Published in Vision Research, 40(16), 2187-2200, 2000.

Ultra-Rapid Categorisation of natural scenes does not rely on colour cues: A study in monkeys and humans

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

Centre de Recherche Cerveau et Cognition (UMR 5549) Faculté de Médecine de Rangueil. 133, route de Narbonne, 31062 – Toulouse - FRANCE

In a rapid categorisation task, monkeys and humans had to detect a target (animal or food) in briefly flashed, previously unseen, natural images. Removing colour cues had little effect on average performance. In both tasks, accuracy and latency of the fastest behavioural responses were unaffected. Impairments were restricted to a mild accuracy drop (in some of the human subjects) and a small mean RT increase (10-15 ms) observed both in monkeys and humans but only in the detection of food targets. We suggest that rapid categorisation might depend on the feed-forward processing of the early coarse achromatic magnocellular information.

Key words*:* colour, natural scenes, categorisation, primate, visual processing

INTRODUCTION

The recognition of objects and scenes appears effortless and almost instantaneous. The complex processing needed to perform an object categorisation task can be extremely fast in humans (Thorpe, Fize, & Marlot, 1996). Subjects could detect an animal in briefly flashed, previously unseen natural photographs with a high rate of success (94% of correct responses) and very short reaction times (median RT: 445 ms). Frontal ERPs differed sharply on target and non-target trials from 150 ms after stimulus onset. As the underlying visual processing probably involves all the stages along the ventral visual pathway - from retina to the highly integrative infero-temporal cortex - the authors argued that the underlying processing should be essentially feed-forward to be compatible with such time constraint. This constraint on processing speed could be even more severe in monkeys; using the same fast visual categorisation task, rhesus macaques can detect food or animal targets in natural scenes with a correct response rate close to 90% but with much faster behavioural RTs (median RT: 290 ms) than humans (Fabre-Thorpe, Richard, & Thorpe, 1998).

Which image features could be used in generating such fast responses? Although this high level of accuracy cannot be explained by the use of one single image feature, the monkeys could use a combination of different low level cues, and among them colour is an obvious candidate. There is evidence for such a view, for example, to determine whether a photograph contained a human being, Capuchin monkeys may have used the presence of a red patch (D'Amato & Van Sant, 1988). But on the other hand, colour cues were shown not to account for the monkeys' performance in a recent study using fish vs. non fish and tree vs. non-tree categorisations (Vogels, 1999a). The importance of colour could depend on whether or not this cue is a diagnostic feature of the target category (Oliva & Schyns, 2000; Tanaka & Presnell, 1999). For instance, in monkeys and humans that had to select photos of kingfishers – a very brightly coloured bird - among photos of other birds, a drop of performance was observed when the pictures were presented in black and white (BW) (Roberts & Mazmanian, 1988).

Colour differences and colour contours may also be used in image segmentation to provide information about object shape and region of interest within individual objects. However, in human object recognition, the role of colour - particularly in the early visual processes leading to fast identification - is still very controversial. Colour appears to interact with object recognition processing when object naming is required but not in verification tasks in which the object name is presented prior to the object (Ostergaard & Davidoff, 1985). Nevertheless, when performing both a verification task and a naming task with either colour photographs or BW drawings, human performance was unaffected by the absence of colour (Biederman & Ju, 1988). The authors argued for a fast access to a coarse structural mental representation of objects; colour would only be used in the recognition of blurred objects, when the shape does not provide enough information for accurate categorisation or in the case of low level vision subjects. When target and non-target items are very similar in shape (i.e. subordinate classification), colour has indeed been shown to be relevant even when naming is not required (Price & Humphreys, 1989). Alternatively, an advantage was observed with coloured images -over BW ones- in a food object naming task but was not enhanced when the task was done with blurred images or when testing low vision patients (Wurm, Legge, Isenberg, & Luebker, 1993). These studies point towards a role of colour in late stages of processing to facilitate object recognition or naming. However, colour has recently been shown to play a role in very early visual recognition processes, in a delayed match to sample task using natural images (Gegenfurtner, 1997).

The aim of the present study is to test whether colour is an important feature in the rapid visual go/no-go categorisation task that we have used both in monkeys and humans. Processing speed is so fast that the system might have reached its limits (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2000). Thus if colour is one of the relevant features used in the early phases of visual processing, the absence of colour cues should induce an impairment in either accuracy or speed of performance. On the other hand, if the fast responses observed in our task can be produced using feed-forward processing of the fastest visual inputs to cortex, they should be independent of colour features, as the earliest responses in the visual cortex originate from the (achromatic) magnocellular stream, whereas the arrival of parvocellular chromatic information is delayed by roughly 20 ms (Nowak, Munk, Girard, & Bullier, 1995).

In the present study, the role played by colour in fast object categorisation was addressed in humans and monkeys using two different categories of target-objects: "food", and "animal". For both categories one might predict that colour could be important for segmenting images, since both food objects (such as ripe fruit) and animals (such as birds or fish) are often brightly coloured. Coloured and BW natural photographs were mixed at random to prevent subjects (humans or monkeys) from adopting different strategies when using chromatic and achromatic stimuli and to allow more direct comparison of performance in the two conditions. Monkeys and humans were tested on the same task with the same set-up and the same stimuli for a comparative study on the relative role played by colour cues in their strategies. Moreover, as monkeys performed hundreds of trials per day, the role of colour in the categorisation task could be analysed on both new and highly familiar natural scenes.

GENERAL METHODS

Subjects

Three rhesus monkeys were trained either on a Food/non Food (Rh1, male aged 6) or on an Animal/non Animal (Rh2 and Rh3, male and female both aged 4) categorisation task.

Two groups of 10 human subjects were also tested, one on each categorisation task, (mean age 37 in the food task and 35 in the animal task, with seven subjects in common). Subjects were mainly students or members of the laboratory staff; they all gave written consent to do the task and reported normal colour vision.

Task and set-up

The subjects (monkeys or humans) sat about 30-35 cm away from a tactile screen in the centre of which pictures were flashed for only 32 ms on a black background with a 1.5-3 s random intertrial between successive images. The subjects placed one hand on a capacitive tactile key located below the screen at waist level to start stimulus presentation. They had 1s to quickly release the button and touch the screen when they detected a target (animal or food) in the flashed image, otherwise they had to keep their hand on the button. This is a very demanding task : humans and monkeys have to make a succession of rapid decisions on the basis of brief stimulus presentations that prevent any foveating eye movements. Correct - go or no-go – decisions were rewarded by a beep noise. In addition, monkeys were given a drop of fruit juice. Incorrect decisions were followed by a 3-4 s display of the incorrectly classified stimuli delaying the next trial and the next possibility of reward 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. Adequate measures were taken to minimise any discomfort to the animals. They were restrained in a primate chair (Crist Instruments, GA USA) during testing and lived in a cage (European normalisation) in between the sessions.

