Rapid categorization of natural scenes in monkeys: target predictability and processing speed

Marc J-M. Macé CA, Ghislaine Richard, Arnaud Delorme[1] and Michèle Fabre-Thorpe

Centre de recherche Cerveau & Cognition, CNRS-UPS - UMR 5549, Faculté de Médecine de Rangueil, 133 route de Narbonne, 31062 Toulouse Cedex, France

CACorresponding author: marc.mace@cerco.ups-tlse.fr

[1]present address: Computational Neurobiology Laboratory, Salk Institute, 10010 N. Torrey Pines Road, San Diego, CA 92037, USA

Received 9 December 2004; accepted 14 January 2005

Three monkeys performed a categorization task and a recognition task with briefly flashed natural images, using in alternation either a large variety of familiar target images (animal or food) or a single (totally predictable) target. Processing time was 20 ms shorter in the recognition task in which false alarms showed that monkeys relied on low-level cues (color, form, orientation…). The 20-ms additional delay necessary in monkeys to perform the categorization task is compared to the 40-ms delay previously found for humans performing similar tasks. With such short additional processing time, it is argued that neither monkeys nor humans have time to develop a fully integrated object representation in the categorization task and must rely on coarse intermediate representations.

Keywords: Categorization, Early Visual Processing, Low-level cues, Macaques, Natural scenes, Recognition, Target predictability.

INTRODUCTION

Although less documented than for pigeons or for avians in general, the ability of monkeys to categorize complex visual photographs has now been shown for a variety of categories from subordinate to superordinate levels such as kingfishers, birds, fish, trees, primates, animals, food objects [1-5]. Baboons have been shown to develop multimodal abstract concepts of human and baboon categories [6] and can make judgments of conceptual identity [7]. When performing categorization tasks with very severe temporal constraints, macaque monkeys are able to produce their motor response with very short reaction times (RT). Their earliest correct responses are observed at a latency of 180 ms [5,8], a delay shown to challenge many models of object processing [9,10]. Performing very similar tasks, human subjects, although very fast, are much slower than monkeys with their earliest behavioral responses observed at about 270 ms after stimulus onset [11].

With such differences in minimal input-output processing time (180ms in monkeys versus 270 ms in humans), one should wonder how similar is the neural processing underlying visual categorization in humans and monkeys. These very short response latencies observed in monkeys might result from the processing of low-level cues rather than the use of abstract representations. Indeed Torralba and Oliva [12] have shown that, in humans, the statistics of low-level features across natural images can be used to prime the presence or absence of objects in the scene and to predict their location before exploring the image. However, the initial use of low-level cues might be as important for monkeys as for humans. Alternative –and non exclusive- explanations could also account for the rapidity of monkeys in these tasks. First, one cannot exclude a speed accuracy trade-off since monkeys are slightly less accurate than humans (about 90% vs 94% correct). Second, it could simply result from shorter conduction delays because of macaques smaller brain dimensions.

This study had mainly two aims. First we wanted to compare in monkeys the visual processing of natural images in two tasks, one during which the monkey might need to rely on the abstract representation of a high-level object category such as “food” or "animal", and another where the target was totally predictable, so that the monkey could respond using a limited number of specific low-level cues stored in short-term memory. Faced with similar tasks, human subjects are faster when target predictability is total [11]. Since in humans and monkeys, both tasks used natural photographs as stimuli and required the same motor response, any differences in the latencies of the motor responses should reflect central processing differences related to task demands. Thus, a second aim of the study was to compare how the different requirements of the two tasks would affect monkey and human performance.

METHODS

Participants: Three rhesus monkeys were trained to perform a rapid go/no-go visual superordinate categorization task with food-objects (Rh1, male aged 7) or animals (Rh2 and Rh3, male and female aged 6 and 5) as targets. These monkeys have already been tested in different experiments that showed their ability to categorize familiar photographs and generalize to new photographs [5] and revealed that color did not play a crucial role in rapid categorization [8]. All procedures conformed to French and European standards concerning the use of experimental animals and the protocols were approved by the regional ethical committee.

