

RAPID PROCESSING OF COMPLEX NATURAL SCENES : A ROLE FOR THE MAGNOCELLULAR VISUAL PATHWAYS?

Delorme A., Richard G., & Fabre-Thorpe M.

Centre de Recherche Cerveau & Cognition

133, route de Narbonne, 31062, Toulouse, France

arno@cerco.ups-tlse.fr, grichard@cerco.ups-tlse.fr, mft@cerco.ups-tlse.fr

Abstract:

Monkeys and Humans were tested on a rapid categorization task of briefly flashed (30ms) previously unseen natural images. They had to respond only when a target (animal or food) was present. Removing color information had no effect on the monkey's performance. For humans, a strong correlation between the use of color information and the RT duration appeared. Accuracy was identical with chromatic and achromatic photographs for the fastest humans, whereas subjects with longer RTs showed an increasing impairment when color was removed. Rapid identification may rely mainly on the fast feed-forward processing of coarse achromatic information in the magnocellular stream. These results impose strong constraints for modeling the first stages of visual processing.

Key words : Natural scenes, categorization, early processing, color, biological visual systems

1. Introduction

The neural modeling of the cerebral processes requires precise knowledge concerning both the kind of information processed by the neurons and the dynamics of neural populations. In the present study, monkeys are faced with an extremely challenging visual task, and the results provide constraints not only on the processing time but also on the earliest visual information necessary to complete the task. The remarkable speed and efficiency of processing in the visual system has recently been demonstrated in humans [1] and monkeys [2]. In both studies, subjects had to detect a given target (i.e. an animal) in previously unseen natural photographs that were flashed very briefly (20-80 ms) in order to avoid eye movement. The task could be completed with a high rate of success (94% of correct responses for humans, 88% for monkeys) and with very short reaction times. In fact, Thorpe et al, [1] showed that the sophisticated visual processing required in such a task could be done in under 150ms by the human visual system. With such short processing times, the authors argued that this sort of rapid categorization must depend mainly on fast feed-forward processing. Furthermore, they have argued that this kind of processing may be achieved using only one

spike per neuron [3], possibly using a code that involves the relative timing of spikes across a population [4]. The subsequent study with macaques increased the strength of their argument, since monkeys had even faster behavioral RT (250 and 350ms) and the faster monkey showed a statistical bias towards correct responses for behavioral RTs as short as 180-200ms. As this value also includes a substantial motor component, visual processing may be reduced to as little as 100ms.

With such constraints on the speed of processing, one might wonder what visual information is used in the early processing of a natural scene. Recent neurophysiological data has demonstrated that neurons in the visual cortex that get inputs from the magnocellular pathways tend to respond roughly 20ms earlier than cells in the chromatically sensitive parvocellular stream [5]. Given this fact, one should predict that any visual processing task which only used the earliest arriving information in each cortical area should effectively be color blind.

To test this idea, the present study directly investigated the role of color information in a very rapid categorization task in both humans and monkeys. We used two types of categorization tasks, one involving "food" as the target category, the other using "animals". For both categories one