Stimuli

All the pictures were natural scenes taken from a large commercial CD database (Corel). Some additional photographs (roughly 10%) were added for the food task in order to increase stimulus variety and to allow further controls to

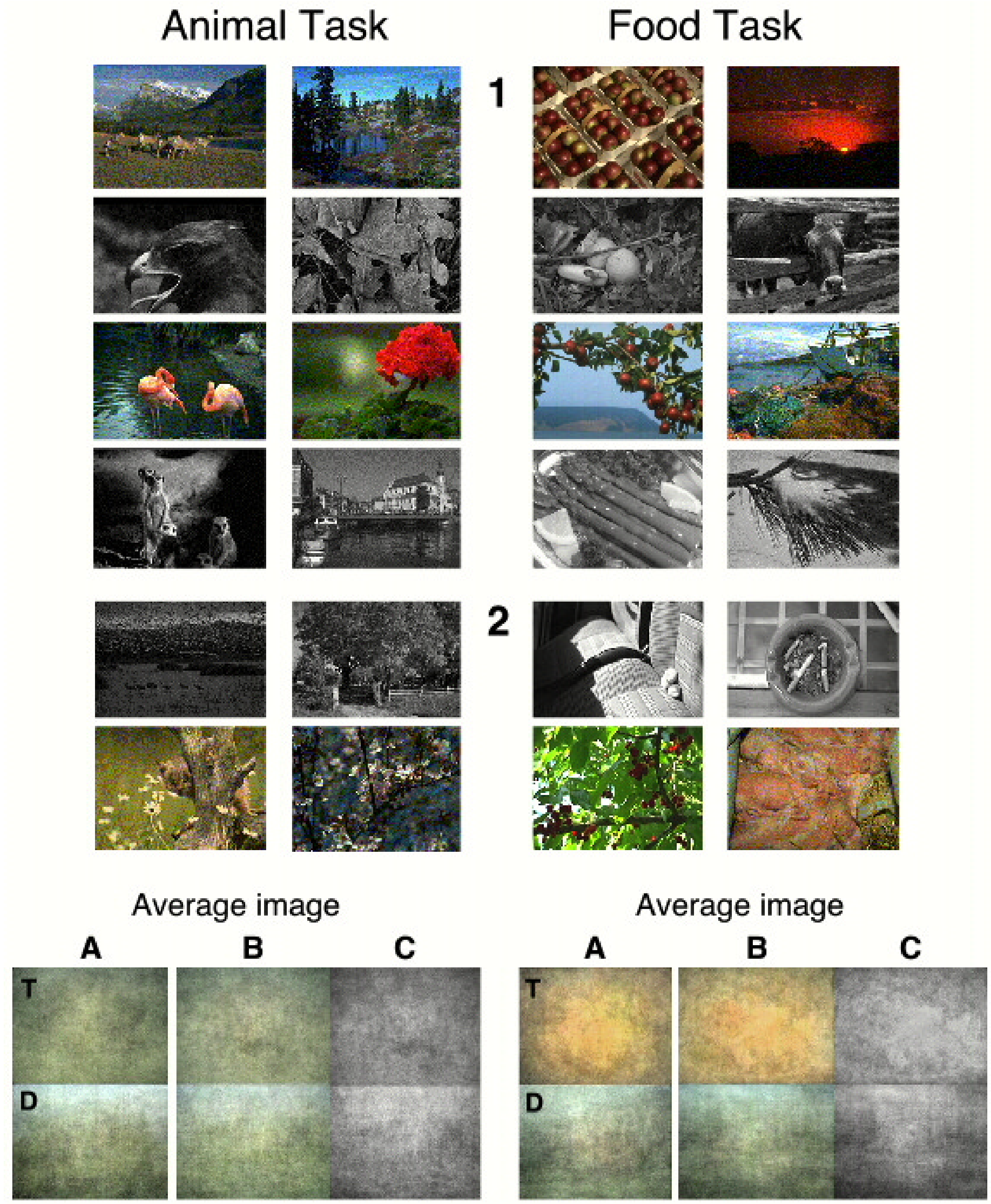

Fig. 1. Examples of the stimuli used in the animal and in the food categorisation tasks. Top : For each of the tasks, 6 targets are presented in the left column and six distractors in the right column. They are illustrated in colour or in BW as they were presented in the tasks. In all 4 columns, the first 4 photographs (1) were accurately classified by both the monkeys and the 10 human subjects, the bottom two (2) induced incorrect responses both in monkeys and in some of the human subjects.Bottom : Average images for targets (T, top row) and distractors (D, bottom row) have been calculated for each task : animal task (on the left) and food task (on the right) and for all horizontal images of each group of 200 photos. These averages allowed to search for a colour bias between sets of images and, within a given set, between targets and distractors. (A) average colour images for the set of photos presented in colour; (B) average colour pictures and (C) average BW images of the set of pictures that was then randomly chosen to be converted in BW images. Average images have been renormalised to reinforce differences that otherwise would remain undetected. For a given task all averaged images were considered together in the renormalisation process for rigorous comparison. Note that a color bias can clearly be seen in between targets and distractors especially in the food task so that the question of the role of colour cues addressed in the present study is definitely pertinent. Note also that there is no obvious colour bias between the set of images that were presented for categorisation either in colour (A) or in BW (B, C).

be performed. Targets and distractors were equiprobable and included both close-ups 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 some of the target category of the other task, landscapes, trees, flowers, objects, monuments, cars... On target trials, the subjects had no a priori knowledge concerning the position, the size or the number of targets in the picture. Moreover, both tasks included targets that were only partly visible, or partially masked in the scene. The photographs were so varied (Fig. 1) that contextual help can effectively be ruled out.

Images (192 x 128 pixels, corresponding to an angular size of about 25°/15°) were mostly horizontal photographs (73%). They were flashed 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. Colour and BW images were both converted from 24-bit colour photographs to 8-bit indexed pixels.

Evaluation of performance and Data analysis

Performance was evaluated both in terms of accuracy and speed. A go response was scored (whether correct on target trials or incorrect on distractor trials) when the subject released the key and touched the screen with the same hand in less than 1 s. The reaction time (RT) - delay between stimuli onset and button release - was recorded for all go responses. A no-go response was scored when the subject had kept pressing the key for over 1 s. Any other response (i.e. releasing the key without touching the screen) was considered as an incorrect response both on target and distractor trials. Accuracy differences were assessed using a standard two-tailed Chi^2 test on correct and incorrect responses; RT distributions were compared using two-tailed Mann-Whitney U test.

Procedure before testing monkeys with new achromatic stimuli

The training steps and first results have been reported previously (Fabre-Thorpe et al., 1998). At the end of the training period (4-6 months), monkeys were able to categorise accurately new stimuli (90,5% in the food task and 87% in the animal task) with very short RTs (mean values: 356 ms and 251 ms respectively for the food and the animal task).