Tasks and protocol: In the rapid categorization task, monkeys were presented with a random succession of different natural scenes, half of which were targets. The monkeys started stimulus presentation by placing one hand on a capacitive tactile key. When a target image was flashed, they had to quickly release the button and touch the screen (go response), otherwise they had to keep their hand on the button (no-go response). They were given a maximum of 1000 ms to respond, after which any response was considered as a no-go response.

In the second task, monkeys were still performing the same go/no-go task, except that only one single (food or animal) target was used and presented among varied non-targets images. As this single target was totally predictable, the task performed by the monkey became a “recognition task”. Targets and non-targets were still equiprobable as in the categorization task.

To perform the tasks, monkeys were sat about 30-35 cm away from a tactile screen. A small fixation point appeared in the center of the screen and pictures were flashed around the fixation point on a black background for only 28 ms; a duration that prevented any exploratory eye movements. The tactile key used to start the sequence of images and to record the motor response was located below the screen at waist level. Two successive images were separated by a random 1.5-3 s intertrial period. Correct (go or no-go) responses were rewarded by a drop of fruit juice and a beep noise. Incorrect decisions were followed by a 3-4 s display of the incorrectly classified stimuli delaying the next trial and allowing time for ocular exploration. The monkeys worked daily for as long as they wanted (1-3 hours), five days a week. At the end of each testing session and during week-ends *ad libitum* water was provided. They were restrained in a primate chair during testing (Crist Instruments, Georgia, USA).

The results presented here were recorded during 20 successive testing sessions. In a given testing session, monkeys performed in alternation the categorization task with the recognition task by blocks of 150 trials. Before the start of the testing session, monkeys performed the categorization task until they were calm and up to their usual level of performance. The testing session started first by a block of 150 trials of the categorization task using 150 different stimuli. Warning that the categorization task was going to become a recognition task was then given through a sequence of 10 trials presenting repetitively the single target image that was going to be presented 75 times randomly among 75 different non-target images in the subsequent recognition block. No warning was given in between the recognition block and the next categorization block.

For each of the 20 sessions, the 150 stimuli of the categorization task together with the 75 non-target stimuli and the single target stimulus of the recognition task were chosen at random from the pool of familiar images that the monkey had already categorized many times (the two monkeys working on the animal/non-animal categorization were tested on the same single target stimuli).

Thus the animals alternated between a task in which training had optimized stimulus processing and a task in which target predictability was total. In each session the monkey alternated categorization task blocks with recognition task blocks, until they stopped working on the task. A minimum of 2 blocks in each task was required for a session. Thus, a session was usually run on one day, exceptionally on two successive days.

Stimuli: All stimuli (examples in Figure 1) used in the tasks were color photographs of natural scenes (Corel CD-ROM library). Targets and distractors included both closeups and general views. Food-targets included photographs of fruit, vegetables, salads, cakes, biscuits, sweets... presented against natural backgrounds. Animal-targets included fish, birds, mammals and reptiles also presented in their natural environments. Distractors included landscapes, trees, flowers, objects, monuments, cars and some targets of the other categorization task....

Images (192 x 128 pixels, corresponding to an angular size of about 25° X 15°) were mostly horizontal photographs (73%). They were flashed for 2 frames at a refresh rate of 60 Hz (non-interlaced), using a programmable graphics board (VSG 2, Cambridge Research Systems) mounted in a PC-compatible computer.

