Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects

Rufin VanRullen, Simon J Thorpe
Centre de Recherche Cerveau et Cognition, UMR 5549, CNRS-UPS, 133 route de Narbonne,
31062 Toulouse Cedex, France; e-mail: rufin@cerco.ups-tlse.fr, thorpe@cerco.ups-tlse.fr
Received 5 January 2000, in revised form 20 December 2000

Abstract. Visual processing is known to be very fast in ultra-rapid categorisation tasks where the subject has to decide whether a briefly flashed image belongs to a target category or not. Human subjects can respond in under 400 ms, and event-related-potential studies have shown that the underlying processing can be done in less than 150 ms. Monkeys trained to perform the same task have proved even faster. However, most of these experiments have only been done with biologically relevant target categories such as animals or food. Here we performed the same study on human subjects, alternating between a task in which the target category was 'animal', and a task in which the target category was 'means of transport'. These natural images of clearly artificial objects contained targets as varied as cars, trucks, trains, boats, aircraft, and hot-air balloons. However, the subjects performed almost identically in both tasks, with reaction times not significantly longer in the 'means of transport' task. These reaction times were much shorter than in any previous study on natural-image processing. We conclude that, at least for these two superordinate categories, the speed of ultra-rapid visual categorisation of natural scenes does not depend on the target category, and that this processing could rely primarily on feed-forward, automatic mechanisms.

1 Introduction

The analysis of the visual scene is to human observers an effortless process that seems almost instantaneous. However, it requires a tremendous amount of processing, such that even the most sophisticated artificial vision systems are very far today from reproducing human performance. Yet a number of recent studies have underlined the fact that this kind of processing is achieved with relative ease and impressive speed by human and nonhuman primates even under the most constraining conditions. In a task where the subject has to decide whether a briefly flashed photograph (presentation time 20 ms) contains an animal or not, humans were able to respond in under 400 ms (Thorpe et al 1996), with a bias towards correct responses starting 280 – 290 ms after stimulus onset (Fabre-Thorpe et al 1998b). Monkeys have proved even faster, with mean reaction times around 250 ms, and a bias towards correct responses starting before 200 ms (Fabre-Thorpe et al 1998b). In the same study, another monkey was able to categorise images containing 'food' versus 'non-food', and, in a recent experiment by Vogels (1999a), two monkeys were trained to categorise natural images of 'trees' versus 'non-trees' with average reaction times inferior to 250 ms. This special kind of rapid visual processing displays a number of interesting properties. First, it does not require foveal vision, and images can be presented randomly at different eccentricities without a concurrent cost in accuracy (Fabre-Thorpe et al 1998a; Thorpe et al 1999). Second, the presence of colour information in the stimulus images is not necessary for this form of rapid visual categorisation (Delorme et al 2000), suggesting a major role of the magnocellular pathways in the underlying information transmission and processing mechanisms. Finally, these rapid reaction times cannot be shortened by familiarity with the stimulus images, even after three weeks of intensive training (Fabre-Thorpe et al 2001).

These results are fundamental for current theories of visual processing because they give an upper limit to the time required by the visual system to analyse a complex scene. Because the reaction time also includes the time needed to generate the motor command, the actual duration of visual processing itself is likely to be much shorter. Indeed, Thorpe et al (1996) demonstrated that event-related potentials (ERPs) to target and distractor images diverge strongly at roughly 150 ms post stimulus. Single-unit recordings in the anterior temporal cortex of rhesus monkeys showed a selectivity to exemplars of the target category ('tree') as early as 80–90 ms after stimulus presentation (Vogels 1999b). Neurons in the inferotemporal cortex (IT) are also known to respond selectively to faces of humans or monkeys after only 100 ms (Perrett et al 1982; Oram and Perrett 1992). The implications of these findings for computational theories of visual processing are difficult to avoid: given the number of synaptic stages between the retina and higher-level visual areas such as IT, processing is very likely to be done with a single feed-forward pass through the system, and with very few spikes emitted by each neuron (Thorpe and Imbert 1989).

While the specific 'ultra-rapid visual processing' described in these studies can certainly not account for the full range of tasks that can be performed by the human visual system, an ability to analyse the visual environment rapidly is undoubtedly a strong and critical component of biological visual systems. The survival of an animal often depends on the speed with which predators, prey, or food can be recognised. Therefore, it is fundamental to determine precisely the conditions under which this form of ultra-rapid visual processing can occur. In particular, the results presented above cannot be generalised yet to the categorisation of all kinds of complex visual scenes because, in the experiments mentioned, the target categories used had a high 'biological' relevance to the behaviour of the primate subject. Animals, food, and even trees have been present in the primates' natural environment for millions of years, and have a strong meaning in terms of the survival of the observer. Therefore it is possible that they constitute a special case for vision, and that their processing might rely on hardwired innate mechanisms, genetically inherited, or at least highly specialised. Indeed, there is some evidence for such differences in that speed of name retrieval differs when identifying pictures of natural objects versus artifacts (Humphreys et al 1999).

