Website publication 16 January 1998

NeuroReport 9, 303–308 (1998)

Rapid categorization of natural images by rhesus monkeys

Michèle Fabre-Thorpe,[CA] Ghislaine Richard and Simon J. Thorpe

Centre de Recherche Cerveau et Cognition (UMR 5549), Faculté de Médecine de Rangueil, 133 route de Narbonne, 31062 Toulouse, France

[CA]Corresponding Author

Two rhesus macaques were tested on a categorization task in which they had to classify previously unseen photographs flashed for only 80 ms. One monkey was trained to respond to the presence of an animal, the second to the presence of food. Although the monkeys were not quite as accurate as humans tested on the same material, they nevertheless performed this very challenging visual task remarkably well. Furthermore, their reaction times were considerably shorter than even the fastest human subject. Such data, combined with the detailed knowledge of the monkey's visual system, provide a severe challenge to current theories of visual processing. They also argue that this form of rapid visual categorization is fundamentally similar in both monkeys and humans.

Key words: Categorization; Feed-forward processing; Object recognition; Rhesus macaque; Visual processing

Introduction

In a task in which human subjects have to decide whether a previously unseen natural image contains an animal, event-related potentials recorded on target and non-target trials showed a clear difference as early as 150 ms after stimulus onset.[1] It has been argued that such rapid processing is only really compatible with feed-forward processing,[2] but the strength of this argument is limited by the relative sparsity of anatomical and physiological data concerning the organization of the human visual system. In contrast, much is known about visual cortical areas and their connections in non-human primates, and in particular in rhesus macaques,[3,4] but few studies have looked at their ability to perform sophisticated visual categorization tasks. Moreover, since Herrnstein and Loveland's pioneering study of visual concept learning in pigeons,[5] the question of whether animals can perform categorical judgements in a similar way to humans has been the subject of lengthy debate.[6–9] Although some experiments have provided evidence that monkeys can learn to categorize a large range of natural stimuli,[7,10,11] other authors have argued that only language-trained animals develop abstract categories, and that other monkeys probably rely on a more image-based code or on the recognition of a small set of relevant features.[9,12]

In the present experiment, we tested the performance of two rhesus monkeys trained on a visual categorization task very similar to the one used by Thorpe *et al.* with humans,[1] and which makes particularly severe demands on the visual system. Their results were compared with the accuracy and speed of humans tested under similar conditions using the same material.

Materials and Methods

Two rhesus monkeys had to perform a visual categorization task in which they had to classify a succession of photographs that appeared in the centre of a tactile screen. Trials started when the animal had positioned one hand on a button located just below the screen at waist level. Pictures were flashed for only 80 ms and, when the picture belonged to the target category, the monkey had 1 s to release the button and touch the middle of the screen ('go' response); if not, it had to keep pressing the button ('no go' response). All correct decisions were rewarded by fruit juice, whereas incorrect decisions were punished by showing the image again for 3 s. This approach is original because it forces the animal to make rapid decisions on the basis of very brief stimulus presentations, with no time for exploratory eye movements.

Each monkey was trained on a different categorization task: food *vs* non-food for one monkey, and animal *vs* non-animal for the other. Targets in the food task included photographs of fruit, vegetables, salads, cakes, biscuits and 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 both tasks included images where they were only partly visible (see examples in Fig. 1). Moreover, the monkeys had essentially no a priori knowledge concerning the position, the size or the number of targets in a picture.

Training was progressive, starting with just 10 images, and gradually adding more images over a period of several weeks until both monkeys were