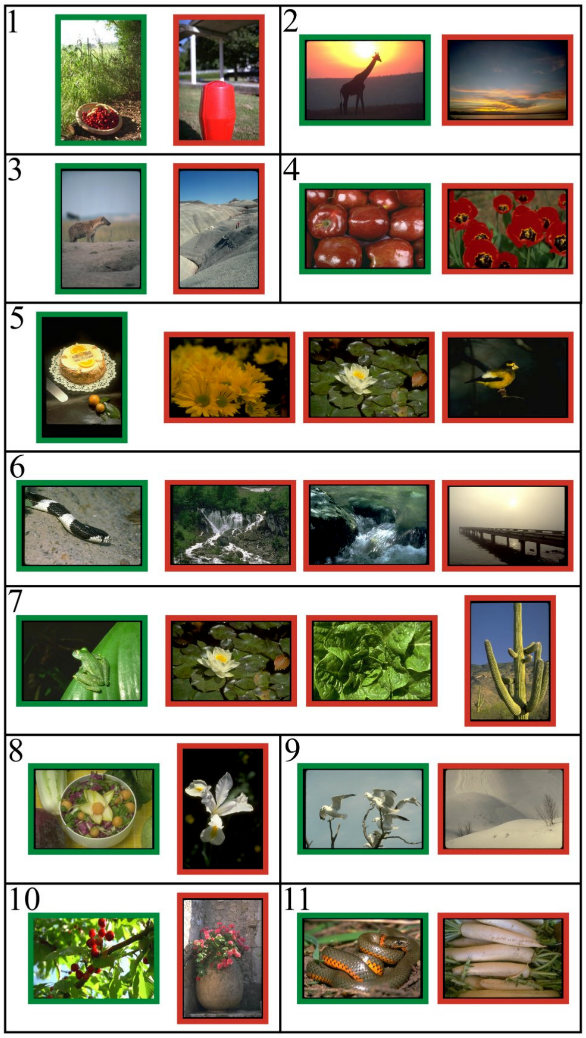

Fig 1. Targets (with a green frame) and associated errors (with a red frame) in the recognition task. The figure shows the high variety of stimuli used in the 20 testing blocks in each of the tasks. On the right of each target image is shown 1-3 non-target photograph(s) that induced a false alarm. Some errors can clearly be related to prevailing color (1-7), global form (4, 6, 11), global orientation (3, 6, 11), color patches or specific form in specific locations (4-6, 8, 10), spatial layout of the scene (3, 6-11) or any combination. Similar natural images were used in the categorization task. Note that, with target n°3, both monkeys made the same error.

RESULTS

We will first present the results obtained on the monkeys in the two tasks, then briefly present the results obtained in a group of human subjects to compare the effect of target predictability on processing speed in both species (the detailed human subject data have been published in Delorme et al., 2004).

Behavioral results. Categorization vs. Recognition tasks: The analysis of the monkey behavioral performance included accuracy, speed of response and a study of the non-target images that incorrectly induced a go response in the recognition task. Results concerning the food-target task will be given for three daily blocks of categorization and recognition tasks: a total of 9000 trials in each task for monkey Rh1. Results concerning the animal-target task will be given for two daily blocks of categorization and recognition tasks: a total of 6000 trials in each task and for each of the two animals.

Accuracy. Although very high in both tasks (92.4 % correct in the categorization task; 96.3 % correct in the recognition task), accuracy was significantly better in the recognition task (two-tailed χ^2, df =1, p < .0001). This effect was present regardless of the target to find (food or animal), and was significant at p < .0001 for each individual subject.

With animal targets, both monkeys were better at responding to target images than at ignoring distractors. This bias in favor of correct go responses was quite pronounced (about 4-5%) and was observed with the same strength in both tasks and for both monkeys (two-tailed χ^2, df=1, p always <.0001).

The third monkey working with food targets did not show the same biased pattern. A small but significant bias (1.3%) was observed in the recognition task in favor of correct go responses whereas a stronger inverted bias (<3%) was seen toward correct no-go responses in the categorization task.

Reaction time (RT). Monkeys were extremely fast in both tasks, but as illustrated in Figure 2, reaction times were always faster for the recognition task than for the categorization task. Overall, for the 3 monkeys, the processing time was shorter by 19 ms (mean RT: 244 ms vs 263 ms). It was shorter by 17 ms for Rh1 working with food objects (mean RT 281 ms vs 298 ms) and by 21 ms (14 ms for Rh2 and 28 ms for Rh3) with animal targets (mean RT 225 ms vs 246 ms). For each individual subject these differences were always significant (two-tailed t-test, p < .0001).

