another alternative explanation based on a biased distribution of selective neurons can account for fMRI experimental results [8]. Furthermore, the visual processing of other classes of objects can be also extremely fast. Indeed, in the context of natural scenes, categorization of biological and non-biological classes of objects such as animals and means of transport have been shown to induce specific brain cerebral activity that is related to categorization at about 150 ms after stimulus onset [16,19]. A processing speed that relies on highly optimized mechanisms [3] and that already sets a strong temporal constraint over the processing of visual information [17] and the underlying neuronal codes [20]. But the comparison with face processing is difficult as the vast majority of the studies addressing this question has been conducted with well-centered isolated faces of about the same size, same view angle and presented over a homogenous background. Thus, the results obtained in such conditions might reflect the very high homogeneity of the stimuli when compared to other object categories and may not apply in real world situations.

To address this question we ran two different experiments using natural scenes as stimuli to compare the processing speed of faces and animals in a fast categorization task. The strength of the inversion effect was evaluated by using upright and inverted natural scenes for both object classes. In the first experiment (Exp1), the natural scenes contained humans (and thus human faces) and animals in the full range of scales from close-ups to far views, presenting one or more targets in any location of the photographs. In the second experiment (Exp2), targets were restricted to close-ups of human face(s) and animal face(s) but still in the context of a natural scene. Accuracy, reaction times (RT) and associated brain activity were recorded during task performance. For detailed data see [11,12].

2. Material and methods

In both experiments, 24 subjects were tested on blocks of 96 stimuli (48 targets and 48 non-targets). They had to process humans as a target in half of the blocks and animals as a target in the other half; 16 blocks (1536 stimuli) were used in Exp1 and
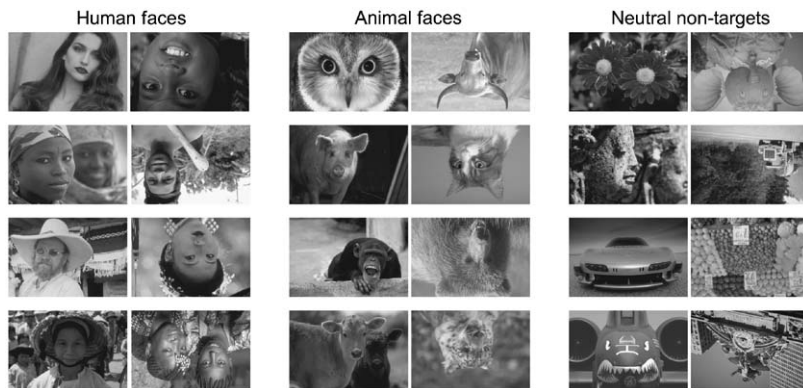
Fig. 1. Natural scenes with face close-ups used in Exp2. Targets were either human faces or animal faces. When performing a given task (i.e. human faces) the distractors included neutral photographs that did not contain a real face, and faces from the non-target category (i.e. animal faces).

8 (768 stimuli) in Exp2. All stimuli were natural scenes chosen to be as varied as possible (Fig. 1) and taken from a vast CD ROM database allowing access to several thousand stimuli. Images sustained about $20° \times 14°$ of visual angle (768×512 pixels) on a screen positioned at 1 m from the subject. They were flashed for only 20 ms in order to avoid eye movements. The 48 non-targets of a series included 24 neutral stimuli that were used in both tasks (landscape, urban scenes, buildings, tools, flowers, fruits, vegetables, plants and some tricky distractors such as dolls, statues or paintings, etc.) and 24 stimuli that belonged to the other class of objects (animals in the human task and humans in the animal task). To study the inversion effect any subset of stimuli was presented 50% in the upright position and 50% upside down. The 96 stimuli of a given block (targets and non-targets, upright and inverted stimuli) were mixed randomly. All stimuli were seen only once by a given subject in one orientation (upright or inverted) and with one status (target or non-target). All conditions were counterbalanced across subjects.

Subjects performed a go/no-go visual categorization task. They pressed a button to start the series and released it as fast as possible when the image contained an object of the target category (animal or human) that had been assigned through verbal instructions.

In order to analyze the N170 ERP component linked to the various stimuli used in our tasks, brain electrical activity was recorded from 32 electrodes mounted in an elastic cap in accordance with the 10–20 system (Oxford Instruments) using a SynAmps amplifier system (Neuroscan Inc.). Signals were digitized at a sampling rate of 1000 Hz and low-pass filtered at 100 Hz. Potentials were referenced on-line to the Cz electrode and averaged-referenced off-line. Two artifact rejections were applied to reject trials with eye movements and trials with excessive activity in the alpha range. ERP were computed separately for correct target trials and correct non-target trials.

3. Results

Both experiments revealed the remarkable object processing efficiency of our visual system and virtually no advantage for human faces over animals.