To study the effects induced by the removal of colour cues, monkeys went through a number of new training and control steps. They were first trained to perform the categorisation task on a set of 200 images (taken from the set of 480-650 of images that they had already experienced) of which half were presented in BW (50 targets and 50 distractors). To prevent monkeys (and humans) from developing a new categorisation strategy in which colour cues would be ignored because stimuli were always -thus predictably- presented in BW, colour and BW images were mixed at random. Under these conditions, the monkeys scored as well as before on colour images and took 1-2 weeks to stabilise accuracy and speed with BW ones. Then, the role of colour cues on the processing of familiar images was studied, during 2-5 successive days, with a second set of 200 familiar images with all conditions equally balanced: (1) half were targets, half distractors, (2) all were seen in colour and in BW and finally (3) half were first presented in BW then in colour the second half being presented first in colour. Only the responses given by the monkey to the first BW and the first coloured presentation of each image were taken into consideration and compared. Results were replicated in the final experiment that is described in detail in the present study. They showed that monkeys can categorise BW images with very little impairment.

Testing performance with new achromatic stimuli

In the final testing phase, monkeys and humans were tested with 400 images that they had never seen before. For each task (Food vs. non-Food and Animal vs. non-Animal), 400 novel images (200 targets and 200 distractors) were chosen from the large database used previously (Fig. 1). For each task, the 400 images chosen were randomly divided into 2 sets of 200 images (100 targets, 100 distractors). For each set, the average horizontal colour image was computed separately for targets and distractors (Fig. 1). These average images illustrate the colour bias that exists between target and distractor images –especially in the food task- as well as the similarity of the two image sets. One set was then randomly chosen to be converted in BW.

Testing monkeys

Monkeys had to categorise daily 20 new images (10 in colour and 10 in BW of which half were distractors and half were targets) mixed with familiar colour and BW images that were randomly selected every day. With familiar images monkeys are rewarded in over 90% of the trials. Mixing new images with familiar ones (i) allowed performance to be compared for new and familiar stimuli, (ii) avoided the impact of response errors on the processing of subsequent images and thus (iii) ensured the stability of the emotional state of the monkey when faced with a new stimulus. As we only consider the response given by the monkey to the first presentation of a new image, this stability is a necessary condition to obtain consistent results. For the same reason, monkeys were tested with 40-100 familiar images before being presented with new images and 2 new images were never shown in immediate succession.

Testing humans

Subjects were all familiar with the task since they already volunteered for other studies using this rapid categorisation task (Fabre-Thorpe et al., 1998; Thorpe et al., 1996). They were tested using the same set-up and the same 400 stimuli that were presented in 4 blocks of 100 new stimuli. As with the monkeys, they were given feed-back about the accuracy of their responses: a beep noise for a correct response and a 3 s re-appearance of incorrectly classified images.

RESULTS

Only mild impairments were observed for both monkeys and humans when colour cues were removed, although colour appeared as a more important feature in the food task. For humans, the accuracy impairment varied from one subject to another indicating that individual subjects may rely differently on colour cues to reach their decision.

Accuracy in monkeys

New images: For the 3 monkeys the average accuracy with previously unseen photographs was 87.2% in colour and 87.3% in BW. Whatever the task (food task: 85.5% correct in colour vs. 87.5% in BW, two-tailed Chi^2=0.343, d.f.=1, p=0.66; animal task: 88% correct in colour vs. 87.2% in BW, two-tailed Chi^2=0.104, d.f.=1, p=0.83) there was no significant difference (Fig. 2). The same result was observed for each of the 2 monkeys (Rh2 and Rh3) tested in the animal task (Table I).

Familiar images: The 3 monkeys had performed a large number of trials with familiar images (1500-1750 trials for each condition). They averaged 95.2% correct in colour and 92.8% in BW (Fig. 2). This mild 2–3% accuracy impairment for BW images is statistically significant (table I). When accuracy was

Accuracy

			Colour			BW	
Food Task			T / D	Total	Chi2	Total	T / D
Familiar							
	Rh1	N trials	844 / 851	1695		1506	735 / 771
		N Cor trials	816 / 800	1616		1386	676 / 710
		%Cor	96.7 / 94	95.3	=.0004	92	92 / 92.1
New							
	Rh1	N trials	100 / 100	200		200	100 / 100
		%Cor	83 / 88	85.5	ns	87.5	91 / 84
Animal Task							
Familiar							
	Rh2	N trials	822 / 820	1642		1646	822 / 824
		N Cor trials	812 / 763	1575		1561	800 / 761
		%Cor	98.8 / 93	95.9	ns	94.8	97.3 / 92.4
	Rh3	N trials	915 / 856	1771		1747	883 / 864
		N Cor trials	893 / 780	1673		1604	862 /742
		%Cor	97.6 / 91,1	94.5	<.005	91.8	97.6 / 85.9
New							
	Rh2	N trials	100 / 100	200		200	100 / 100
		%Cor	94 / 87	90.5	ns	91.5	91 / 92
	Rh3	N trials	100 / 100	200		200	100 / 100
		%Cor	83 / 88	85.5	ns	83	86 / 80

Reaction Time (ms)

			Colour			BW	
Food Task			mean	med	U test	med	mean
Familiar							
	Rh1	Cor Go-trials	310	297	<.0001	309	322
		InC Go-trials	342	320		318	351
New							
	Rh1	Cor Go-trials	312	297	ns	306	324
		InC Go-trials	346	316		325	356
Animal Task							
Familiar							
	Rh2	Cor Go-trials	262	258	ns	258	264
		InC Go-trials	285	276		278	283
	Rh3	Cor Go-trials	247	240	ns	241	248
		InC Go-trials	271	256		266	266
New							
	Rh2	Cor Go-trials	270	266	ns	269	281
		InC Go-trials	314	292		300	350
	Rh3	Cor Go-trials	268	252	ns	242	262
		InC Go-trials	329	293		308	356

Table I. Monkeys' performance with new and familiar images in both colour and black and white (BW) conditions. On the left part of the table, the accuracy performance is shown for each of the two tasks (Food Task and Animal Task) and each of the three monkeys (Rh1, Rh2 and Rh3). For New and Familiar targets and distractors (T/D) the total number of trials (N trials), and the percentage of correct trials (% Cor) are indicated. For familiar images the total number of correct trials (N Cor trials) is also given. The same indication are given for targets and distractors pooled together (Total) with, in the central column the statistical result of a Chi^2 test between the colour and the BW conditions. On the right part of the table, the reaction time (expressed in ms) is also given for each of the two tasks (Food Task and Animal Task) and for each of the three rhesus monkeys (Rh1, Rh2 and Rh3). For new and familiar images, the mean and median (med) reaction times are indicated for all correct go-responses (Cor Go-trials) and all incorrect go-responses (InC Go-trials). The RT distributions of correct go-responses obtained in the two conditions (Colour and BW) have been compared using a two-tailed Mann-Whitney U test; results are shown in the central column (U test).

considered separately for targets and distractors, all monkeys showed a significant bias towards go responses with colour stimuli (two-tailed Chi^2, d.f.=1, p<.01 for Rh1, p<.0001 for Rh2 and Rh3). This bias persisted with BW photographs in the animal task (Rh2, Rh3) but disappeared for Rh1 that was tested on the food categorisation. The main accuracy effect associated with the absence of colour cues is seen in the detection of targets in the food task.