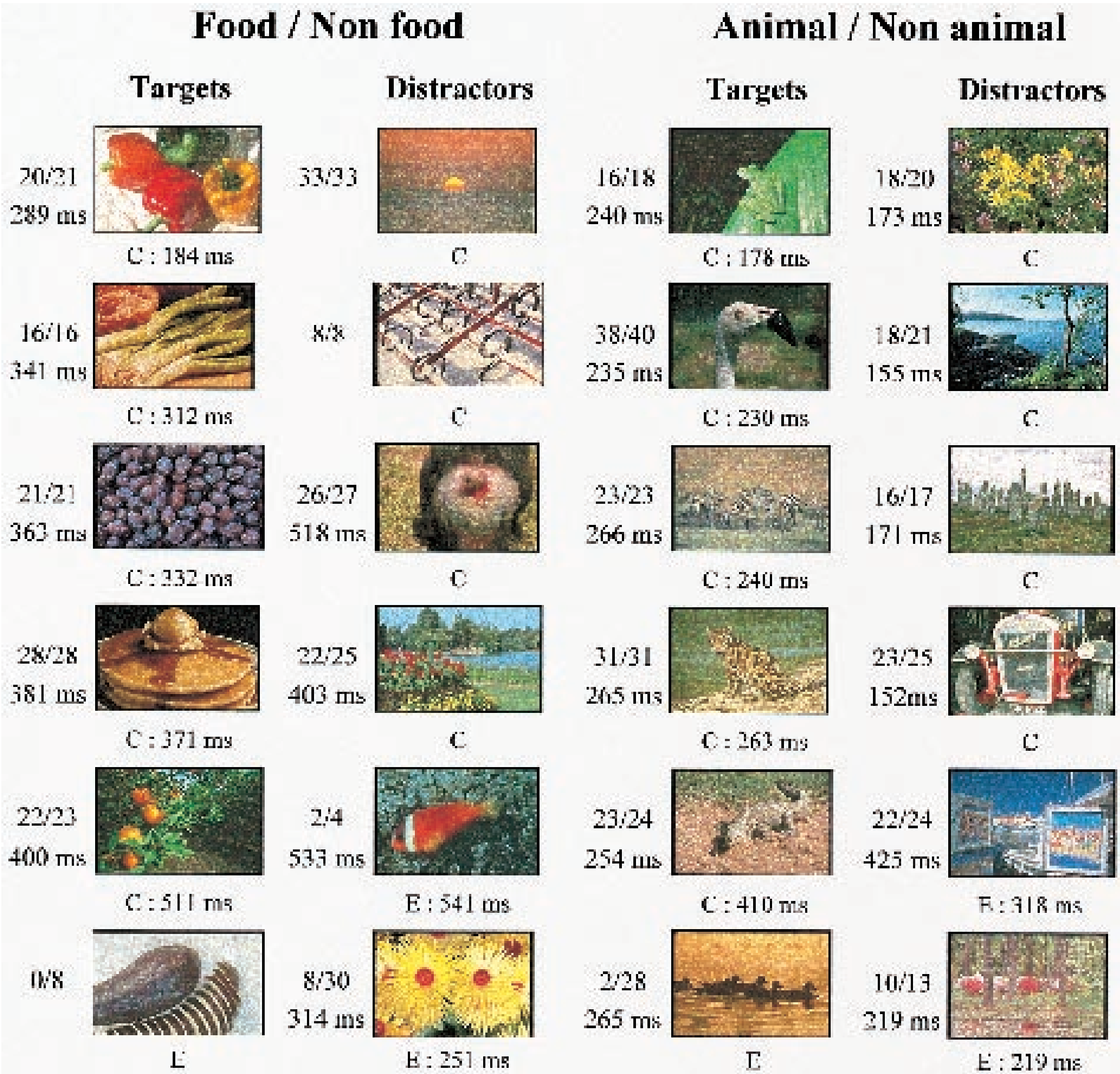

FIG. 1. Examples of the stimuli used in the two categorization tasks. The images (192 × 128 pixels, 8-bit colour) were presented for five frames at a refresh rate of 62 Hz (non-interlaced), corresponding to a presentation time of 80 ms, using a programmable graphics board (VSG 2, Cambridge Research Systems) mounted in a PC-compatible computer. The monkey was typically 25–30 cm from the screen so that the image approximate angular size was 15/25°. Most of the images were oriented horizontally, although some were oriented vertically. All the pictures were natural scenes taken from a large commercial CD-ROM library allowing access to several thousand stimuli, although some additional photographs (roughly 10%) were added for the food *vs* non-food categorization task in order to allow further controls to be performed. Six horizontal targets and distractors are shown for each task. Below each stimulus is indicated the monkey's behaviour to the very first presentation of the image (C for correct, E for error) with the reaction time for go responses. The numbers on the left of each stimulus give the total number of correct responses, the total number of presentations and the mean reaction time when appropriate. The six target examples illustrate the range of reaction times obtained starting with the shortest reaction time at the top. An example of a missed target is shown at the bottom. The six distractor examples also illustrate the range of behavioural responses observed and include pictures for which the monkeys showed a strong tendency to respond as if they were targets (bottom).

