

## On the speed of natural scene categorisation in human and non-human primates

Michèle Fabre-Thorpe, Ghislaine Richard,  
and Simon J. Thorpe

*Centre de Recherche Cerveau et Cognition, Toulouse, France*

**Abstract.** Two rhesus monkeys were tested in a visual categorisation task in which they had to respond to the presence of a given target (food for one monkey, animals for the other). All stimuli were natural images taken from a vast CD ROM data bank that were flashed for only 80 ms (to avoid eye movements). Every day, among the images that they had already experienced, the macaques were shown 5 new targets and 5 new distractors; their responses to the first presentation of each new image were analysed. Monkeys categorised with a high rate of success (90.5% correct for the food task and 84% for the animal task), and with very fast reaction times (median RTs were 344 ms and 245 ms, respectively). A group of 10 humans that categorised the same images in similar conditions responded somewhat more accurately (95% correct in both tasks) but much more slowly (median RT on the food and animal tasks: 446 ms and 422 ms). Moreover, the images incorrectly classified by humans and monkeys overlapped considerably. A last experimental study demonstrated that the macaques' performance was equally good when stimuli were proposed in black and white. These results show that monkeys can make extremely fast and reliable classifications of previously unseen stimuli, even without color, probably on the basis of abstract representa-

---

Correspondence should be sent to Michèle Fabre-Thorpe, Centre de Recherche Cerveau et Cognition, UMR 5549, Université Paul Sabatier, Faculté de Médecine de Rangueil, 133 route de Narbonne, 31062 Toulouse cedex, France (e-mail: mft@cerco.ups-tlse.fr).

tions that overlap with those of humans. Moreover, the fast responses recorded show that such rapid categorisation cannot rely on time consuming visual processing and may well involve a single feed-forward pass through the visual system.

**Key words:** Primates, visual categorisation, visual processing, object recognition.

---

## INTRODUCTION

In the monkey, the anatomy and physiology of the visual system is well documented. The existence of two main streams of processing in the extrastriate cortex is now widely accepted, with a ventral stream primarily involved in the analysis of object features and a dorsal stream more involved in the perception of space and the spatial relationships between objects (Ungerleider & Mishkin, 1982). Much is known about the interconnection of the numerous visual cortical areas (Felleman & Van Essen, 1991). Moreover, neurophysiological data have shown that high order cortical areas along the ventral stream such as the superior temporal sulcus contain neurons that are selectively activated by the presentation of complex stimuli such as specific objects (Logothetis et al., 1995) or even faces with latencies as short as 80-100 ms (Perrett et al., 1982). In contrast, relatively few authors have looked at the performance of the monkey's visual system in sophisticated visual tasks such as object recognition and categorisation. The few studies that do exist have shown that monkeys can indeed learn to categorise a large range of stimuli, but they leave unclear the mechanisms used to perform their classification. Some authors have argued that only language-trained apes can develop abstract representations, and that other monkeys probably rely on the recognition of small sets of relevant features (Premack, 1983). In fact, since a very early study by Herrnstein and Loveland (1964), a controversy on the nature of the concepts that animals can acquire and their similarity to those of humans has developed (Premack, 1983; Roberts & Mazmanian, 1988; Vaughan, 1988; Vaughan & Greene, 1984). Among the relevant features that could play an essential role in the monkeys' decisions, the colour cues may be preponderant. For example, in the d'Amato and Van Sant experiment (D'Amato & Van Sant, 1988) capuchin monkeys may have decided whether a photo-

graph contained a human being on the basis of relatively simple features such as the presence of a red patch. On the other hand, the categorisation performance of pigeons with black and white photographs has been shown to be either as accurate as their performance with chromatic stimuli (Herrnstein & Loveland, 1964), or reduced to chance level (Watanabe, 1997). However, the role played by colour cues is still unclear, as these results, obtained in a bird whose vision is pentachromatic, cannot be extended to old world monkeys which have trichromatic vision close to that of humans (Dulai et al., 1994).