New vs. familiar images : Training improves accuracy in both tasks by about 5-10% in the chromatic condition and 3-8% in the achromatic condition.

Monkeys are extremely good and equally efficient at categorising new stimuli whether in colour or not. Colour is clearly not needed to explain the monkeys performance in this rapid categorisation task. Accuracy being similar when both BW and colour images are seen for the first time, the slight advantage for coloured stimuli that appears with familiar images might be due to the monkeys larger experience with chromatic images than with their BW counterparts.

Accuracy in human subjects

For the two groups of humans, the global average categorisation accuracy (Fig. 2) with previously unseen photographs was about 2% higher with colour photographs (93.2% in colour and 91% in BW). This tendency was not significant in the food task (90.7% correct in colour vs. 89% in BW, two-tailed Chi^2=2.8, d.f.=1, p=0.0942) but reached significance in the animal task (95.7% correct for colour images and 93% correct for BW image, two-tailed Chi^2=13.6, d.f.=1, p=0.0002). When the rate of correct responses was analysed separately for targets and distractors a difference appeared between the two tasks. In the animal task, the global 2.5-3% accuracy decrease observed with BW stimuli was seen for both targets and distractors. On the other hand, in the food task, the ability to detect targets dropped by 5.9% with BW stimuli (two-

FOOD TASK

Accuracy

	Colour			BW	
	T/D	Total	Chi 2	Total	T / D
F1 *	93 / 90	91,5	ns	90	90 / 90
F2*	93 / 82	87,5	ns	89,5	89 / 90
F3*	90 / 92	91	ns	86	76 / 96
F4*	91 / 89	90	ns	89,5	88 / 91
F5	93 / 84	88,5	ns	87,5	86 / 89
M1 *	97 / 83	90	ns	88	94 / 82
M2*	95 / 89	92	ns	88,5	89 / 88
M3*	96 / 92	94	ns	91,5	89 / 94
M4	93 / 93	93	ns	93	92 / 94
M5	93 / 85	89	ns	86,5	82 / 91
Total	93,4 / 87,9	90,7	ns	89	87,5 / 90,5
Chi 2	<.0001				<.04

RT on correct go-responses

	mean	med	U test	med	mean
F1 *	461	446	<.05	462	478
F2*	376	360	=.0004	396	412
F3*	506	493	<.05	508	521
F4*	390	368	<.06	384	403
F5	406	394	<.008	412	427
M1 *	410	397	=.09	401	428
M2*	449	438	ns	438	453
M3*	467	449	ns	474	481
M4	468	457	ns	465	479
M5	438	421	<.06	451	463
Total	437	427	<.0001	439	453

ANIMAL TASK

Accuracy

	Colour			BW	
	T/D	Total	Chi 2	Total	T / D
F1 *	98 / 96	97	ns	94,5	96 / 93
F2*	95 / 96	95,5	ns	94	94 / 94
F3*	94 / 94	94	<.06	88,5	90 / 87
F4*	96 / 94	95	ns	92	91 / 93
F6	98 / 99	98,5	<.02	94	94 / 94
M1 *	99 / 93	96	ns	96	97 / 95
M2*	90 / 97	93,5	ns	90,5	89 / 92
M3*	98 / 96	97	ns	94,5	94 / 95
M6	97 / 96	96,5	ns	93	92 / 94
M7	97 / 91	94	ns	92,5	95 / 90
Total	96,2 / 95,2	95,7	=.0002	93	93,2 / 92,7
Chi 2	ns				ns

RT on correct go-responses

	mean	med	U test	med	mean
F1 *	419	411	ns	408	419
F2*	345	330	ns	333	361
F3*	524	520	ns	503	513
F4*	385	376	ns	374	388
F6	469	455	ns	470	477
M1 *	387	381	ns	380	387
M2*	445	444	ns	440	445
M3*	442	429	ns	443	457
M6	427	417	ns	411	430
M7	363	346	ns	348	368
Total	420	412	ns	415	424

Table II. Individual human performance in the food task (left) and in the animal task (right). Female (F) and male (M) subjects that participated in each task are listed on the left; the asterisk indicates that the same subject participated in both experiments. For each of the subjects and for both colour and black and white (BW) images, the accuracy is given as the percentage of correct responses; the mean and the median (med) RT are given in milliseconds. In both tasks, human subjects had a tendency to categorise coloured images better than BW ones. Statistical comparison using a Chi^2 is shown in the central columns. However it only reached significance in the animal task (two-tailed Chi^2=13.6, d.f.=1, p=0.0002) and mainly as a result of two of the subjects (F3 and F6). The bias towards go responses is not significant in both colour and BW conditions for the animal task. It is highly significant in the food task when stimuli are in colour (p<.0001) and tends to disappear in the BW condition (p<.04). For RTs, in the animal task, the difference between conditions never reached significance at the global or at the individual level. In the food task, the average RT difference between the two conditions was highly significant (two-tailed Mann Whitney U=359702, p<0.0001) and reached significance for most of the subjects.

tailed Chi^2=20.1, d.f.=1, p<.0001) whereas the rate of correct no-go responses increased by 2.6% with BW distractors (almost reaching significance two-tailed Chi^2=3.5, d.f.=1, p<.07) which partly compensate for the accuracy drop with targets. This detection impairment for food targets ranged from 1% to 14% (Table II, 1% for subject M4, 14% for subject F3) whereas the impairment range was much smaller with animal targets (1-5%). Note that the same impairment in detecting food targets was seen for familiar images in monkeys.

As a global result, this study shows that, in rapid categorisation tasks, removing colour cues from the stimuli has on average, very little effect on human accuracy. It also shows that human subjects rely more heavily on colour to detect food targets relatively to animal targets. However, a large amount of variability between subjects was seen when considering individual performances. Whereas most subjects showed an impaired accuracy when categorising BW photographs, for some of them the global accuracy was identical in both conditions (Table II, subjects F2 and M4 in the food task; M1 in the animal task). For the subjects tested, the strength of the accuracy impairment induced by the removing of colour cues appeared correlated with RTs. The fastest subjects categorised equally well chromatic and achromatic stimuli whereas the accuracy advantage observed with chromatic stimuli increased progressively for subjects with longer RTs (correlation : R=0.81 in the animal task, R=0.72 in the food task). Colour could be a relevant feature in the visual processing leading to decision only for subjects responding with relatively long reaction times.