performing the categorization task well. The ability to perform such a task could be the result of a fairly simple stimulus-reinforcement association, so to test whether the animals were indeed capable of true categorization, we began a period in which we tested the response of the monkeys to the very first presentation of 200 new images. At this point they were already working with a relatively large set of images (280 in the case of the food task, 340 in the case of the animal task). Each day, they were shown 10 new photographs (five targets and five distractors), mixed with 90 familiar images (50 stimuli selected at random from the pool of all photographs already seen by the monkey, the remaining 40 being the new stimuli introduced on the 4 preceding days). The 100 stimuli were used repeatedly during a daily session but always in a different order to avoid any sequential learning.

Results

Despite the severe demands made on the visual system by such a task, the monkeys' performance on the very first presentation of the 200 new stimuli was remarkably good. Accuracy was 90.5% correct on the food task and 84% on the animal task despite the fact that both monkeys were biased towards making go responses (respectively 94% and 99% correct on targets and 87% and 69% correct on distractors). Furthermore, the monkeys were outstandingly fast in performing on go trials (Fig. 2) with mean reaction times of 356 ms and 251 ms respectively for the food and the animal tasks (median values 344 ms and 245 ms). Some of this rapidity may involve a speed-accuracy trade off since the faster monkey also showed the greater bias towards making go responses. However, even early behavioural responses in this monkey were clearly biased towards targets demonstrating that such responses were not random. Indeed, if behaviour was random for go responses with very short reaction times, correct go responses should be as common as incorrect go responses as target and non-target stimuli are equiprobable. This is not the case and go responses with reaction times in the range 210–230 ms were much more common on target than on distractor trials (two-tailed $\chi^2 = 6.9$, d.f. = 1, $p < 0.04$) arguing strongly that at least on some trials, the entire visuo-motor sequence can be completed in little more than 200 ms (Fig. 2).