In the present study, we investigated the performance of two monkeys on a go/no-go visual categorisation task using natural photographs flashed for just 80 ms on a video monitor. One of the monkeys was trained to respond to the presence of food and the other to the presence of an animal. More specifically we tried to tackle three questions: (1) How fast and accurate are monkeys in such a task? (2) How close are they to human performance? (3) How much do they rely on colour cues to perform the task?

## MATERIAL AND METHODS

Two male rhesus monkeys (aged 4 and 2 and a half years) were trained on a task very similar to the one used by Thorpe et al. (Thorpe et al., 1996) with humans. Monkeys were shown a series of photographs on a tactile screen. In order to start the sequence of stimuli, the animal had to put one hand on a button located below the screen at waist level. The pictures were flashed in the centre of the screen for only 80 ms and the monkey had to perform a go/no-go visual categorisation task. When the image belonged to a given target category, the monkey had to quickly release the button and touch the screen (in less than 1 s); otherwise it had to keep its hand on the button. Any correct go or no-go decision was rewarded with fruit juice, whereas incorrect decisions were punished by showing the image again for 3 s. The demands of this task on the visual processing abilities of the monkey are particularly severe. It forces the animal to make a series of rapid decisions on the basis of very brief stimulus presentations. The 80 ms duration used was short enough to prevent the monkeys from making any foveating eye movements. This had the added advantage of encouraging the animals to make behavioural responses that were as rapid as possible - allowing