Here we present an experiment where human subjects had to categorise complex scenes on the basis of the presence or absence of a clearly artificial category: 'means of transport'. The target images were natural colour photographs chosen to be as varied as possible, including many examples of cars, trucks, trains, boats, aeroplanes, helicopters, hot-air balloons, etc (see examples in figures 2 and 3). To allow direct comparison between performance in this task and performance in a categorisation task involving a 'natural' category, subjects had to alternate between blocks of 96 images where the target category was 'means of transport' and blocks of 96 images where it was 'animal'. Furthermore, half of the distractor images in each task were targets for the other task, ie means of transport in the animal categorisation task, and animals in the means-of-transport task. The results presented in the next sections demonstrate that visual categorisation of artificial objects in complex scenes is neither slower nor less accurate than categorisation of biologically relevant images such as animals.

2 Materials and methods
2.1 Procedure
The subjects were sixteen volunteers, eight males and eight females, with ages ranging from 21 to 50 years. They were seated in a dark room, at approximately 120 cm from a computer screen piloted from a PC computer. Two categorisation tasks with a go/no-go paradigm were performed in alternation by all subjects. In each task they viewed 10 series of 96 colour images, half of which were targets and half of which

were distractors. Whether the first block was an animal or a means-of-transport block was randomly decided for each subject. Then the subjects alternated between blocks of 96 images of one task and of the other. A trial was organised as follows: a white fixation point (smaller than 0.1 deg) was drawn in the middle of the black screen; subjects had to press a touch-sensitive button to start the trial; the image, which subtended roughly 10 deg, was flashed at the centre of the screen for 20 ms; no back-ward masking was applied after image presentation; subjects had to release the button within 1 s if the image contained an element of the target category, and maintain pressure otherwise. The intertrial interval was 2 s plus or minus a random delay of no more than 200 ms, to prevent the subjects from locking their response to the expected time of presentation. Note that the very short presentation time ensures that there was no possibility of making exploratory eye movements during image presentation. An important issue is the changes in the experimental procedure that we made compared to previous similar studies (eg Thorpe et al 1996). The purpose of these changes was to shorten the reaction times as much as possible. Therefore, the image size was enlarged to 10 deg (instead of 5 deg) to increase stimulus energy, and we used a touch-sensitive plate instead of a computer mouse to record the subject's response.

2.2 Images
The pictures were complex colour scenes taken from a large commercially available CD-ROM library allowing access to several thousand stimuli. The images in each category were chosen to be as varied as possible. The animal category included pictures of mammals as well as birds, fish, insects, and reptiles. The means-of-transport category included images of cars, trucks, trains, civil and military aeroplanes, helicopters, boats, hot-air balloons, and even rockets. Subjects had no a priori information on the size, position, or number of the targets in a single photograph. There was also a very wide range of distractor images, which could be outdoor or indoor scenes, natural land-scapes or street scenes with buildings and roads, pictures of food, fruits, vegetables, or plants, houses, man-made objects or tools. Examples of the photographs are shown in figures 2, 3, and 4. In each block, subjects were presented with 96 photographs—48 targets and 48 distractors. Half of the distractors were targets from the other task, ie vehicles in the animal task, and animals in the means-of-transport task. Further-more, half of the vehicle images (24 per series in the means-of-transport task, 12 per series in the animal task) were images of cars and the other half of different means of transport, to allow further data analysis of intracategory differences. To prevent effects of image-specific learning, an image was presented once for each subject and could not appear both as a target in one task and a distractor in the other task. Image sequences were randomised so that different subjects were presented differ-ent image sequences over the same series of 96 trials, and that the presentation of a target was not predictable. Finally, all subjects alternately viewed 10 blocks of 96 trials for each task, making a total of 1920 images. The order of appearance of the series was randomly determined for each subject, as well as whether they would start with an animal or a means-of-transport categorisation series.