Monkeys alternated between the categorization and the recognition tasks by blocks of 150 trials. Performance speed and accuracy were analyzed separately for each section of 50 trials to determine the stability of the performance throughout the progress of the 150 trials block. The performance was very stable showing that the monkey modified its behavioral strategy as soon as the task changed.

d' analysis. The average slower speed in the categorization task could result from the presence of some difficult photographs that need longer to process. Thus, although the average processing time could be shorter in the recognition task, the latencies of the earliest responses might be similar in both tasks. This is clearly not the case and this shortening of RT latencies concerns the whole range of motor responses from the very first responses triggered. RT distribution in the recognition task can be seen (Figure 2) as a shift of the entire RT distribution obtained on the categorization task toward shorter latencies. The dynamic d’ calculated for both tasks and for each monkey (Figure 2) illustrates clearly that the earliest responses triggered have shorter latencies in the recognition task compared to the categorization task.

The minimal reaction time, evaluated as the first time bin from which correct hits significantly outnumbered false alarms, reflects the minimal processing time necessary to reach decision in each task. The difference in minimal RT between the 2 tasks was 20 ms on average and was significant for each animal. It is clearly visible on the d’ curves presented for each monkey in Figure 2. Thus the increase of latencies seen on the mean processing time is also present with the same strength on the very first behavioral responses observed.