Speed of performance in monkeys

New images: In both tasks (Table I for individual results), RTs for correct go responses did not depend on whether the images were shown in colour or in BW (Fig. 2). In the animal task, mean RT was 269 ms (median: 259 ms) in colour and 271 ms (median: 261 ms) in BW (two-tailed Mann Whitney U=15269, p=0.68); in the food task mean RT was 312 ms (median: 297 ms) in colour and 324 ms (median: 306 ms) in BW (two-tailed Mann Whitney U=3651, p=0.31). The RT distributions for new BW and colour photos are illustrated for both tasks (Fig. 3). In the animal task they are identical for both BW and colour conditions, and the absence of colour cues had no consequence on the earliest correct go responses that are seen with latencies as short as 200 ms in both cases. Note that these responses cannot be simply considered as just random anticipations because, as targets and distractors are equiprobable, correct and incorrect anticipated go-responses should be equally distributed. From 200 ms on, correct go-

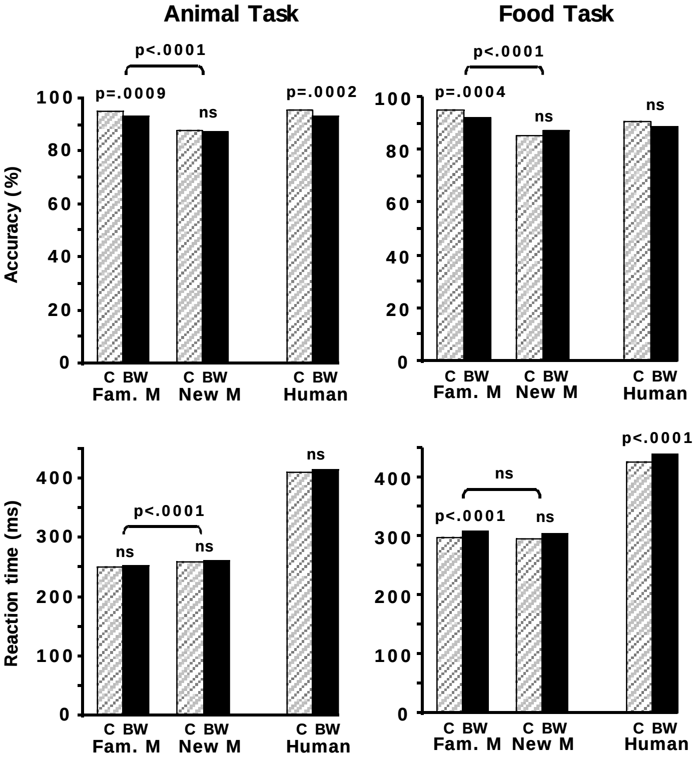

Fig. 2. Monkeys' and humans' global categorisation performance. Familiar images for monkeys (Fam. M); new images for monkeys (New M), new images for the group of 10 Humans (Human), images presented in colour (C; black bars) and in grey levels (BW; hatched bars). Top : accuracy expressed as the percentage of correct responses, for the Animal Task (left) and the Food Task (right). Note that monkeys categorise equally well new colour and BW images and that a very mild advantage is observed with colour pictures for monkeys tested with familiar images and for humans (the statistical significance is given by a Chi^2). Training leads to improved accuracy as shown by the higher scores observed for monkeys tested with familiar images; this is true in both colour and BW and with the same statistical significance. Bottom : The speed of response is illustrated for all conditions by the median reaction time (expressed in ms) for correct go-responses. Note the absence of an effect in the animal task and the 10-15 ms RT increase when food objects have to be detected in BW. This increase is significant (see text). Note that the effect of training on performance speed is only visible in the animal task.

responses significantly outnumbered incorrect go responses both in colour (12 correct vs. 1 incorrect go responses in the bin width 200-220; p<0.02) and in BW (12 correct vs. no incorrect go responses in the bin width 200-220; p<0.005). Thus, visual processing must have been completed and cannot exceed 100-120 ms as the 200 ms behavioural RT also includes time for decision and a large motor component. In the food task, the small delay seen for the earliest latencies is not very conclusive since only one monkey was tested (a total of 88 correct go responses in colour and 91 in BW) and the effect was not replicated with the larger sample of familiar photographs.

Familiar images: Concerning the performance speed, results with chromatic and achromatic familiar photographs confirmed the observations made with new images (Fig 2). No difference was seen in the animal task whereas, in the food task, RTs for correct go responses towards BW images were about 10 ms longer (median: 297 vs. 309 ms, mean: 310 vs. 322 ms) with the large sample of trials with familiar images (over 1500 trials in either condition), this shift towards longer latencies is significant (two-tailed Mann Whitney U=242504, p<0.0001). The RT distributions obtained in the food task (Fig. 3) shows that: (1) the earliest responses statistically biased towards correct go responses are produced at the same latency (220-230 ms) in both conditions, (2) the peak of the RT distribution for correct go responses is sharper for chromatic stimuli (variance 3321 in colour and 4326 in BW) with a clear mode in the bin width 270-280 ms. The effect associated with colour cues is mainly observed in the range 250-300 ms within which colour appears to facilitate the detection of food objects as more colour targets (an additional 9% relatively to the BW condition) are detected. On the other hand, in the achromatic condition, the RT distribution for correct go responses has no clear mode and extends towards long latencies responses with a greater percentage of go responses triggered after 400 ms in BW (6.5% in colour vs. 11% in BW).

New vs. familiar images : Whereas training induced a clear improvement in performance accuracy, there is little effect (if any) on the speed at which such a task is performed. In the food task, familiar and previously unseen photographs are categorised at exactly the same speed. The only effect was seen in the animal task in which a small RT decrease (10 ms) was observed with familiar images and in both monkeys. The fact that extensive training with photographs fails to speed up the neural processes underlying performance suggests that the speed at which monkeys categorise new images is already near to optimal.

In summary, the results obtained clearly demonstrate that for previously unseen images, the monkeys' abilities to categorise colour and BW images are almost identical for both accuracy and speed of response. A mild facilitation in the detection of food targets can be seen when colour cues are available, but this facilitation appears from 250 ms on. The experiment also shows that familiar images are categorised with higher accuracy than new ones but tend to be processed at about the same speed.

Speed of performance in human subjects

Speed of response: Data obtained with human subjects also showed that animal detection was not speeded up when colour cues were available whereas a mild effect could be seen in the detection of food targets. In the animal task (Fig. 2) the comparison of the overall RT distributions for correct go responses with either coloured or BW targets showed no statistical difference (RT in colour: mean=420 ms, median=412 ms; RT in BW: mean=424 ms, median=415 ms; Mann Whitney U=441635, p=0.58). The same result was obtained when comparing RT distributions individually for each subject (table II). In the food task, the results were clearly different. A statistically significant RT increase was seen with BW images (RT in colour: mean=437 ms, median=427 ms; RT in BW: mean=453 ms, median=439 ms; Mann Whitney U=359702, p<0.0001) although it was – as for the monkey - very small (about 15 ms). This result was confirmed at the individual level: the RT increase was seen in all subjects and reached significance for seven of them.

