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

Guillaume A. Rousselet, Marc J-M. Macé & Michèle Fabre-Thorpe
Centre de recherche Cerveau et Cognition (UMR 5549, CNRS-UPS)
133 route de Narbonne, 31062 Toulouse, Cedex, France

Categorization of objects can be extremely fast. Indeed, in the context of natural scenes, 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 categorisation at about 150 ms after stimulus onset (Thorpe et al., 1996; VanRullen & Thorpe, 2001). But among objects, the status of human faces might be very special. Growing evidence suggests that a specific module might be dedicated to faces. Faces would be processed faster than any other object and, because of the observed strength of the inversion effect, face configuration might play an essential role in rapid face recognition. However, face studies are usually conducted with well-centred isolated faces and the results obtained in such conditions might not apply in real world situations, when faces are seen at different distances, under various angles and in competition with numerous other stimuli.

To address this question we ran two different experiments using natural scenes as stimuli to compare (1) the processing speed of faces and animals in a fast categorization task and (2) the strength of the inversion effect for both object classes with upright and inverted natural scenes. In experiment 1, the natural scenes contained humans (and thus human faces) and animals in the full range of scales from close-up to far views presenting one or more targets in any location of the photograph. In experiment 2, targets were restricted to close-ups of human face(s) and animal face(s) but still in the context of a natural scene.

In the "contextual human face task" of experiment 1, 24 subjects were tested on 16 blocks (8 blocks with humans as a target, 8 blocks with animals as a target) of 96 stimuli (48 targets and 48 non targets). All 1536 stimuli were chosen to be as varied as possible, they were natural scenes taken from a vast CD ROM data base allowing access to several thousand stimuli. The 48 non-targets included 24 neutral stimuli that could be used in both tasks (landscape, urban scenes, buildings, tools, flowers, fruits, vegetables, plants and some tricky distractors such as dolls, statues or paintings…) and 24 stimuli that belong 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 over the group of subjects. The same conditions apply to experiment 2, but subjects were tested on 8 blocks using 768 stimuli.

In all tasks, we used a *go/no-go* visual categorization task in which subjects were shown the 96 images in sequence (768 by 512 pixels, sustaining about 20° by 14° of visual angle) on a screen positioned at 1 m from the subject. Images were flashed for only 20ms in order to avoid the use of eye movements. The subjects had to press a button to start the series and to release the button as fast as possible when the image belonged to the target category (animal or human) that had been assigned to him through verbal instructions.

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

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

Assessed with the mean RT and statistical comparison on RT distributions, human face processing only showed a speed advantage (about 10 ms) with upright stimuli in Exp2 (382 ms vs. 392 ms). When comparing the minimal processing time in all tasks (a value computed as the latency at which *go*-responses are statistically biased in favour of correct *go*-responses -as targets and non-targets were equally probable) this value turned out to be identical in all cases (260 ms) except for upright close-ups of human faces in Exp 2 (250 ms).

Thus, in a fast categorisation task, the processing of human faces appears to have very little advantage (if any) over processing of animals. Human faces might be computed 10 ms faster than animal faces, but a predictive model (Rao & Ballard, 1997) taking in consideration the extreme similarity of human faces compared with the large variability of animal faces might just account for this result. The second important result concerns the very slight inversion effect seen with this fast categorization task. Performance appears almost unaffected when stimuli are presented upside down: the maximal impairment was seen in Exp1 that used various target scales but even there, the impairment in accuracy was below 2% for both types of targets. The increase in reaction time with inversion was slightly more pronounced for humans (23 ms) than for animals (13 ms).

These results argue against a hardwired face mechanism that would be more efficient than other non-face object mechanisms in daily life (Tarr & Cheng, 2003). They 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. The data suggest that evidence accumulate very quickly and efficiently during the categorization of visual objects, without necessarily requiring a specialized face module or mental rotation mechanisms. 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.