3.1. Accuracy and speed of performance

Accuracy scores were very similar for human and animal targets. This was true in both experiments with upright stimuli, (96.4% vs. 96.3% and 97.7% vs. 97.9%, respectively in Exp1 and Exp2) and with inverted stimuli (94.7% vs. 94.8% and 97.2% vs. 96.9%).

These accuracy scores are very high since subjects had to categorize successive stimuli with no a priori information concerning the type (target or non-target) and the orientation (upright or inverted) of the photograph that was going to be flashed for only 20 ms! The inversion effect was indeed very limited as the very weak decrease ($< 2\%$) seen in Exp1 was even smaller in Exp2 where it only reached significance for animal faces (1%).

Regarding the speed of processing, the results showed that subjects were not only very accurate but also very fast, regardless of whether the stimuli were shown upright or upside down.

The analysis of mean RT and RT distributions demonstrated that the processing of upright human faces and animals was done at the same speed when targets were shown at different scales (Exp1: 382 ms in both cases). With face close-ups, a non-significant 10 ms speed advantage was seen for human faces relatively to animal faces (Exp2: 382 ms vs. 392 ms) (Fig. 2, top row A).

Another index of processing speed is the minimal processing time. As targets and non-targets were equally probable, this value is computed as the latency from which on go-responses are always statistically biased in favor of correct go-responses. Minimal processing time was found at 260 ms for upright and inverted humans and for animals in Exp1. With close-ups of faces in Exp2 the minimal processing time was again found at 260 ms with upright human faces, but was increased by 10 ms (270 ms) with animal faces whether upright or inverted and by 20 ms (280 ms) with inverted human faces.

Thus, in the fast go/no-go categorization task used in our studies, the processing of upright human faces appears to have very little advantage (if any) over the processing of animals. Close-ups of human faces might be computed 10 ms faster than animal faces (Fig. 2, top row A), but the extreme similarity of human faces compared with the large variability of animal faces might account for this result.

The second important result concerns the slight inversion effect on accuracy and processing speed in the categorization tasks. Rather surprisingly, performance appears almost unaffected when stimuli are presented upside-down. The maximal accuracy impairment was seen in Exp1 that used a large range of target scales but even there, the impairment in accuracy was below 2% for both types of targets. Concerning the inversion effect on processing speed, a significant increase of mean RT was found for all kind of targets and was always more pronounced for humans than for animals (23 ms vs. 13 ms in Exp1; 14 ms vs. 10 ms in Exp2). However, as illustrated by the
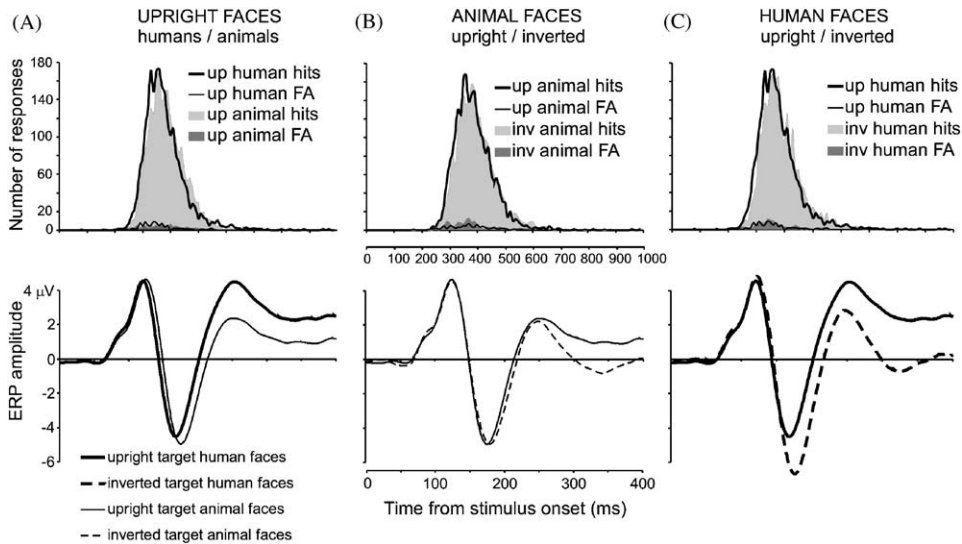
Fig. 2. Behavioral and EEG results with face close-ups used in Exp2. Top row shows the RT distributions (10 ms bins) of all go responses whether correct (hits) or incorrect (false alarm FA) in the various task conditions. Bottom row shows the grand-average ERP on electrode T6 to illustrate the effects observed on the N170. (A): Comparison of data obtained on human and animal faces when shown upright. (B): Comparison of data obtained on animal faces when seen upright or inverted. (C): Comparison of data obtained on human faces when seen upright or inverted.

overlap of the RT distributions shown in the top row of Fig. 2 (B, C), this increase in processing speed was very limited.