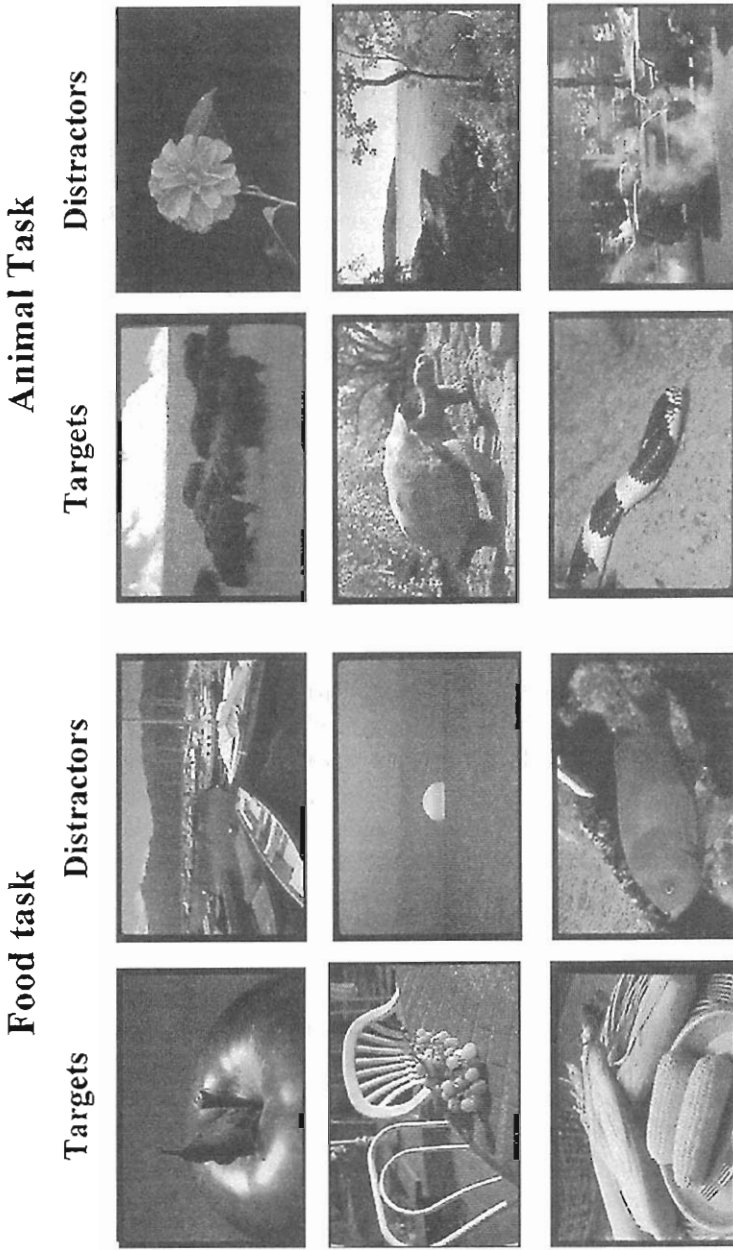


Figure 1. Examples of natural stimuli used, in colour in the food/non-food categorisation task and in the animal/non-animal categorisation task. To illustrate the diversity of the stimuli used in the task, three targets on which both the monkey and all 10 humans responded correctly are shown for each task; three distractors are also shown. On the top two, both the monkey and the 10 humans responded correctly; on the bottom, the monkey and at least 5 of the 10 humans responded incorrectly.

them additional time to explore the images would certainly have resulted in substantially longer reaction times. Each of the two rhesus monkeys worked on a different categorisation task: food vs. non-food for one monkey, and animal vs. non-animal for the other one. All of the pictures were natural scenes (Figure 1) taken from a vast CD ROM data bank. Some additional photographs (roughly 10%) were added for the food vs. non-food categorisation task in order to allow for further controls. Part of these results have been published elsewhere, with other examples of the new targets and distractors used (Fabre-Thorpe et al., 1998). The targets in the food vs. non-food task included fruit, vegetables, salads, cakes, biscuits, and sweets, presented against natural backgrounds. The targets in the animal vs. non-animal task included fish, birds, mammals, and reptiles, again presented in their natural environments. The distractors included landscapes, trees, flowers, monuments, cars, and the target category of the other monkey. The monkeys had no a priori knowledge concerning the position, size, or number of targets in a picture. Moreover, both tasks included targets that were only partly visible. Thus the slides were so varied that contextual help can also be legitimately ruled out.

Both monkeys were naive and had no previous experience with colour photographs or go/no-go procedures. Training was progressive, starting with just 10 targets, and gradually adding distractors and new images over a period of several weeks until both monkeys could work daily with 100 different images (50% distractors and 50% targets), 10 of which had never been seen before (5 distractors and 5 targets). The ability to perform such a task during the initial stages of training was probably the result of a fairly simple stimulus-reinforcement association; on the other hand, a relatively large set of new images was added daily to encourage the monkeys to look for a new strategy. The 100 stimuli that were used on any particular day could be seen repeatedly, as long as the monkey wanted to work, but the order in which they appeared was always randomised to avoid sequential learning. Among these 100 photographs were the 10 stimuli that had never been seen before by the monkey. These were mixed with 90 familiar images (50 stimuli selected at random from the pool of all photographs already experienced by the animal, the remaining 40 being the new stimuli introduced on the four preceding days). The monkeys' categorisation accuracy was assessed by analysing their responses to the first presentation of each new image. Performance on these new images was closely monitored and training

was continued until there was no further improvement in accuracy. At this point, performance was analysed in terms of both accuracy and speed for 200 new stimuli presented on 20 subsequent sessions. Note that in addition to the new views of known objects in different contexts, the new stimuli also included food objects (e.g., walnuts, pineapple, cabbage) and animal objects (e.g., crocodile, tortoise, shark) that the monkeys had never experienced in the course of the training.

Performance was compared to that of 10 humans working under similar conditions on the same material. Finally, we also evaluated how the categorisation performance of the monkeys was affected by chromatic and achromatic stimuli, both on familiar images seen for the first time in black and white and on a pool of 400 new images (200 in color and 200 in black and white).

## RESULTS

In less than three months, both monkeys could generalise to photographs that they had never seen before. They had experienced 280 stimuli for the food task and 340 for the animal task when we started collecting the data reported below.