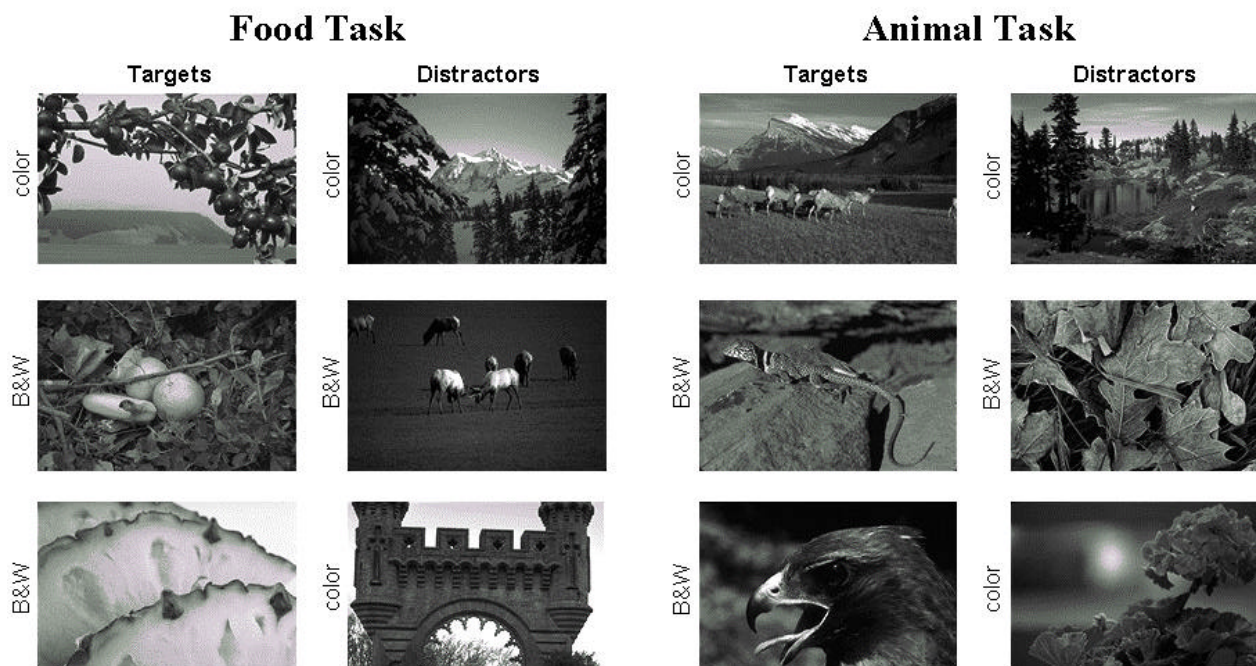


Figure 1 : Examples of natural stimuli used, in the food/non food categorization task and in the animal/non animal categorization task. To illustrate the diversity of the stimuli used in the task, three targets and three distractors, on which both the monkey and all 10 humans responded correctly are shown for each task. They were proposed either in color or in black and white (B&W) as indicated on the figure.

might predict that color could be very important for segmenting images, since both foods and animals are often brightly colored.

2. Material and methods

Two rhesus monkeys (M1 and M2) and two groups of 10 humans (GH1 and GH2) performed a visual categorization task in which they had to classify a succession of photographs that appeared in the center of a tactile screen. Trials started when the subject had positioned one hand on a button located just below the screen at waist level. Pictures were flashed for only 30 ms and, when the picture belonged to the target category, the subject had 1s to release the button and touch the screen; if not, he had to keep pressing the button. Correct decisions were signaled by a tone that was associated with a fruit juice reward in the case of the monkeys, whereas incorrect decisions were punished by showing the image again for 3s.

Each monkey was trained on a different categorization task : food vs. non food for M1, and animal vs. non animal for M2. All the pictures were natural scenes (Fig 1) taken from a large commercial CD-ROM library (Corel database) allowing access to several thousand stimuli, although additional

photographs were added for the food vs. non-food categorization task in order to allow further controls to be performed. Targets in the food task included photographs of fruit, vegetables, salads, cakes, biscuits, sweets ... presented against natural backgrounds. In the animal task, targets included fish, birds, mammals and reptiles in their natural environments. In both tasks the distractors included landscapes, trees, flowers, monuments, cars, and the target category of the other monkey. The targets were extremely varied and the monkeys had essentially no a priori knowledge concerning the position, the size or the number of targets in a picture. The images (192 x 128 pixels, 8-bit color converted from 24 bit color for both black and white and color photographs) were presented for 2 frames at a refresh rate of 62 Hz (non-interlaced), corresponding to a presentation time of 32 ms, using a programmable graphics board (VSG 2, Cambridge Research Systems) mounted in a PC-compatible computer. The monkey was typically at 25-30 cm from the screen so that the approximate angular size of the image was 15/25°.