3.2. Brain activity: associated N170

A large N170 was observed in response to both kinds of faces in Exp2. Its amplitude was much larger than the one recorded with human or animal pictures in Exp1 or with other varied control natural scenes used as non-targets in both experiments. This finding confirms the sensitivity of the N170 to facial features. In the present paper we will focus on the modulation of the N170 in Exp2. One of the main questions addressed in this experiment concerned the sensitivity of the N170 to human and animal facial characteristics (Fig. 2, bottom row). For both targets, there was a non-significant tendency for a higher N170 peak amplitude with animal faces and the peak latency was significantly 6 ms shorter for human faces compared to animal faces.

The N170 recorded on inverted target trials showed two main effects. The N170 peak amplitude was enhanced only with inverted human faces and was delayed for both human and animal faces. This peak latency increase was more pronounced for human faces (6 ms) than for animal faces (3 ms). Thus, although behavioral differences between upside-down human and animal faces were very limited, the pattern of N170 inversion effects was clearly specific to human faces.

Finally, we addressed the question of the effect of task status on the N170. In both experiments, when subjects performed a given task (e.g. human faces), half of the non-target images were photographs used as targets in the other categorization task (e.g. animal faces). We were thus able to compare the N170 elicited by human and animal faces when processed as target or as non-target. Our analyses did not reveal any significant difference related to task status at the level of the N170 and all results reported above with human and animal targets were also seen when photographs were processed as non-targets.

4. Discussion

The high accuracy and speed of the human subjects in the two tasks are impressive and show how sophisticated fast mechanisms can be in the ventral pathway of the human brain, even when targets and non-targets shared global features as in the case of animal and human face close-ups. The high accuracy level reached in both experiments might be explained by the fact that faces constitute a very special object class, processed differently by our visual system, with no interference with other object categories. Despite this specificity, we found no clear evidence in favor of a computational advantage for human faces that would make them easier or faster to detect; thus our results argue against a hardwired face mechanism that would be more efficient than other non-face object mechanisms [15]. When presented in a full range of scale in the context of natural scenes, upright human faces and animals were being processed at the same speed. This result might be extended (at least) to another non-biological category of objects, as "means of transport" were shown to be categorized as fast as animals in a preceding study [19]. The only small advantage for upright human faces was found in Exp2, in which close-ups of human faces were processed 10 ms faster than animal faces, an effect that did not reach significance.

The ERP analysis showed that the N170 was not specific to human faces as already stated previously by others [10], since we also observed this component with animal faces. However, the N170 peaked at a shorter latency with human faces. This could be taken as evidence for the disruption of configural mechanisms dedicated to the processing of upright human faces, but could also simply reflect the higher variability in animal faces.

Whereas the N170 was not specific to human faces, the strength of the inversion effect was quite specific to human face pictures. Surprisingly, despite a larger N170 inversion effect for human faces compared to animal faces, the analysis of behavioral performance revealed very little effect of stimulus inversion associated with these two categories. Accuracy decreased from 0.5% to 1.7% with inversion, the maximum accuracy deficit being reached for both humans and animals in Exp1. The cost in processing speed was very limited although slightly more pronounced for humans (14–23 ms) than for animals (10–13 ms). With such temporal constraints, very little time would be available to implement a mental rotation mechanism during the time course of the categorization process.

The inversion effect is usually taken as evidence that face processing relies preferentially on configural mechanisms distinct from the part-based mechanisms used in the processing of other objects. Another alternative explanation states that recognition speed depends on the accumulation rate of object selective neuronal activity [9]. Neurons in higher-level occipito-temporal visual areas respond to complex stimuli such as animals and faces. Through experience, more neurons would selectively respond to upright animals and human faces or upright face features compared to inverted views. 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. It follows that, on average, it will take slightly more time to reach the decisional threshold with inverted than with upright stimuli. As a consequence, the stronger inversion effect for faces often explained by the specificity of face processing can find an alternative explanation in the accumulation rate of selective neural activity. In Exp2, if we take into account the large variability of animal faces and the more homogeneous set of human faces, the same reasoning applies to explain the slightly delayed responses for upright animal faces versus upright human faces.

These results obtained here can be interpreted in the framework of a single object processing system in the ventral pathway whose performance is modulated by expertise, level of recognition and information availability [9,13]. The data suggest that evidence accumulates very quickly and efficiently during the categorization of visual objects, without necessarily requiring a specialized face module or mental rotation mechanisms. Furthermore, the relatively weak inversion effects observed might indicate that the representations underlying categorization in our tasks are relatively coarse, at least coarser than several high-level properties that have been reported to be strongly affected by inversion [14]. The performance of such categorization task might rely on diagnostic features of intermediate complexity [18].