### **How fast and accurate are monkeys in such a task?**

Given the severe challenge that such a task represents for the visual system, the monkeys' performance on the very first presentation of the 200 new stimuli was remarkably good. First in terms of accuracy, the percentage of correct responses reached 90.5% on the food task and 84% on the animal task (Figure 2A). Secondly in terms of speed (Figure 3), the monkeys' median reaction time (RT) distributions for correct go-responses were 344 ms and 245 ms for the food and animal tasks, respectively (mean reaction times of 356 ms and 251 ms). The RT measured the delay between stimulus onset and movement onset (when the hand broke contact with the button) and was computed for each go response, whether correct or incorrect.

### How close is the performance of monkeys and humans?

Human performance has rarely been tested under the same conditions and with the same material as those used with monkeys (Herrnstein et al., 1976; Roberts & Mazmanian, 1988; Schrier & Brady, 1987). In this study, 10 human subjects were tested in conditions that were essentially the same as those used with the monkeys. They were tested on the same two sets of 200 new photographs (with the exception that, for humans, the new stimuli were not mixed with familiar ones), with the same negative feedback (the incorrectly classified image was shown again for 3 s); the positive feedback for correct responses was a tone for humans whereas the monkeys had both a tone and a fruit juice reward.

Humans were significantly more accurate ( $\chi^2$  for correct vs. incorrect responses:  $p < .004$  for the food task and  $p < .0001$  for the animal task) than monkeys, averaging 95% correct on both tasks (Figure

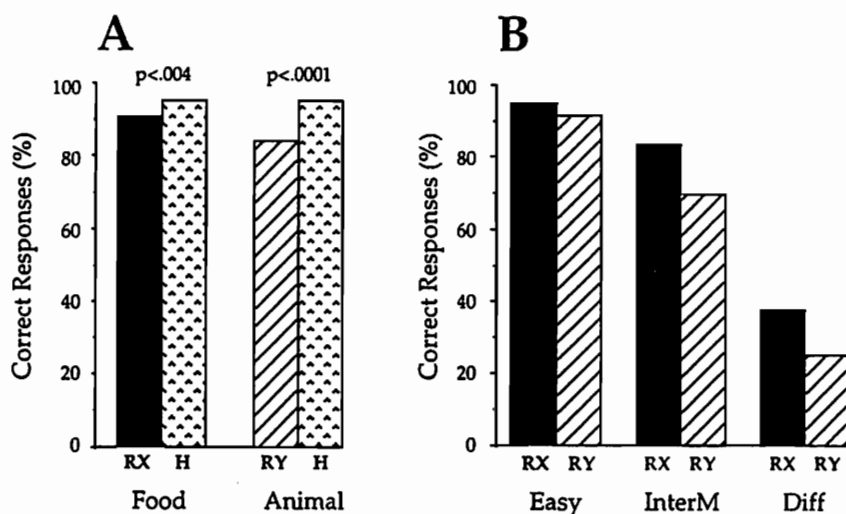


Figure 2. Comparison of human and monkey accuracy on the categorisation tasks. A: percentage of correct responses is shown for each monkey (RX and RY) and for the group of humans (H). Solid bars: monkey working on the food/non-food categorisation task; hatched: monkey working on the animal/non-animal categorisation task; stippled: mean performance of the corresponding group of 10 humans. B: accuracy achieved by each monkey (RX and RY) for the three different groups of images (easy, intermediate, and difficult) classified on the basis of human performance.