3 Results
3.1 Primary results
Reaction times and percentage-correct scores are shown for each subject in table 1. Note that, although it was clearly not the purpose of the present study, we could not demonstrate any systematic pattern of differences between male and female subjects. Performance appears remarkable, given the complexity and the very short presentation times of the photographs. The percentage of correct responses is close to 95% in both tasks, and the median reaction time slightly above 350 ms. Some of the subjects show

Table 1. Percentage-correct scores, and mean and median reaction times (RTs) are presented for each of the sixteen subjects in both tasks, as well as the time after stimulus presentation (± 5 ms) when the distribution of reaction times for correct responses becomes significantly (χ^2 test over 10 ms periods, 1 df, $p < 0.001$ for more than five consecutive periods) different from the distribution for false positives (discrimination time t_{discr}). Mean values over the group of subjects are also shown. The last line shows the performances obtained when the statistics are calculated from the complete data set obtained from all subjects (here the discrimination time is obtained using a χ^2 test over 10 ms periods, 1 df, $p < 0.001$ for more than 30 consecutive periods).

| Subject	Animal task				Transport task			
	score	mean RT	median RT	t_{discr}	score	mean RT	median RT	t_{discr}
	% correct	ms	ms	ms	% correct	ms	ms	ms
BAT	98.12	413.6	400	335	96.35	424.1	411	345
CHA	97.08	397.9	387	325	97.92	431.1	412	335
DEL	89.58	332.1	321	285	85.62	348.2	328	295
FAB	97.92	358.1	347	305	98.44	374.9	350	305
FIZ	91.77	389.3	372	315	93.23	410.5	397	325
GIL	96.56	467.3	458	395	94.48	472.3	457	375
LAG	96.46	440.5	429	375	95.31	446.5	432	365
MAR	96.77	305.2	293	265	91.77	336.6	307	265
MAS	90.21	293.3	290	265	87.19	316.4	296	275
PAQ	92.60	332.0	317	285	90.73	335.4	318	275
PER	95.62	398.5	380	325	95.94	395.5	369	315
PET	93.44	382.1	376	335	93.23	399.1	392	305
TAY	93.33	334.0	327	275	92.81	336.1	332	265
THO	93.02	327.5	316	275	95.00	328.4	306	255
TRZ	96.98	357.3	344	285	97.08	351.4	333	285
VAN	92.71	296.0	288	255	92.60	309.1	293	245
Mean	94.51	364.0	352.8	306.2	93.61	376.0	358.3	301.9
Overall	**94.51**	**363.8**	**350**	**225**	**93.61**	**375.6**	**357**	**245**

even better performance, with median reaction times under 300 ms and percentage-correct scores close to 93% in both tasks. We also indicate the time at which 'go' responses on targets significantly outnumber 'go' responses on distractors at the $p < 0.001$ level (this time will now be referred to as 'discrimination time'). For this we applied a χ^2 test using 10 ms bins (1 df), and determined the first bin for which $p < 0.001$ for at least five consecutive bins in the case of data from a single subject, and for at least thirty consecutive bins when using the data pooled from all sixteen subjects. Because of the strictness of the criteria used, no correction was made for repeated testing. The obtained discrimination time can be inferior to $250 - 260$ ms in some of the subjects.

When considering the performance for all subjects, one can ask whether there is a significant difference between the animal and the transport tasks. Table 2 shows that there is no such difference (significance is defined as $p < 0.05$ for a paired t-test across subjects, 15 degrees of freedom) for either percentage-correct scores or median reaction time. There is also no significant difference in the moment at which correct 'go' responses significantly outnumber false positives (discrimination time). Thus, subjects can perform the transport categorisation task as quickly and as efficiently as the animal task. This constitutes the primary result of this study. The statistical difference between the two tasks observed for mean reaction times only reflects the fact that there are more long-latency responses in the transport task, but the lack of a difference for the medians implies that neither task is actually faster than the other.

Table 2. Statistically significant differences between the two tasks (paired t-test, df $= 15$). Discrimination time (t_{discr}) denotes the time (± 5 ms) when correct 'go' responses significantly (χ^2 test over 10 ms periods, 1 df, $p < 0.001$ for more than five consecutive periods) outnumber false positives.