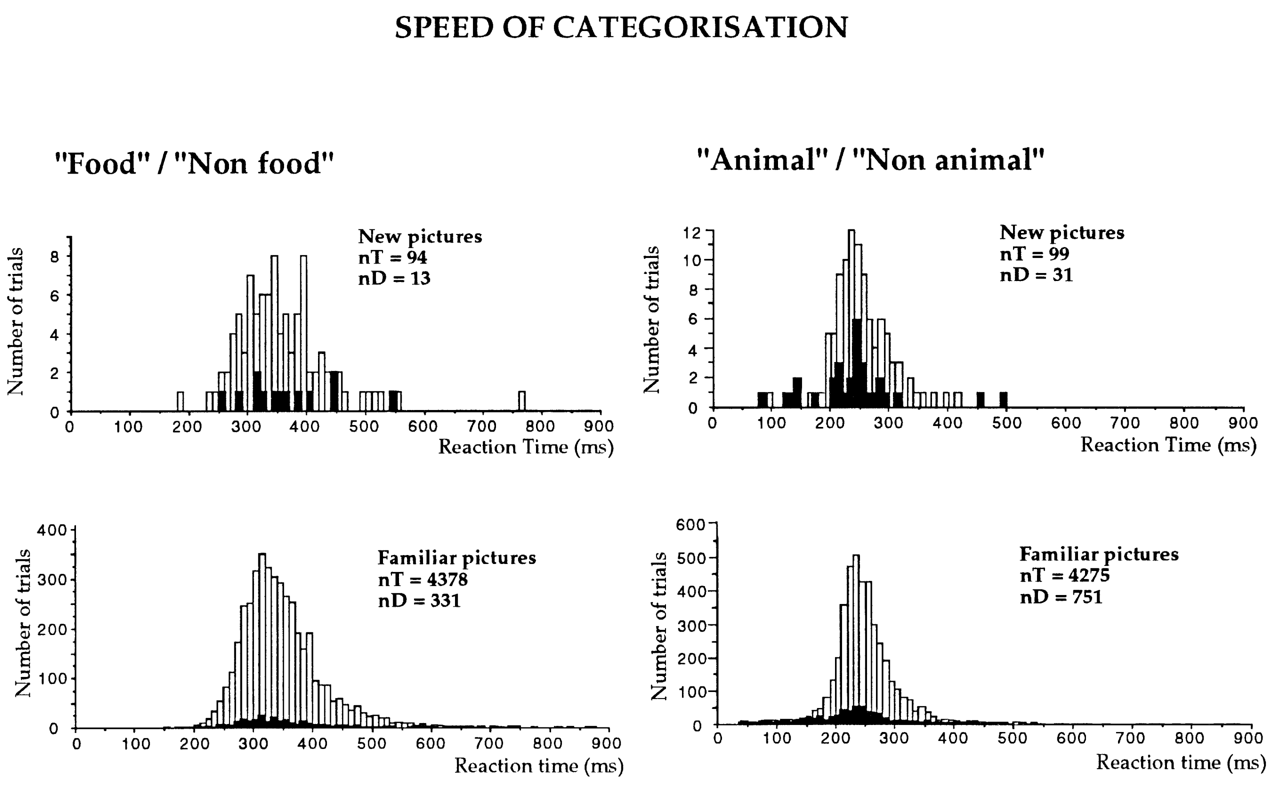

FIG. 2. Behavioural reaction times measured for both monkeys on correct (white) and incorrect (black) go responses performed towards new pictures (top) and familiar pictures (bottom) are shown together with the number of correct go responses towards target (nT) and the number of incorrect go responses towards distractors (nD). A statistical bias towards correct go responses (compared with incorrect go-responses) appears as early as 290–310 ms for the food task and from 210–230 ms for the animal task and shows that the visuo-motor processing can be done in such a very restricted amount of time. This bias was found even earlier, respectively 230–240 ms and 180–190 ms for the categorization of familiar pictures. This may simply be due to the larger number of trials available since no statistical difference was found between the distributions of reaction times to new and familiar pictures.

Table 1. Performance on the two categorization tasks. Performance of the monkeys in the food *vs* non-food and the animal *vs* non-animal tasks. The numbers of correct and incorrect responses are given separately for both novel and familiar targets and distractors. Mean and median reaction times in ms are shown separately for go reponses to both targets and distractors.

	Correct responses	Incorrect responses	Mean RT (ms)	Median RT (ms)
Food/non-food				
New targets	94	6	356	344
New distractors	87	13	367	352
Familiar targets	4378	246	348	336
Familiar distractors	4245	331	371	339
Animal/non-animal				
New targets	99	1	251	245
New distractors	69	31	241	243
Familiar targets	4275	170	249	243
Familiar distractors	3566	751	269	239

The effects of learning on processing time could also be evaluated by comparing the monkey's performance on previously unseen photographs with its performance on the same stimuli when they were familiar (from their third presentation on). Performance was analysed on 4700–5000 trials. Accuracy was higher but the effect was small (93.7% correct for the food task and 89.4% for the animal task) and was essentially due to a reduction of incorrect go responses on distractor trials as no statistically significant improvement was seen for targets (see Table 1). More importantly, there was no reduction in reaction times when the targets had become familiar (Fig. 2). The median reaction time (336 ms in the food task and 243 ms in the animal task) were 8 ms and 2 ms shorter than when new targets were used and the comparison of the two RT distributions showed no statistically significant difference (two tailed t-test, d.f. = 4470, t = 1.02, n.s. for the food task, and d.f. = 4372, t = 0.317, n.s. for the animal task). For the faster monkey, correct go responses significantly outnumbered incorrect go responses from 180 ms onward (two-tailed $\chi^2 = 31.2$, d.f. = 1, $p < 0.0001$ for the period 180–190 ms). As no statistical difference was found when comparing the RT distributions of correct go responses performed toward either new or familiar targets, this even shorter input–output delay could be explained by the very large number of trials that were used when analysing performance with familiar targets. Such results underline just how remarkably efficient the monkeys were at categorizing photographs on the first presentation, and suggest that similar processing mechanisms are used for classifying both familiar and novel stimuli.

The monkeys' ability to categorize accurately novel stimuli effectively ruled out the use of simple stimulus–reward association. This conclusion is reinforced by the fact that certain familiar stimuli were consistently categorized in the wrong class (see the consistent failure to respond to the bottom food target in Fig. 1), despite the fact that such behaviour was punished by a 3 s re-presentation of the image, giving ample opportunity to learn its features.