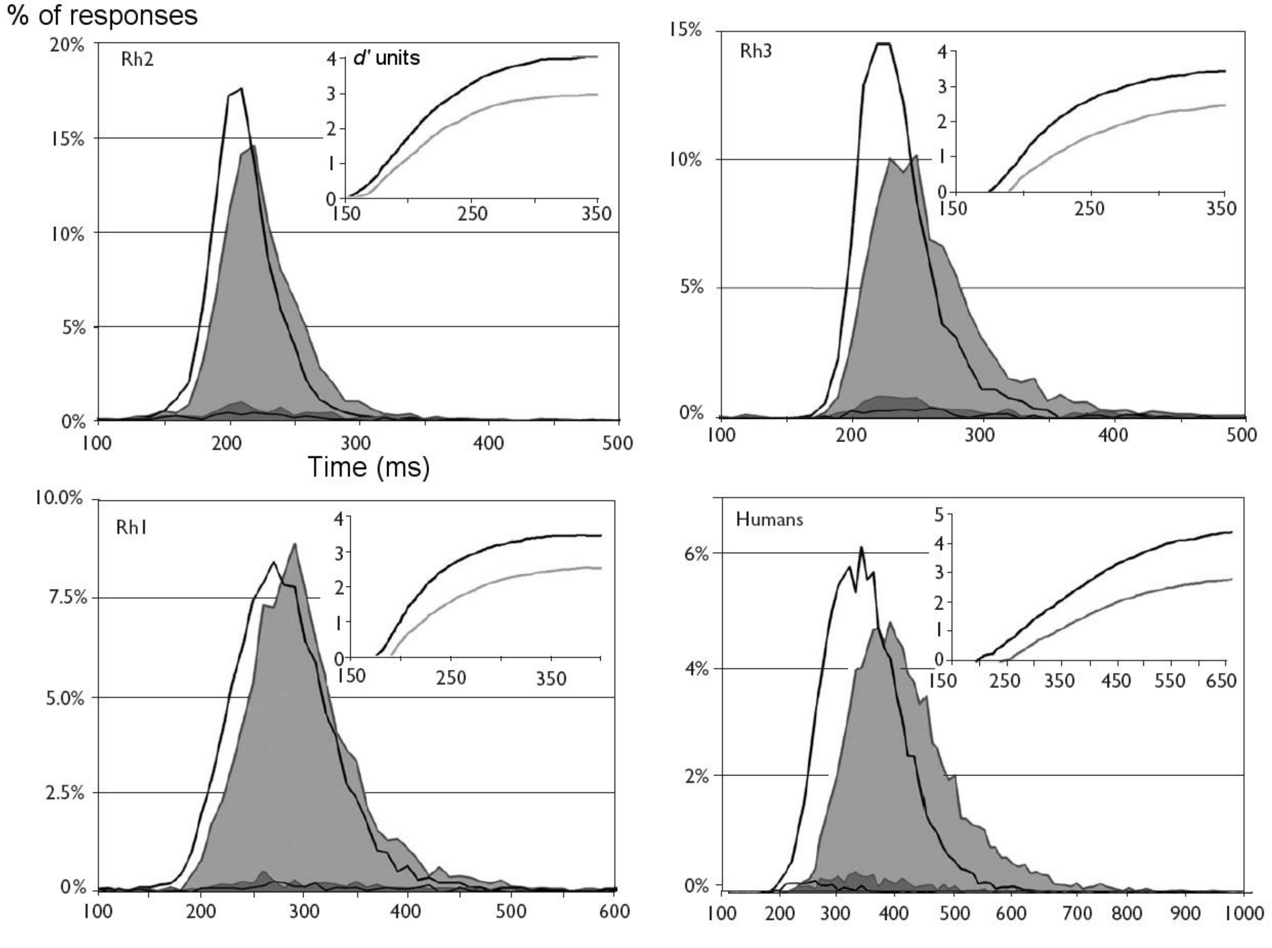

Fig 2. Overall reaction time distributions of go responses in both the categorization task (gray traces and shaded distribution) and the recognition task (black lines) for each of the 3 monkeys (Rh1, Rh2 and Rh3) and for the group of 14 human subjects. The two thick traces are for correct go responses towards targets, the two thin traces for false alarms induced by non-target stimuli. Rh2, Rh3, and humans were tested with an animal task, Rh1, with a food-object task. In the top right hand corner, a "dynamic" *d'* (see details in [23]) is calculated for each task from the cumulative number of hits and false alarms at each successive 10ms time bin (gray trace: categorization task; black trace: recognition task). Note that although the scales of the axes are not identical in all illustrations, RT distributions and corresponding *d'* share the same time scale.

Errors. A question that needs to be raised concerns the kind of errors that are produced in both tasks. So far, on all data collected when monkeys were performing the categorization task in this study and others, it has not been possible to determine the reasons for these false alarms. However, in the recognition task, the non-target images that induced errors often shared some obvious low-level properties with the memorized target image. These features (see Figure 1) appear to be often related to the prevailing color, the global form of objects or their coarse orientation and sometimes to the spatial layout or complexity of the whole scene… When performing the recognition task, monkeys could thus rely on low-level memorized visual cue(s). Because the distractors in the recognition task were chosen at random in the pool of familiar images already seen by the monkeys, it was rare that monkeys Rh2 and Rh3 had to deal with the same distractor image when looking for the same single target image. Thus it is worth noting that with the target image n°3 (Figure 1) both monkeys made the same error that can be seen as induced by the prevailing color and/or the global layout of the scene.

Behavioral results. Monkeys vs. humans: The neural mechanisms underlying visual categorization are still poorly understood and their similarity in humans and animals is very controversial. Therefore it is interesting to compare monkey and human performance in these two tasks to determine whether or not increasing target predictability has the same behavioral consequences in both species. A group of 14 humans had been studied on the same two tasks using animal targets as monkeys Rh2 and Rh3 of the present study (see Delorme et al., 2004 for a detailed report of their performance and associated cerebral activity).

As with monkeys, human accuracy was higher in the recognition task (98.7 %) than in the categorization task (93.1 %). Humans were slightly better at ignoring distractors in the categorization task, whereas they were better at responding to targets in the recognition task. Thus humans and monkeys showed the same bias in the recognition task but the discrepancy in the categorization task merits further investigation in humans. Re-analysis of the human data in the categorization task showed that individual biases in humans were tightly correlated with response speed. Humans favoring speed over accuracy would display the same bias as Rh2 and Rh3, the two fastest monkeys. Conversely human subjects favoring accuracy over speed would display a bias similar to those of monkey Rh1, the slowest monkey. Thus, this performance bias is similar in humans and monkeys and is highly dependent of the individual speed-accuracy trade-off of the subject.