Performance was studied using 400 images that the monkeys had never seen before (200 in color and 200 in black and white out of which 50% were distractors and 50%

were targets). For comparison, the GH1 and GH2 groups were tested using the same apparatus and the same 400 photographs on which M1 and M2 had respectively been tested. Human had to categorize 4 blocks of 100 new stimuli whereas monkeys were given 20 new photographs every day that were mixed with 80 familiar images on which they had been trained. The 20 new stimuli and the 80 familiar images were half targets and half distractors, half in color, half in black and white.

3. Results

For both monkeys and humans, the accuracy and speed of categorization reached when tested on the 200 colored photographs were compared with those obtained on the 200 photographs presented in black and white (Fig 2). The data show that removing color information has only marginal effects on this task.

3.1 In monkeys :

In order to avoid any learning effect, the results only consider the response made by the monkeys to the first presentation of each new image. Accuracy was similar with chromatic and achromatic new stimuli (respectively 90.5% vs. 91.5% of correct responses in the animal categorization and 85.5 vs. 87.5% in the food categorization) with very little effect on reaction time (median RT for color and black and white stimuli : 246 ms vs. 249 ms for the animal task and 278 vs. 287 ms for the food task). The small differences

observed on both accuracy and RT were not statistically significant.

3.2 In humans :

Considering the mean performance of each group of 10 subjects, the categorization accuracy was slightly lower with achromatic stimuli (93% vs. 95.7% of correct responses in the animal task and 89% vs. 90.7% in the food task), but the difference was only significant in the case of the animal task ($\chi^2 = 14.13$, d.f. = 1, $p = .0002$). However, as in monkeys, the absence of color cues had essentially no effect on reaction times (median RT : 392 ms in color vs. 395 ms in black and white for the animal task, respectively 407 ms vs. 419 ms for the food task). The difference was only statistically significant in case of the food task (two tailed t test, d.f. = 1807, $t = -4.3$, $p < .0001$) but it was very small indeed (12 ms).

The analysis of individual performances showed that some of the subjects were as accurate with black and white stimuli as they were with colored ones. It also showed a strong correlation between the accuracy advantage displayed by humans when categorizing colored stimuli and their median (or mean) reaction time (Fig 3). Thus the fastest subjects had no (or very little) accuracy impairment when color information was removed, and the accuracy impairment with achromatic stimuli increased progressively in subjects with longer RTs.

When considering the monkeys' performance together with the human performance, monkeys behave like the fastest humans, showing short RTs and no advantage for colored stimuli.