2A), with a range of 92.5% to 98.5% on the food task and 93% to 99.5% on the animal task. We were specifically interested in comparing the errors made by both humans and monkeys. In fact, the overlap between the monkeys' and the humans' errors was substantial: 11 out of 19 errors for the monkey working on food vs. non food categorisation, and 20 out of 32 for the other monkey were reproduced by at least one and as many as nine humans (Figure 1). The photographs were then classified in three categories on the basis of human performance: easy images that were categorised correctly by all 10 humans (156 on the food task and 140 on the animal task), intermediate images that 1 to 3 human subjects categorised incorrectly (36 on the food task and 46 on the animal task), and difficult images on which 6-9 human subjects made a mistake (8 on the food task and 4 on the animal task). The performance of each monkey was then analysed separately for each pool of images (Figure 2B). Easy images were categorised with a high rate of success by monkeys (94.9% and 91.4% for the food task and the animal task, respectively), the score was significantly lower (83.3% and 69.6%) for intermediate images ( $df = 1$ ,  $\text{Chi}^2 = 5.76$ ,  $p < .02$  and  $15.1$ ,  $p = .0001$ ), and even lower (37.5% and 25%) for the set of difficult images ( $df = 1$ ,  $\text{Chi}^2 = 34.3$ ,  $p < .0001$  and  $18.4$ ,  $p < .0001$ ). These results suggest that the humans and monkeys were using similar categorisation strategies.

We also compared the speed of performance in humans and monkeys. The human subjects were much slower than the monkeys in initiating their movement. Human median reaction times were 446 ms (402-503 ms) for the food task and 422 ms (383-504 ms) for the animal task. Thus, monkeys were 100-180 ms faster than humans (Figure 3).

The results obtained with humans seem to be very consistent, as they are close to those published in a recent study using a very similar task in which humans had to detect the presence of animals in briefly flashed photographs (Thorpe et al., 1996). In that study, they obtained mean correct answer scores of 94% with a median RT of 445 ms.

### **How much do monkeys rely on the processing of colour to perform the task?**

Among the various features of an object, colour may be an important one for identifying food objects or animals. To study how much the two monkeys relied on colour cues, they were trained to perform the catego-



risation task on 200 very familiar photographs (that they had already seen many times) randomly presented either in colour or in black and white. When they were used to the task, we studied their performance on another pool of 200 familiar pictures. Among these, 100 were presented first in colour then in black and white (condition C-BW), whereas the other 100 were presented first in black and white and then in colour (condition BW-C). Statistical analyses were performed to compare the performance on the first presentation in black and white with the first presentation in colour of the same daily session (Figure 4). The monkey working on the animal vs. non-animal task was also tested on a true categorisation task, i.e., with 400 photographs never seen

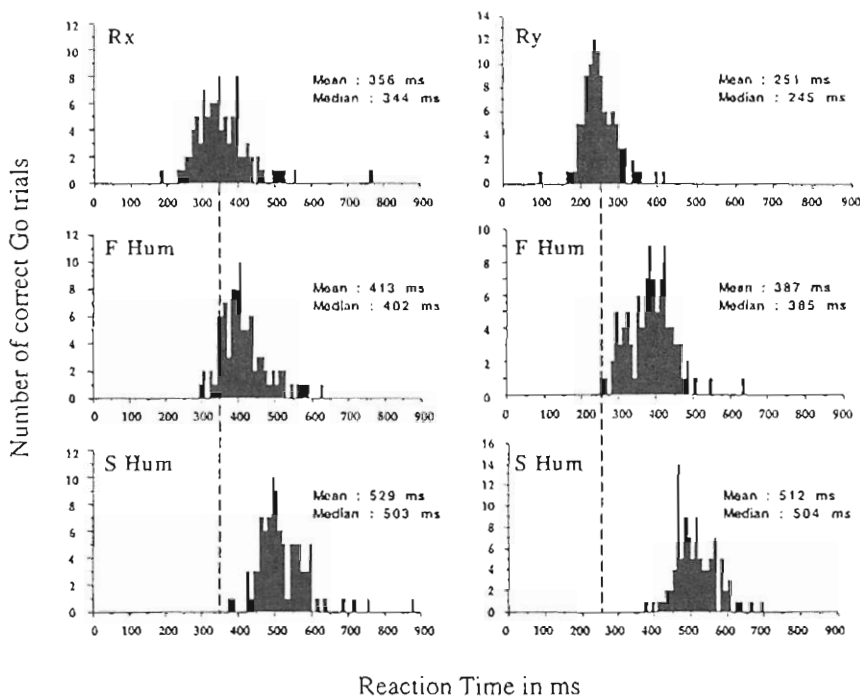


Figure 3. Comparison of human and monkey speed on the categorisation tasks. On the left, food task; on the right, animal task. The bargraphs of the reaction times for correct go responses are shown for the two monkeys (RX and RY), and for the fastest (F Hum) and slowest (S Hum) humans in each group of 10 subjects. Mean and median reaction times are indicated for each bargraph.