Further insight was provided by testing 10 human subjects on the same two sets of 200 images in the same conditions except that they were only tested with new photographs. Although humans were more accurate than monkeys (95% correct on both tasks, a value close to the 94% reported earlier[1]) there was substantial overlap between the types of errors that were made (Fig. 3) suggesting that both human and monkeys were using similar categorization strategies. Images that were categorized correctly by all 10 humans (156 in the food task and 140 in the

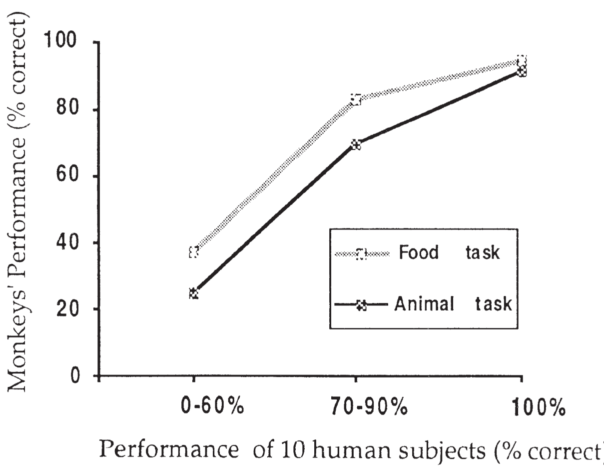

FIG. 3. Comparison of error rate between each monkey and 10 human subjects tested in the same experimental set-up (with the exception that positive reinforcement was given by a tone). The complete set of images used in each categorization task was presented to the 10 subjects. Mean percentage correct, mean and median reaction times were respectively 94.8%, 445 ms and 435 ms for the food task and 94.2%, 437 ms and 425 ms for the animal task. Images were pooled in three classes (from 0–100% correct) according to the number of human subjects that had given a correct answer. The accuracy of the monkey was then computed for each of the three classes of images.

animal task) were categorized with a high rate of success by monkeys (94.9% and 91.4%). This rate was lower (83.3% and 69.6%) for images that 1–3 human subjects categorized incorrectly (36 in the food task and 46 in the animal task) and even lower (37.5% and 25%) for difficult images on which 6–9 human subjects made a mistake (eight in the food task and four in the animal task). Even more strikingly, the monkeys had median reaction times that were 100–180 ms faster than the human subjects both in the present study where median reaction times were 447 ms (401–504 ms) and 422 ms (383–503 ms) for the food and the animal tasks, and in the earlier study[1] (median value 445 ms).

Discussion

One novel feature of the current experiment is the use of briefly flashed stimuli. Virtually all previous studies on categorization have used long presentation times (typically 4–30 s), and it may be that such long presentations allow the use of much more varied strategies including ocular explorations which might emphasize the differences between humans and monkeys. In contrast, the brief presentations used here may encourage subjects to make rapid instinctive decisions on the basis of the first rapid pass through the system. Under these conditions, categorization performance is still surprisingly good and appears remarkably similar in human and non-human primates.

What conclusions can be drawn from the fact that monkeys are faster than humans and can perform such sophisticated visual categorization tasks in as little as 220 ms? Typical latencies for visual responses in higher order visual areas in the monkey temporal lobe are in the range 80–140 ms[13,14] and neurones selectively responding to food have been reported in other structures with somewhat longer latencies.[15,16] However any behavioural reaction time also includes a substantial motor component that corresponds to the initiation and execution of the motor response (probably 80–100 ms or more).[17] Thus, it is clear that the speed with which monkeys can perform the visual categorization tasks leaves little time for anything other than a single feed-forward pass through the visual system. In fact, reaction times observed here are in the same range as those reported in various choice reaction time tasks, even those requiring very limited visual processing.[18] This is surprising, because it implies that the sophisticated visual processing needed to categorize complex natural images does not involve substantial overheads compared with other 'simpler' tasks. Our results are therefore difficult to reconcile with theories of visual processing that rely on sequential attention-based mechanisms such as those proposed for visual search tasks,[19] or which require time-consuming processes such as the large-scale use of feed-back, or the development of oscillatory activity. The implication is that object recognition and visual categorization of complex natural images relies essentially on massively parallel processing, and not on processes in which individual objects in the scene are checked and compared sequentially with internally stored sets of information. This idea is strengthened first by the clear ability of monkeys to generalize not only to new views of known items but also to items they had never seen before, and second by the failure of familiarity to induce a reduction in reaction time.