Humans are extremely efficient and can reach high accuracy scores with short RTs in both tasks. Compared to the animal task, the food task appears more difficult: the global accuracy is lower and the RTs are longer. Moreover, whereas the earliest responses start around 280 ms in the animal categorisation, they are delayed by at least 30 to 40 ms in the food categorisation (Fig. 3). Removing colour did not affect these earliest behavioural latencies. On the other hand, as it was already the case for monkeys in the food task, a greater percentage of go responses are triggered late (after 500 ms) with achromatic stimuli (17.5% in colour vs. 23,3% in BW) which is not the case in the animal task. In fact, the effect of removing

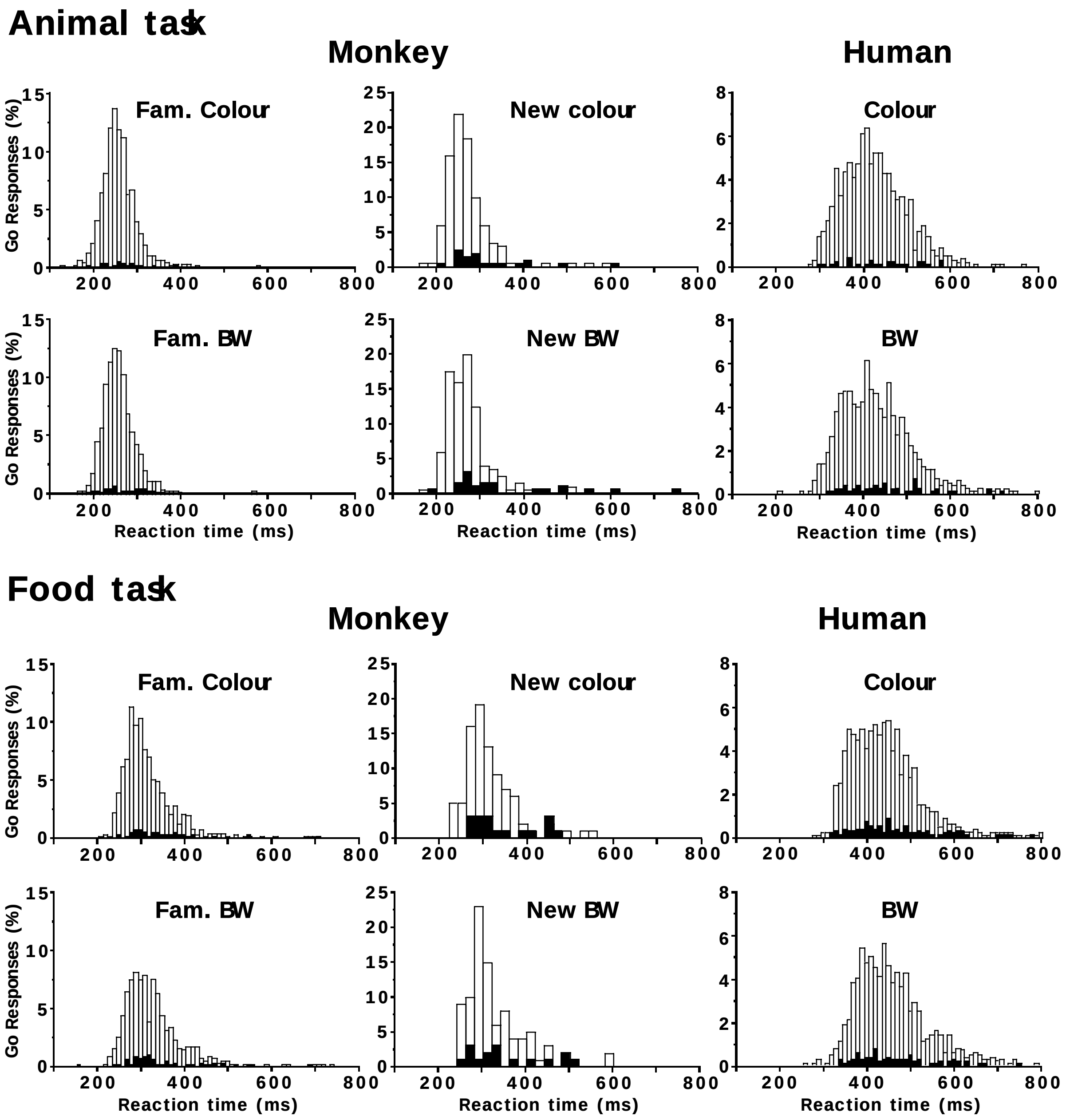

Fig. 3. Reaction time distribution of correct go responses and false positives for humans and monkeys. For each task : animal task (top) and food task (bottom), the RT distributions are shown for chromatic (top row) and achromatic (bottom row) image presentations. In each graph, the empty histogram corresponds to the correct go-responses whereas incorrect go-responses (false positives) are shown in black. Left column: monkey RT distributions (10 ms bin width) for go responses towards familiar (Fam.) stimuli. Central column : monkey RT distributions (20 ms bin width) for go responses with new photographs. Right column, RT distributions (10 ms bin width) for the group of 10 human subjects tested with the same previously unseen images. Reaction time is expressed in ms and in each bin width, go-responses are expressed as a percentage of all (correct and incorrect) go-responses.

colour cues was very mild, most stimuli were accurately and quickly identified in BW with, at the most, a 2% decrease in global accuracy and a 15 ms increase in RT. With the very varied total of 800 natural scenes proposed in these tasks, colour does not seem essential to allow fast and accurate categorisation of natural images. When available, colour cannot speed up the visual processes underlying the earliest, ultra-rapid, behavioural responses.

Comparison between humans and monkeys

When comparing the speed at which humans and monkeys perform the task, monkeys are much faster than humans; on average this advantage reaches roughly 130 ms in both tasks. Monkeys appear to behave like the fastest humans, combining short RTs and nearly no advantage for coloured stimuli. Note that (1) like humans, monkeys are faster to perform the animal categorisation task, (2) like humans, monkeys' go responses are delayed only in the food categorisation of BW stimuli, (3) this delay reaches about 10 ms in monkeys, a value compatible with the 15 ms delay observed in humans, (4) as for humans, colour appears more important as a feature for detecting food than animals, (5) monkeys and humans tend to make incorrect decisions in response to the same stimuli both in chromatic and achromatic conditions (Fig. 1).

GENERAL DISCUSSION

The first results replicate the data previously reported (Fabre-Thorpe et al., 1998; Thorpe et al., 1996): both monkeys and humans are fast and accurate at categorising natural images that they have never seen before, even without contextual help and without using eye movements. However studying the processing of 800 previously unseen natural scenes by humans and monkeys, and an additionnal 800 familiar natural scenes by monkeys, the main finding is that this sort of rapid visual processing of natural images is only very mildly affected by the removal of colour information. Humans' impairment without colour cues varies from one individual to another (some of them bein unaffected) and monkeys are fast and rely very little on colour.