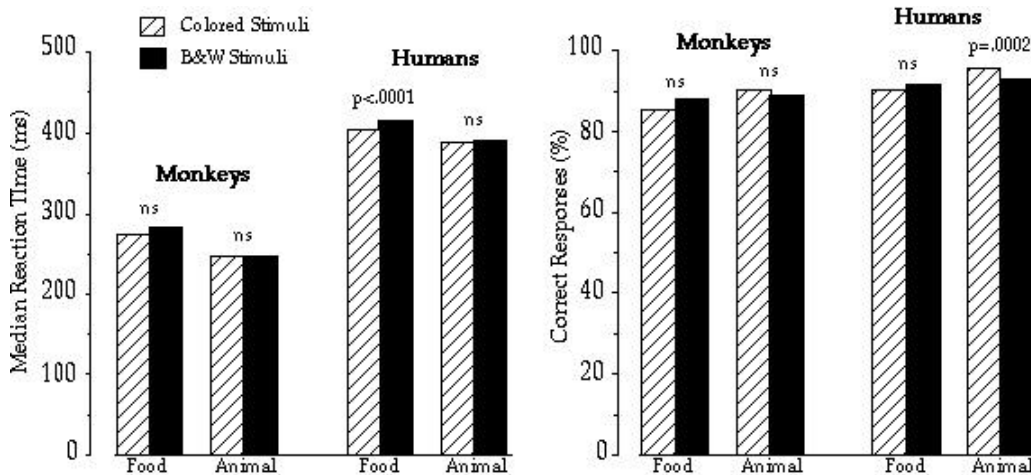


Figure 2 : Performance of the two monkeys and mean performance of each group of 10 Humans with chromatic (hatched bar) and achromatic (black bar) stimuli in the two categorization tasks (food vs. non food and animal vs. non animal). The median reaction time of correct go-responses is shown on the left, the percentage of correct (go and no-go) responses is shown on the right.

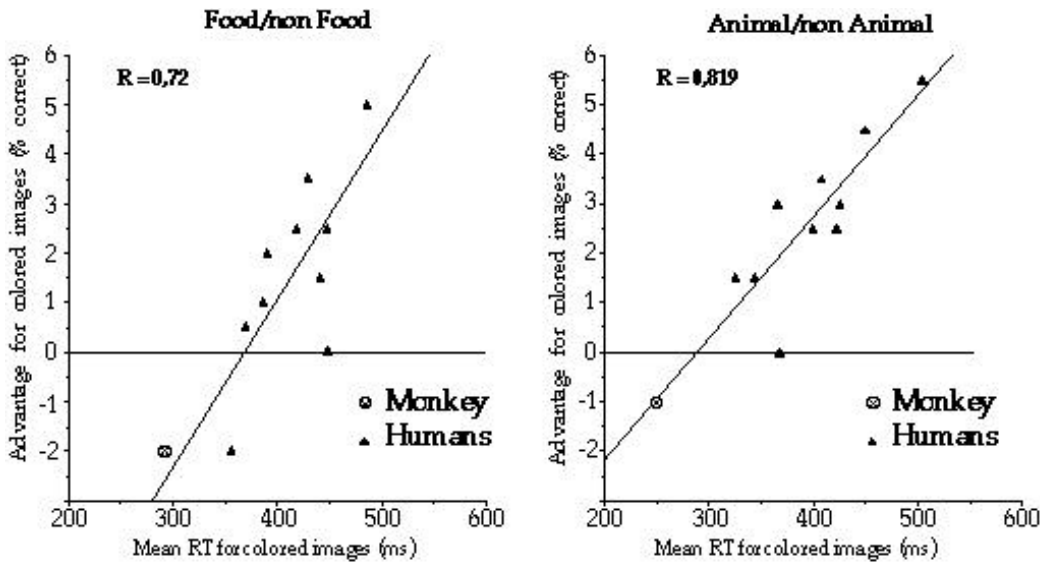


Figure 3: The advantage for color images (expressed in percentage correct) is shown as a function of the mean RT for the 10 human subjects tested on each categorisation task (food vs. non food and animal vs. non animal). The monkey's performance is shown for information but was not included for calculating the correlation coefficient.

4. Discussion

What implications do these findings have for the nature of the neuronal mechanisms underlying rapid visual categorization? The results obtained here show how efficient the primate visual system is and can give some important clues for the development of artificial visual system concerning the relative importance of various visual information in object recognition. They first replicate the results obtained previously [1, 2] : both monkeys and humans are fast and accurate at categorizing natural images that they have never seen before, even without contextual help or without eye movement. Indeed, reaction times are so fast that, at least in this sort of task, it would appear that object recognition cannot depend heavily on time consuming iterative processing.

The other main finding is that removing the color information from the natural images had only mild effects on rapid categorization in both monkeys and humans. The role of color in object recognition is the subject of a large debate. The visual system is very sensitive to color [6]; color has been reported to interact with other information to improve pattern recognition [7, 8] and can even be used early in delayed match to sample tasks using natural images [9]. Furthermore, Mel [10] has developed artificial systems for object recognition that make extensive use of color information. On the other hand, Biederman and Ju [11] found that color did not play a direct role in object recognition except when the object shape is degraded. They argue for a fast access to a coarse structural mental

representation of the object [12].