Overall, these findings imply that rapid categorization of natural images in monkeys must rely, as in humans, on the existence of abstract categorical concepts. In this case, why are monkeys faster than we are? It could be that they use simpler categorical concepts requiring less processing, but the similarity between the types of errors made by monkeys and humans argues against this view. Alternatively, they may use less time consuming verifications before reaching a decision, a result compatible with their somewhat lower accuracy: in other words, their faster responses could reflect a straight speed–accuracy trade-off. Finally, it could be a simple question of the physical dimensions of the brain. Recent data suggest that the conduction velocities of intracortical connections may be considerably slower than previously thought, possibly as slow as 1 m/s or less.[4,20] If so, monkeys could be faster simply because the conduction delays between input and output are shorter.

Conclusion

The present work demonstrates the remarkable efficiency of the primate visual system in processing complex natural scenes. Even with very brief presentations which effectively rule out the use of exploratory eye movements and with a very varied set of photographs to eliminate contextual effects, monkeys can accurately categorize images that they have never seen before. Moreover they can perform such tasks remarkably quickly. In some cases, the delay between the visual input and the onset of the motor output was as short as 180–220 ms. This result shows that the time for completing visual processing in such a demanding task can be very short and challenges theories of object recognition that rely on time consuming sequential processes. Moreover this form of rapid visual categorization seems to involve similar kinds of mechanisms in both human and non-human primates.

References

1. Thorpe SJ, Fize D and Marlot C. *Nature* **381**, 520–522 (1996).
2. Thorpe SJ and Imbert M. Biological constraints on connectionist models. In: Pfeifer R, Schreter Z, Fogelman-Soulié F and Steels L, eds. *Connectionism in Perspective*. Amsterdam: Elsevier, 1989: 63–92.
3. Van Essen DC, Anderson CH and Felleman DJ. *Science* **255**, 419–423 (1992).
4. Nowak LG and Bullier J. The timing of information transfer in the visual system. In: Kaas J, Rocklund K and Peters A, eds. *Extrastriate Cortex in Primates*. New York: Plenum Press, 1997: in press.
5. Herrnstein RJ and Loveland DH. *Science* **146**, 549–551 (1964).
6. Herrnstein RJ. Levels of categorization. In: Edelman GM, Gall WE and Cowan WM, eds. *Signal and Sense: Local and Global Order in Perceptual Maps*. New York: Wiley-Liss, 1990: 385–414.
7. Roberts WA and Mazmanian DS. *J Exp Psychol: Anim Behav Proc* **14**, 247–260 (1988).
8. Vaughan W Jr and Greene SL. *J Exp Psychol: Anim Beh Proc* **10**, 256–271 (1984).
9. Premack D. *Behav Brain Sciences* **6**, 125–167 (1983).
10. Schrier AM and Brady PM. *J Exp Psychol: Anim Behav Proc* **13**, 136–143 (1987).
11. D'Amato MR and Van Sant P. *J Exp Psychol: Anim Behav Proc* **14**, 43–55 (1988).
12. Savage-Rumbaugh ES, Rumbaugh DM, Smith ST and Lawson J. *Science* **210**, 922–925 (1980).
13. Oram MW and Perrett DI. *J Neurophysiol* **68**, 70–84 (1992).
14. Perrett DI, Rolls ET and Caan W. *Exp Brain Res* **47**, 329–342 (1982).
15. Rolls ET, Sanghera MK and Roper Hall A. *Brain Res* **164**, 121–135 (1979).
16. Thorpe SJ, Rolls ET and Maddison SP. *Exp Brain Res* **49**, 93–115 (1983).
17. Kalaska JF and Crammond DJ. *Science* **255**, 1517–1523 (1992).
18. Luce *RD. Response Times: Their Role in Inferring Elementary Operations.* Oxford: Oxford University Press, 1986.
19. Treisman A. Q *J Exp Psychol: Hum Exp Psycho*l **40**, 201–238 (1988).
20. Frégnac Y and Bringuier V. Spatio-temporal dynamics of synaptic integration in cat visual cortical receptive fields. In: Aertsen A and Braitenberg V, eds. *Biological Basis and Computational Theory of Vision*. Amsterdam: Elsevier, 1996: 1–57.

ACKNOWLEDGEMENTS: This work was supported by the Human Capital and Mobility program of the EEC, the CNRS, the GIS 'Sciences de la Cognition' and the Conseil Regional Midi-Pyrénées.

Received 9 September 1997; accepted 20 November 1997