before. The methods were as described in the material and methods section, except that instead of showing 10 new photographs in every session (mixed among 90 familiar ones), 20 new stimuli were proposed every day to the monkey, of which 10 (5 targets and 5 distractors) were in colour and 10 were in black and white (mixed among 80 familiar ones). Performance on the 200 new chromatic stimuli and the 200 new

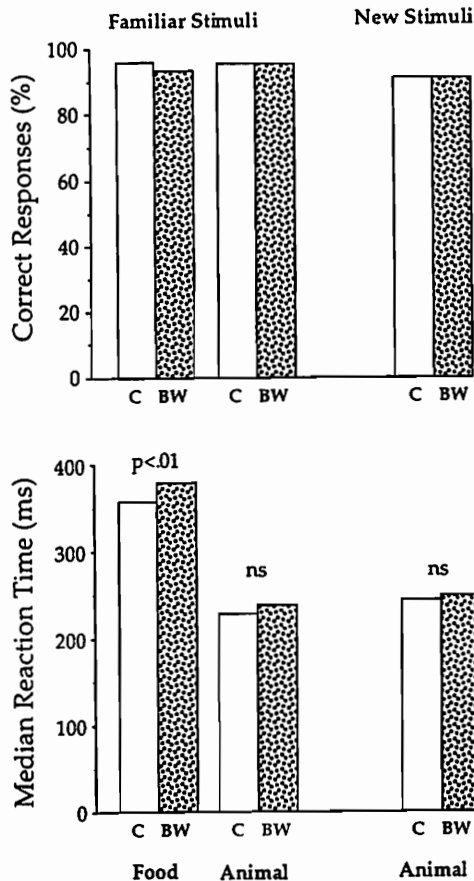


Figure 4. Performance of the two monkeys working with black and white images. Top: mean percentage of correct responses. Bottom: median reaction time in ms. Performance on coloured stimuli (C) is shown by the empty columns, and on black and white stimuli (BW), by the stippled pattern. Data for familiar stimuli are shown on the left for both the food task and the animal task. Data for new stimuli are shown on the right for the animal task.

achromatic stimuli was compared (Figure 4). There were no differences between the C-BW and the BW-C conditions. Thus we will only report the results for black and white (BW) versus coloured (Col) stimuli.

Monkeys were as accurate at categorising familiar achromatic and chromatic stimuli (93.5% vs. 96% correct for the food task and 95.5% correct in both conditions for the animal task). Concerning the latency of movement onset, there was an increase that was statistically significant for the monkey working on the food task: the medians of the RT distributions for correct go-responses were 358 ms for Col stimuli and 379 ms for BW stimuli (means 368 ms and 393 ms, two-tailed *t*-test, *df* = 189, *t* = -2.6, *p* < .01). This increase was almost significant for the other monkey: the medians of the RT distributions for correct go-responses were 228 ms for Col stimuli and 238 ms for BW stimuli (means 230 ms and 243 ms, two-tailed *t*-test, *df* = 197, *t* = -1.86, *p* < .065). The very slight effects induced by removing the colour information in the image imply that this sort of rapid categorisation task does not depend heavily on colour cues, since the task can be done remarkably well with BW stimuli. This conclusion was further reinforced by the data obtained for the categorisation of new images by the monkey working on the animal task. Accuracy on classifying the 400 new stimuli was 91% in both conditions, but there was no reaction time difference between the two conditions (two-tailed *t*-test, *df* = 183, *t* = -1.59, ns). The median RT value was: 246 ms for Col stimuli and 249 ms for BW stimuli; the mean RTs were 250 ms and 261 ms, respectively.