The use of colour information in between tasks and species

Colour appears as a more relevant feature in the categorisation of food objects. The use of colour features to determine whether an item belongs to a category might depend on the pertinence of colour in identifying objects from that category. Although very few objects could be recognised on the basis of their colour only, colour could be more or less “diagnostic” in the recognition of certain categories of object (Biederman & Ju, 1988; Oliva & Schyns, 2000; Tanaka & Presnell, 1999). Monkeys were shown to encode pictures of fruits mainly along two dimensions : the type of fruit (apples or grapes) and their colour (red or yellow), ignoring other dimensions like their size or number (Sands, Lincoln, & Wright, 1982). For food, the colour is less arbitrary – i.e. varying in a smaller range - than in other categories like animals for example. This is shown by the yellow to orange predominance in the average food images that were computed from our sets of photographs (see Fig. 1). Colour could also help decision making in the case of ambiguous photographs such as close-ups of a rose vs. a green salad. Using top-down influences, the visual system could be "pre-set" to detect an object with special colour attributes; this could explain why RT are globally 10-15 ms shorter with chromatic stimuli. On the other hand, for animal categorisation, colour is clearly not essential, perhaps because it has no "diagnostic" value. Although it can be argued that cage-reared monkeys may not rely on animal colour to generalise their training, the fact that the same results were found in humans shows that fast detection of animal does not depend on colour cues.

Colour could also be used to help segmentation of the target-objects from the background. In that case, the contribution of colour may depend on how well objects can be segregated from their background and it could be argued that fast categorisation responses may only be seen when colour is not necessary for target segregation. In fact, the natural scenes used in our studies are very complex and we recently showed that ultra-rapid visual processing is not restricted to "easy to process" animal-targets (Fabre-Thorpe et al., 2000). It therefore appears that, at least for the animal task, short latency behavioural responses can be observed with the vast majority of targets. In the case of the food task, the effect of colour is seen around the mean RT when more chromatic targets are detected. Thus colour may not be used in the earliest processing stages but may be involved in later steps that can help improve object detection. In primates, trichromacy is thought to have evolved for segregating flowers and fruit from background foliage (Mollon, 1989; Regan et al., 1998) and probably plays a vital role in every day tasks such as choosing ripe fruit. In our data, the small 10-15 ms RT increase observed when colour is removed is mainly due to a small subset of targets that take abnormally long to detect in greyscale. In absence of colour, it is possible that more detailed analysis would be necessary for target detection. Thus the additional delay might be explained if, as suggested by (Smid, Jakob, & Heinze, 1997), it is faster to combine the coarse information about an object shape with its colour than to analyse its detailed local shape features. However, even if colour is used in the detection of some targets, and if it can improve pattern recognition in some cases (Syrkin & Gur, 1997), it is clearly not the most crucial aspect of the object used by monkeys and humans to perform the rapid categorisations studied here.

Neural correlates in IT

It seems likely that the processing leading to food or animal categorisation involves all the processing stages along the ventral visual pathway known to play a crucial role in object recognition (Milner & Goodale, 1993; Ungerleider & Mishkin, 1982). Thus, the data obtained in the present study support the idea, previously developed (Fabre-Thorpe et al., 1998; Thorpe et al., 1996; Thorpe & Imbert, 1989), that visual information processing in this task is mainly feed-forward. For the monkeys' earliest behavioural responses, visual processing must be restricted to roughly 100 ms. This finding is in agreement with neuronal responses in IT that typically have onset latencies of 80-100 ms (Oram & Perrett, 1992; Perrett, Rolls, & Caan, 1982) and with the short latency (100 ms) of the differential IT responses between target and non-target stimuli reported in categorisation tasks (Vogels, 1999b). The feed-forward aspect of processing is also supported by the similar latencies of the responses triggered towards new and familiar stimuli. Extensive training could have been expected to allow the bypass of some processing loops resulting in a RT decrease. This floor effect on the processing speed of natural scenes is also seen in humans (Fabre-Thorpe et al., 2000).

The minor role played by colour cues reported in the present study is in agreement with the characteristics of IT neuronal responses. IT neurones respond to faces, and specific objects (Booth & Rolls, 1998; Logothetis, Pauls, & Poggio, 1995; Perrett et al., 1982). Some can selectively respond to several objects that belong to the same category (Vogels, 1999b). Indeed, in monkeys performing a categorisation task with trees as targets, a quarter of IT cells responded differentially to tree vs. non tree stimuli. These responses are sensitive to image scrambling, but largely invariant to stimulus transformations such as changes in position or size. They are observed whether the objects are presented as coloured pictures, line-drawings, silhouettes or illusory contours (Chadaide, Kovàcs, Köteles, & Benedek, 1999), with simplified versions of the objects or with a combination of their features (Kobatake & Tanaka, 1994; Tanaka, 1997). IT cells are generally reported to be selective to shape irrespectively of colour, and only a very small subset of neurones needs a combination of shape and colour information to respond with maximal amplitude (Chadaide et al., 1999; Tanaka, Saito, Fukada, & Moriya, 1991). This view is also supported by two recent studies. In the first one (Booth & Rolls, 1998), macaques had been given a number of new objects to manipulate in their cages. Cellular recording in IT showed a small subset of totally "view-invariant" neurones, suggesting that objects rather than the visual features were coded. Moreover, most of them exhibited similar responses for colour or greyscale object images. The second study (Vogels, 1999b) used a tree vs. non tree categorisation in macaques and although the colour content of the image could affect the average response of IT cells, chromatic and achromatic stimulus presentations often elicited similar neuronal responses. Thus processing in IT cells is consistent with the mild effect observed in our task when colour cues are removed.

Fast processing using the magnocellular pathway ?

The hypothesis that the use of colour features is time consuming is supported by at least three different results reported here. (1) humans with the largest accuracy advantage for colour stimuli were those which had the longest mean reaction times, (2) both monkeys and humans are slower in performing the food task in which colour is a more relevant feature, (3) the earliest behavioural responses do not depend on colour cues. Visual information can reach cortical area V1 using either the magnocellular (M) system or the parvocellular (P) system. Traditionally, the M system has been associated with the extraction of structure from motion whereas the P system is thought to be concerned with the fine analysis of static images. A possible explanation for the late use of colour is based on the FACADE model (Bradski & Grossberg, 1995; Grossberg, 1994) and the sequential use of parvocellular information : boundaries would first be formed and then control the surface filling-in of brightness and colour information. Boundaries could directly activate object recognition processes but the use of colour would take longer. Alternatively, we would like to suggest that in our task, the absence of a clear accuracy advantage when colour cues are available could be related to differences in the temporal dynamics of processing in the M and P pathways of the visual system. Chromatic information in the parvocellular stream reaches visual cortex roughly 20 ms after the magnocellular inputs that mainly transmit motion and luminance based information (Nowak & Bullier, 1997; Nowak et al., 1995). This delay could be even longer if, as suggested recently, colour vision depends upon the koniocellular pathway (Calkins & Sterling, 1999). As has been argued elsewhere, the sort of rapid visual categorisation performed in our tasks could depend on the unidirectional processing of the first 10-20 ms of activity in each cortical area, the analysis would be based on coarse and near colour-blind magnocellular information. It follows that colour would only be important for images that could not be accurately categorised on the basis of such early information.