Concerning response speed, mean RT was shortened by 63 ms in human in the recognition task relatively to the categorization task, a larger difference than the one found in monkey. But as in mokeys, the entire RT distribution was also shifted toward shorter latencies. As illustrated in figure 2, the fastest responses were observed at about 220 ms in the recognition task, corresponding to a 40 ms speed advantage for this task (20 ms in monkeys). The analysis of the stimuli that induced errors in the recognition task also showed, as reported for monkeys in the present study, that humans were most likely relying on low-level visual characteristics of the memorized target.

DISCUSSION

One of the aims of the present study was to determine the time necessary for monkeys to process a natural image on the basis of low-level cues and to evaluate how much more processing they would need to perform a categorization task in which they presumably rely on more abstract representations. Indeed, the results obtained show that the time needed by monkeys to process natural images does depend on the task performed. First, in the recognition task in which monkeys had to recognize a given target image, the visual similarity between the target and the non-target scenes that induced erroneous go response, strongly suggests that –as intended- monkeys relied on low-level cues that varied from target to target. Then, the comparison of monkeys performance in the recognition task relatively to the categorization task showed that monkeys, were both more accurate (by about 4%) and faster when they responded to a single-target image. Testing humans with similar tasks had similar effects on performance but with larger amplitude: their accuracy was increased by 5.6% and their processing time was decreased by 40 ms when considering the earliest responses produced and by about 60 ms for mean RT. The processing required for deciding whether an animal is present in a natural scene, takes at least an additional delay of 20 milliseconds in monkeys and of 40-60 ms in humans.

Why is additional processing time longer in humans than in monkeys? A first interpretation lies in the stimuli used to run the categorization task. Whereas monkeys were tested with familiar images that they had already categorized many times, human were tested with stimuli that they all saw for the first time [11]. However this discrepancy can only explain the increased difference seen when comparing mean reaction times (60 ms). When humans are tested with both new and familiar images, they produce their earliest responses at exactly the same latencies. The only effect of familiarity is to shorten RT of long latencies responses with a resulting effect of decreasing the mean reaction time by about 20 ms [13]. Thus the additional processing delays that should be compared between the two species are the delays seen on the earliest responses: 20 ms in monkeys vs 40 ms in humans. A straightforward cognitive interpretation concerns the type of representations used by monkeys in the categorization task that might be less abstract and more figurative than in humans. But an alternative interpretation is to consider that the 20 ms delay in monkeys is simply the homologue of the 40 ms delay in humans. In fact, monkeys always produced their motor response faster than humans. This has mainly been reported for ocular movements. Express saccades for example are seen at 70 ms in monkeys and 100 ms in humans [14]; vergence reflex or tracking systems are observed at latencies of 55-60 ms in monkeys and 80 ms in humans [15-17]. Visuo-motor responses would thus be produced by monkeys at latencies that are about two thirds of the human latencies. This is true also in the categorization task performed with familiar images that is used in the present study: monkeys have a mean RT of about 263 ms whereas human mean RT was observed at 424 ms [13]. Given that intracortical connections have been shown to be very slow [18,19], the differences between monkey and human latencies could just be due to differences in brain sizes and reflect that, in monkeys, less time is lost in transferring information within a given cortical area or along the different cortical areas [5,9].

Thus during the additional processing time required in between the recognition task (use of low-level cues) and the categorization task (abstract representation?), visual computations made by monkeys and humans might be very similar. This 20-40 ms temporal cost appears very limited when considering the discrepancy in task complexity. On the one hand, this additional delay argues strongly that, in the categorization task, monkeys and humans have to process visual information further than the simple low-level statistical differences in between target and distractor image sets shown by Torralba and Oliva [12]. On the other hand, this additional processing time is so short that neither monkeys nor humans would have time to rely on fully integrated high-level object representations. Thus when responding in the categorization task, fast responses might rely on very coarse intermediate object representations. This is in agreement with experimental series showing that humans can categorize natural scenes at extreme eccentricities [20] and that natural scenes categorization in human and monkeys is very robust even when using achromatic stimuli at very low contrast [21][Macé MJ-M et al., in preparation]. It might be that categorization tasks in which a detailed object representation is necessary would induce a more spectacular increase in reaction time both in humans and in monkeys. Further experiments are needed to evaluate such an interpretation.