Acknowledgements

References

[1] S. Bentin, T. Allison, A. Puce, E. Perez, G. McCarthy, Electrophysiological studies of face perception in humans, J. Cognitive Neurosci. 8 (1996) 551–565.

[2] M. Eimer, The face-specific N170 component reflects late stages in the structural encoding of faces, Neuroreport 11 (2000) 2319–2324.

[3] M. Fabre-Thorpe, A. Delorme, C. Marlot, S. Thorpe, A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes, J. Cognitive Neurosci. 13 (2001) 171–180.

[4] N. George, B. Jemel, N. Fiori, B. Renault, Face and shape repetition effects in humans: a spatio-temporal ERP study, Neuroreport 8 (1997) 1417–1423.

[5] R.J. Itier, M.J. Taylor, Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: a repetition study using ERPs, Neuroimage 15 (2002) 353–372.

[6] N. Kanwisher, Domain specificity in face perception, Nat. Neurosci. 3 (2000) 759–763.

[7] Y. Mouchetant-Rostaing, M.H. Giard, C. Delpuech, J.F. Echallier, J. Pernier, Early signs of visual categorization for biological and non-biological stimuli in humans, Neuroreport 11 (2000) 2521–2525.

[8] H. Op de Beeck, J. Wagemans, R. Vogel, Can neuroimaging really tell us what the human brain is doing? The relevance of indirect measures of population activity, Acta Psycholog. 107 (2001) 323–351.

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

[10] B. Rossion, I. Gauthier, M.J. Tarr, P. Despland, R. Bruyer, S. Linotte, M. Crommelinck, 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 (2000) 69–74.

[11] G.A. Rousselet, M.J.-M. Macé, M. Fabre-Thorpe, Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes, J. Vision 3 (2003) 440–455.

[12] G.A. Rousselet, M.J.-M. Macé, M. Fabre-Thorpe, Animal and human faces in natural scenes: how specific to human faces is the N170 ERP component, J. Vision 4 (2004) 13–21.

[13] P.G. Schyns, Diagnostic recognition: task constraints, object information, and their interactions, Cognition 67 (1998) 147–179.

[14] M.J. Tarr, H.H. Bülthoff, Image-based object recognition in man, monkey and machine, Cognition 67 (1998) 1–20.

[15] M.J. Tarr, Y.D. Cheng, Learning to see faces and objects, Trends Cogn. Sci. 7 (2003) 23–30.

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

[17] S.J. Thorpe, M. Fabre-Thorpe, Seeking categories in the brain, Science 291 (2001) 260–263.

[18] S. Ullman, M. Vidal-Naquet, E. Sali, Visual features of intermediate complexity and their use in classification, Nat. Neurosci. 5 (2002) 682–687.

[19] R. VanRullen, S.J. Thorpe, The time course of visual processing: from early perception to decision-making, J. Cognitive Neurosci. 13 (2001) 454–461.

[20] R. VanRullen, S.J. Thorpe, Surfing a spike wave down the ventral stream, Vision Res. 42 (2002) 2593–2615.

Guillaume Rousselet is a postdoctoral fellow in the department of psychology at McMaster University, under the supervision of Allison Sekuler and Patrick Bennett. He did his Ph.D. in Cognitive Neuroscience at the Centre de Recherche Cerveau et Cognition in Toulouse, France. His background is in Biology and Cognitive Sciences. He is working on the rapid visual processing of objects and natural scenes in human subjects. Ongoing projects involve shape representation and limits of parallelism in the simultaneous processing of objects in young and old adults, as well as background categorization and face representation.

Marc Macé is a Ph.D. student in Cognitive Neuroscience at the Centre de Recherche Cerveau et Cognition in Toulouse, France. He studied Biology and Cognitive Sciences in France and in the States. Using psychophysics and EEG recordings in humans and monkeys, he works on the fast processing of natural scenes in categorization tasks at various hierarchical category levels. Challenging the visual system with manipulated stimuli, he documents the role played by the coarse magnocellular visual information in fast object recognition.

Michèle Fabre-Thorpe (Ph.D.; "French Thèse d'Etat") is a CNRS Research Director. She studied the visual control of reaching movement in the "Institut des Neurosciences" in Paris. In 1993 she moved to Toulouse to help set-up the Centre de Recherche Cerveau et Cognition directed by Michel Imbert. Since, her research interests are focused on the fast visual mechanisms that allow humans and monkeys to process natural scenes and perform high-level cognitive tasks such as object categorization with a speed that is a severe constraint to models of visual processing. Her studies use a comparative approach in humans and monkeys with a variety of methods such as behavioral, evoked potentials and event related fMRI techniques.