Coarse to fine processing has already been proposed by a number of authors (Carpenter & Grossberg, 1987; Parker, Lishman, & Hughes, 1992; Schyns & Oliva, 1994), and in a recent study (Sugase, Yamane, Ueno, & Kawano, 1999) the discharge of IT neurones in response to faces has been shown to have a fast phasic component related to the presentation of a face regardless of its other characteristics and a second tonic component developing with a 40-50 ms delay which is linked to finer information concerning for example, the owner of the face (human or non human) or the facial expression. The coarse magnocellular information might thus be sufficient to get access to a global shape representation that might be adequate in most cases in our task for the ultra-rapid categorisation of natural scenes – i.e. the fast behavioural responses produced in our task both by monkeys and humans. Such fast process could be used as a header to improve further processing of colour and fine details.

Acknowledgements :
This work was supported by the CNRS, by the Cognisciences Program, and by the Midi-Pyrénées Region. Authorization for experiments with humans (CCPPRB N° 9614003).

References

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. ***Cognitive psychology, 20***, 38-64.

Booth, M. C. A., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. ***Cerebral Cortex, 8***, 510-523.

Bradski, G., & Grossberg, S. (1995). Fast-Learning VIEWNET Architectures for Recognizing Three-dimentional Objects from Multiple Two-dimensional Views. ***Neural Networks, 8***, 1053-1080.

Calkins, D. J., & Sterling, P. (1999). Evidence that circuits for spatial and color vision segregate at the first retinal synapse. ***Neuron, 24***, 313-321.

Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. ***Computer Vision, Graphic and Image Processing, 37***, 54-115.

Chadaide, Z., Kovàcs, G., Köteles, K., & Benedek, G. (1999). Selectivity of macaque inferior temporal neurons for shapes with different surface attributes. ***Perception, 28 (suppl.)***, 97.

D'Amato, M. R., & Van Sant, P. (1988). The person concept in monkeys (***Cebus apella)***. ***Journal of Experimental Psychology: Animal Behavior Processes, 14***(1), 43-55.

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2000). A limit to the speed of processing in Ultra-Rapid Visual Categorization of novel natural scenes. ***Journal of Cognitive Neuroscience***, in press.

Fabre-Thorpe, M., Richard, G., & Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. ***Neuroreport, 9***(2), 303-308.

Gegenfurtner, K. R. (1997). Sensory and cognitive contributions of color to the recognition of natural scenes. ***Investigative Ophthalmology and Visual Science, 39***, S156.

Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. ***Perception & Psychophysics, 55***(1), 48-121.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. ***Journal of Neurophysiology, 71***(3), 856-867.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. ***Curr. Biol., 5***(5), 552-563.

Milner, A. D., & Goodale, M. A. (1993). Visual pathways to perception and action. ***Progress in Brain Research, 95***(317), 317-337.

Mollon, J. D. (1989). "Tho she kneeled in that place were she grew...". The use and origins of primate colour vision. ***J. Exp. Biol., 146***, 21-38.

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.

Nowak, L. G., Munk, M. H. J., Girard, P., & Bullier, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. ***Visual Neuroscience, 12***, 371-384.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. ***Cognitive Psychology, in press***.

Oram, M. W., & Perrett, D. I. (1992). Time Course of Neural Responses Discriminating Different Views of the Face and Head. ***Journal of Neurophysiology, 68***(1), 70-84.

Ostergaard, A. L., & Davidoff, J. B. (1985). Some effects of color on naming and recognition of objects. ***Journal of Experimental Psychology : Learning, Memory, Cognition, 11***(3), 579-587.

Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. ***Perception, 21***, 147-160.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. ***Experimental Brain Research, 47***(3), 329-342.

Price, C. J., & Humphreys, G. W. (1989). The effects of surface detail on object categorization and naming. ***The Quaterly Journal of Experimental Psychology A, 41***(4), 797-827.

Regan, B. C., Julliot, C., Simmen, B., Viénot, F., Charles-Dominique, P., & Mollon, J. D. (1998). Frugivory and colour vision in Alouatta seniculus, a trichromatic platyrrhine monkey. ***Vision Research, 38***, 3321-3327.

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 Processes, 14***(3), 247-260.

Sands, S. F., Lincoln, C. E., & Wright, A. A. (1982). Pictorial similarity judgments and the organization of visual memory in the rhesus monkey. ***Journal of Experimental Psychology: General, 111***(4), 369-389.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges : Evidence for time and scale dependent scene recognition. ***Psychological Science, 5***, 195-200.

Smid, H. G. O. M., Jakob, A., & Heinze, H. J. (1997). The organisation of multidimensional selection on the basis of color and shape : An event-related brain potential study. ***Perception and Psychophysics, 59***(5), 693-713.

Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. ***Nature, 400***(6747), 869-873.

Syrkin, G., & Gur, M. (1997). Colour and luminance interact to improve pattern recognition. ***Perception, 26***, 127-140.

Tanaka, J. W., & Presnell, L. M. (1999). Color diagnosticity in object recognition. ***Percept. Psychophys., 61***(6), 1140-1153.

Tanaka, K. (1997). Mechanisms of visual object recognition: monkey and human studies. ***Current Opinion in Neurobiology, 7***(4), 523-529.

Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. ***Journal of Neurophysiology, 66***, 170-189.

Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. ***Nature, 381***(6582), 520-522.

Thorpe, S. J., & Imbert, M. (1989). Biological constraints on connectionist models. In R. Pfeifer, Z. Schreter, F. Fogelman-Soulié, & L. Steels (Eds.), ***Connectionism in Perspective.*** (pp. 63-92). Amsterdam: Elsevier.

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-USA: MIT Press.

Vogels, R. (1999a). Categorization of complex visual images by rhesus monkeys. Part 1 : behavioural study. ***European Journal of Neuroscience, 11***, 1223-1238.

Vogels, R. (1999b). Categorization of complex visual images by rhesus monkeys. Part 2 : single-cell study. ***European Journal of Neuroscience, 11***, 1239-1255.

Wurm, L. H., Legge, G. E., Isenberg, L. M., & Luebker, A. (1993). Color improve object recognition in normal and low vision. ***Journal of Experimental Psychology : Human perception and performance, 19***(4), 899-911.