REFERENCES

1. Roberts WA and Mazmanian DS. Concept learning at different levels of abstraction by pigeons, monkeys, and people. *J Exp Psychol Anim Behav Process* **14**, 247-260 (1988).
2. D'Amato MR and Van Sant P. The person concept in monkeys (Cebus apella). *J Exp Psychol Anim Behav Proc* **14**, 43-55 (1988).
3. Schrier AM. Learning-set formation by three species of macaque monkeys. *J Comp Physiol Psychol* **61**, 490-492 (1966).
4. Vogels R. Categorization of complex visual images by rhesus monkeys. Part 1: behavioural study. *Eur J Neurosci* **11**, 1223-1238 (1999).
5. Fabre-Thorpe M, Richard G and Thorpe SJ. Rapid categorization of natural images by rhesus monkeys. *Neuroreport* **9**, 303-308 (1998).
6. Martin-Malivel J and Fagot J. Cross-modal integration and conceptual categorization in baboons. *Behav Brain Res* **122**, 209-213. (2001).
7. Bovet D and Vauclair J. Judgment of conceptual identity in monkeys. *Psychon Bull Rev* **8**, 470-475 (2001).
8. Delorme A, Richard G and Fabre-Thorpe M. Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. *Vision Res* **40**, 2187-2200 (2000).
9. Thorpe SJ and Fabre-Thorpe M. Neuroscience. Seeking categories in the brain. *Science* **291**, 260-263. (2001).
10. VanRullen R and Thorpe SJ. Surfing a spike wave down the ventral stream. *Vision Res* **42**, 2593-2615. (2002).
11. Delorme A, Rousselet GA, Macé MJ-M and Fabre-Thorpe M. Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Brain Res Cogn Brain Res* **19**, 103-113 (2004).
12. Torralba A and Oliva A. Statistics of natural image categories. *Network* **14**, 391-412. (2003).
13. Fabre-Thorpe M, Delorme A, Marlot C and Thorpe S. A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *J Cogn Neurosci* **13**, 171-180. (2001).
14. Fischer B and Weber H. Express Saccades and Visual Attention. *Behav Brain Sci* **16**, 553-567 (1993).
15. Busettini C, Fitzgibbon EJ and Miles FA. Short-latency disparity vergence in humans. *J Neurophysiol* **85**, 1129-1152. (2001).
16. Busettini C, Miles FA and Krauzlis RJ. Short-latency disparity vergence responses and their dependence on a prior saccadic eye movement. *J Neurophysiol* **75**, 1392-1410. (1996).
17. Miles FA. The neural processing of 3-D visual information: evidence from eye movements. *Eur J Neurosci* **10**, 811-822. (1998).

18. Nowak LG and Bullier J. In: Rockland KS, Kaas JH and Peters A, eds. *Extrastriate visual cortex in primates*. New York: Plenum Press, 1997:205-241.
19. Frégnac Y and Bringuier V. In: Braitenberg AAV, ed. *Brain Theory - Biological basis and computational principles*Elsevier, 1996:143-199.
20. Thorpe SJ, Gegenfurtner, K. R., Fabre-Thorpe, M., Bulthoff, H. H. Detection of animals in natural images using far peripheral vision. *Eur J Neurosci* **14**, 869-876. (2001).
21. Macé MJ-M, Thorpe SJ and Fabre-Thorpe M. Rapid categorisation of achromatic natural scenes: how robust at very low contrasts? *Eur J Neurosci*, in press.
22. Rousselet GA, Macé MJ-M and Fabre-Thorpe M. Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *J Vis* **3**, 440-455. (2003).

Acknowledgment: This work was supported by the CNRS, the University Toulouse III, the Integrative and Computational Neuroscience ACI and the Cognitique program of the French government. Marc Mace and Arnaud Delorme were supported by an MRT grant from the French government.