## DISCUSSION

The data showed that monkeys can quickly and accurately classify previously unseen natural images on the basis of a single brief presentation, and that colour cues are not necessary for this performance. Moreover the photos that humans find difficult overlap considerably with the ones that caused problems for the monkeys. These data have implications both for visual information processing in object recognition, and for the strategy that underlies classification in monkeys.

Although one of the two monkeys was much slower, the fast behavioural reaction times recorded in the fastest animal show that the time between the visual input and the motor output can be very short, sometimes as short as 220 ms (Figure 3). In this group of very fast re-

sponses, correct go-responses statistically outnumbered incorrect ones, whereas targets and distractors were otherwise equiprobable (Fabre-Thorpe et al., 1998), showing that the visual analysis of the stimuli can be completed. This RT value includes not only the time required for visual processing but also a substantial motor component. If this motor component is estimated at around 100 ms, visual processing would have to be completed in 120 ms. This fits with the onset latencies of high-order visual neurons in the monkey temporal lobe, which are typically in the range of 80-140 ms (Perrett et al., 1982), and with a recent study reporting that in humans, visual processing in categorisation tasks could be done in under 150 ms (Thorpe et al., 1996). With so little time, the visual processing required for this task must be essentially feed-forward and massively parallel. This implies that, at least in the sorts of tasks used here, object recognition cannot depend heavily on time-consuming iterative processing. The faster response recorded for monkeys compared to humans could be explained by a simple speed-accuracy trade-off, in that the decreased reaction time in monkeys was associated with a decline in accuracy. It could also be the result of the smaller physical dimensions of the brain. Recent data suggests that the conduction velocities of intracortical connections may be as slow as 1 m/s or less (Nowak & Bullier, 1997). It could therefore be that monkeys are faster simply because the conduction delays are shorter than in humans.

The comparison between monkeys and humans also shows that the images that were difficult for the humans to classify were also those on which monkeys performed badly. This suggests that monkeys and humans made decisions on the basis of similar cues. In our experiment we used briefly flashed stimuli, whereas all previous studies have used long presentation times. It may be that such brief presentations, which rule out the use of exploratory eye movements, encourage fast decisions and emphasize the similarities between humans and monkeys. In any case, it seems difficult to escape the conclusion that monkeys can generalize a rule learned during training, probably on the basis of abstract representations that overlap with those of humans. The fact that the absence of colour cues had no effect on performance is also in agreement with this interpretation. If monkeys were using a conjunction of relevant features to make their decision, colour could be one of the preponderant ones, as suggested previously (D'Amato & Van Sant, 1988), and black and white stimuli should have been more difficult to classify. The fact that one monkey could categorize new black and white stimuli mixed with

coloured ones without any obvious impairment implies that its decisions did not require the use of colour cues, even when the task was designed so that the monkey had no information on the attributes of the next stimulus and could not anticipate the availability of colour cues. However, one cannot rule out the possibility that some other low-level visual features such as spatial frequency and orientation composition play an important part in the categorisation of such images. Nevertheless, it seems unlikely that this explanation could account for all of the present results. Firstly, it would be impossible to explain how the monkeys could respond with a high rate of success, given that the targets in the animal task were also used as distractors in the food task and vice versa. If the animals were simply responding to some simple features, then it clearly could not have been the same features in the two tasks. Secondly, if combinations of low-level features are used, there would have to be a large number of different sets in order to account for the accurate classification of the wide range of stimuli used. Finally, the results reported here suggest that this form of rapid visual categorisation is remarkably similar in humans and monkeys, since monkeys performed nearly as well as humans and showed a similar pattern of errors; the implication would be that humans also perform their visual categorisation on the basis of similar combinations of features. Although the precise nature of these processes is still unclear, the high success rate achieved on the task by the two macaques seems to show that this form of rapid categorisation does not require the verbal and language skills that are normally associated with the categorisation performance of humans and language trained apes.