In our case we would like to suggest that the absence of a color advantage could be related to differences in the temporal dynamics of processing in the magnocellular and parvocellular pathways of the visual system. It has recently been shown that chromatic information in the parvocellular stream reaches visual cortex roughly 20 ms after the magnocellular inputs which are essentially concerned with transmitting luminance based information [5, 13]. If, as has been argued, this sort of rapid visual categorization depends on a feed-forward sequence of processing in which only the first 10 or so ms of activity in each cortical area can be used, it is clear that the first wave of processing in the visual system would be necessarily color-blind. It follows that color would only be important for images that could not be accurately categorized on the basis of the first wave of processing, or in situations where the final decision is not made very rapidly. This is consistent with the fact that, in our study, the humans that had the largest color advantage were those which had the longest mean reaction times.

Such data thus appears consistent with the hypothesis that even complex form categorization tasks can be performed on the basis of largely feed-forward processing mechanisms. Further support for this idea comes from simulation studies using SpikeNET which have shown that it is indeed possible to perform tasks like localizing faces in natural images on the basis of relatively simple feedforward architectures [14].

References

- [1] S. J. Thorpe, D. Fize, C. Marlot, Speed of processing in the human visual system, *Nature* 381, (1996) 520- 522.
- [2] M. Fabre-Thorpe, G. Richard, S. J. Thorpe, Rapid categorization of natural images by rhesus monkeys, *Neuroreport* 9, (1998) 303-308.
- [3] S. J. Thorpe, M. Imbert, Biological constraints on connectionist models., in: R. Pfeifer, Z. Schreter, F. Fogelman-Soulié, L. Steels, Eds, *Connectionism in Perspective.*, (Elsevier, Amsterdam, 1989) 63-92.
- [4] S. J. Thorpe, Spike arrival times : a highly efficient coding scheme for neural networks, in: R. Eckmiller, G. Hartman, G. Hauske, Eds, *Parallel processing in neural systems*, (Elsevier, Amsterdam, 1990).
- [5] L. G. Nowak, M. H. J. Munk, P. Girard, J. Bullier, Visual latencies in areas V1 and V2 of the macaque monkey, *Visual Neuroscience* 12, (1995) 371-384.
- [6] A. Chaparro, C. F. Stromeyer, E. P. Huang, R. E. Kronauer, R. T. Eskew Jr, Colour is what the eye sees best, *Nature* 361, (1993) 348-350.
- [7] G. Syrkin, M. Gur, Colour and luminance interact to improve pattern recognition, *Perception* 26, (1997) 127-140.
- [8] L. H. Wurm, G. E. Legge, L. M. Isenberg, A. Luebker, Color improve object recognition in normal and low vision, *J. Exp. Psychol. (Human perception and performance)* 19, (1993) 899-911 .
- [9] B. S. Tjan, A. I. Ruppertsberg, H. H. Bülthoff, Early use of configural information in rapid scene perception, *Perception* 27 (suppl), (1998) 153.
- [10] B. W. Mel, SEEMORE : Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition, *Neural Computation* 9, (1997) 777-804.
- [11] I. Biederman, G. Ju, Surface versus edge-based determinants of visual recognition, *Cognitive Psychology* 20, (1988) 38-64.
- [12] I. Biederman, Recognition-by-components: A theory of human image understanding, *Psychological Review* 94, (1987) 115-147.
- [13] L. G. Nowak, J. Bullier, The timing of information transfer in the visual system, in: J. Kaas, K. Rocklund, A. Peters, Eds, *Extrastriate cortex in primates*, (Plenum Press, New-York, 1997) 205-241.
- [14] R. VanRullen, J. Gautrais, A. Delorme, S. J. Thorpe, Face detection using one spike per neurone, *Biosystems*, (1998) (in press).