### RÉSUMÉ

Deux singes rhésus ont effectué une tâche de catégorisation visuelle dans laquelle ils devaient répondre à la présence d'une cible donnée (aliment pour l'un, animal pour l'autre). Chaque stimulus était une image naturelle provenant d'une collection disponible sur CD ROM et n'était présenté que pendant 80 ms afin d'éviter toute exploration oculaire. Chaque jour, parmi des images déjà vues, 5 nouvelles cibles et 5 nouveaux distracteurs étaient proposés. La réponse à la première présentation de chaque nouvelle image était analysée. Le taux de réussite

des deux singes sur 200 nouvelles images a été élevé (90.5% pour la catégorisation "aliment" et 84% pour la catégorisation "animal"). De plus, les temps de réaction des réponses étaient très brefs (respectivement 344 ms and 245 ms de TR médian). La catégorisation des mêmes images a été effectuée dans des conditions semblables par 10 sujets humains qui se sont révélés plus précis (95% de réussite dans chaque tâche) mais considérablement plus lents (TR médian : 446 ms et 422 ms pour les classifications "aliment" et "animal"). De plus, les images induisant des erreurs de classification étaient souvent communes aux sujets humains et aux singes. Une dernière série expérimentale a montré que les performances des singes étaient similaires lorsque les images étaient proposées en noir et blanc. Les macaques peuvent effectuer des catégorisations rapides et précises d'images qu'ils n'ont jamais vues auparavant, même en absence de couleur, probablement sur la base de représentations similaires à celles de l'homme. Leur rapidité montre que le traitement visuel requis doit être massivement parallèle et principalement feed-forward.

## REFERENCES

- D'Amato, M. R., & Van Sant, P. (1988). The person concept in monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behavior*, *14*, 43-55.
- Dulai, K. S., Bowmaker, J. K., Mollon, J. D., & Hunt, D. M. (1994). Sequence divergence, polymorphism and evolution of the middle-wave and long-wave visual pigment genes of great apes and old world monkeys. *Vision Research*, *34*, 2483-2491.
- Fabre-Thorpe, M., Richard, G., & Thorpe, S. (1998). Rapid categorization of natural images by rhesus monkeys. *Neuroreport*, *9*, 2, 303-308.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*, 1-47.
- Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concept in the pigeon. *Science*, *146*, 549-551.
- Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior*, *2*, 285-302.
- Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, *5*, 552-563.
- Nowak, L. G., & Bullier J. (1997). The timing of information transfer in the visual system. In J. Kaas, K. Rocklund, & A. Peters (Eds.), *Extrastriate cortex in primates* (pp. 205-241). New York: Plenum Press.

- Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, *47*, 329-342.
- Premack, D. (1983). The codes of man and beasts. *The Behavioral and Brain Sciences*, *6*, 125-167.
- Roberts, W. A., & Mazmanian, D. S. (1988). Concept learning at different levels of abstraction by pigeons, monkeys, and people. *Journal of Experimental Psychology: Animal Behavior*, *14*, 247-260.
- Schrier, A. M., & Brady, P. M. (1987). Categorization of natural stimuli by monkeys (*Macaca mulatta*): effects of stimulus set size and modification of exemplars. *Journal of Experimental Psychology: Animal Behavior*, *13*, 136-143.
- Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520-522.
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behaviour* (pp. 549-585). Cambridge, MA: MIT Press.
- Vaughan, W. J. (1988). Formation of equivalence sets in pigeons. *Journal of Experimental Psychology: Animal Behavior*, *14*, 36-42.
- Vaughan, W. J., & Greene, S. L. (1984). Pigeon visual memory capacity. *Journal of Experimental Psychology: Animal Behavior*, *10*, 256-271.
- Watanabe, S. (1997). Visual discrimination of real objects and pictures in pigeons. *Animal Learning Behavior*, *25*, 185-192.

Received 2 October, 1997

Accepted January 31, 1998