	Difference
	animal task versus transport task
Score/% correct	ns ($p = 0.082$)
Mean reaction time	$p < 0.001$
Median reaction time	ns ($p = 0.067$)
Discrimination time, t_{discr}	ns ($p = 0.186$)

More detailed analysis of the reaction-time histograms (figure 1) shows that the distributions for the animal and transport tasks are statistically indistinguishable (ie $p > 0.1$ for every 10 ms period, using a χ^2 test with 1 degree of freedom) before $280 - 290$ ms. Responses with reaction times below 220 ms are probably due to anticipations, because the distributions of responses to targets and distractors are roughly equal. However, from $220 - 230$ ms in the animal task, and $240 - 250$ ms in the transport task (difference not significant, see table 2) the proportion of responses to targets significantly ($p < 0.001$, see table 1 for a description of the statistical test) outnumbers that for distractors, whereas targets and distractors have the same probability of presentation (50%): this means that the same categorisation tasks with a response time limit of 250 ms would yield a performance significantly above chance! This time window includes not only the time required for visual processing by itself, but also the decision process and the motor output. This severely limits the time available for visual processing, and therefore is a very important result in regard to current computational theories of visual processing.
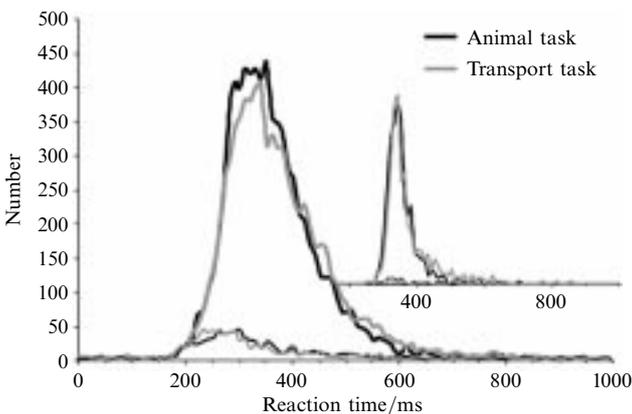
Figure 1. Histograms of reaction times for the two tasks. The reaction times for the false positives are plotted as thin lines. The small inserted curve shows data from subject FAB. Note the low number of errors, the good match of the curves for the two tasks, and the narrowness of the histogram.

The examples of images that were categorised the most rapidly, as well as the images that were incorrectly categorised by the most subjects (figures 2 and 3) illustrate the complexity of the two tasks. Target objects are often highly visible, but there does not seem to be any preferred size, ('canonical') orientation, nor any preferred subcategory (car images represented half of the set of vehicle images, thus a high proportion of cars would be expected). Since half of the distractor images belonged to the target

Animal task

(a) 308.9   16   309.3   16   316.2   16   316.4   16   316.9   16

246.4   5   225.7   4   229.5   4   236.5   4   237.2   4

(b) 241.2   4   244.2   4   248.7   4   251.2   4   254   4

(c) *255.5*   *6*   *267*   *5*   *282*   *5*   *290.8*   *7*   *293.3*   *6*

(d) -   *16*   -   *15*   -   *8*   -   *7*   -   *5*

281   1   283   1   287   1   289   1   292   1

(e) *292*   *1*   293   1   294   1   294   1   294   1

Figure 2. Examples of images presented in the 'animal' categorisation task. Original images were presented in colour. Mean reaction times and number of responses (out of sixteen subjects) are mentioned when available below each image, except for (d), which shows the number of missed trials (out of sixteen subjects). Errors are presented in italics. (a) Fastest targets. Target images having the shortest mean reaction times. Note that all of these images were correctly categorised by all subjects. (b) Targets categorised by the most subjects under 270 ms. The number of responses and mean reaction times are calculated only on those trials with response times inferior to 270 ms and superior to 150 ms (to sort out obvious anticipation trials). (c) False positives, from five or more subjects, with the shortest reaction times. (d) Most commonly missed targets. The number of subjects that did not respond to these targets is indicated. (e) Fastest trials from subject FAB, including potential false positives.

303.6   16   309.5   16   312.8   16   316.9   16   317.7   16
(a)

236.3   6   241.3   6   254.3   6   235.4   5   245.4   5

250.6   5   254.8   5   256.4   5   222.5   4   229.2   4
(b)

261.2   5   261.3   6   288   5   298.4   5   299.8   6
(c)

-       13   -   12   -   11   -   11   -   10
(d)

258   1   265   1   *269   1*   283   1   286   1

287   1   288   1   288   1   *289   1*   295   1
(e)

Figure 3. Examples of images presented in the 'transport' categorisation task. Notations as in figure 2. See text for details.

category of the other task, they are as likely to appear in the false-positive trials as other 'regular' distractors. However, images of aircraft or helicopters on a sky background (figure 2c) seem to be frequently mistaken for animals (most probably birds or flying insects). The targets missed by the most subjects seem to have in common the small size and uncentred position of the target objects in these images. Nevertheless, these objects were perfectly visible since the images subtended 10 deg. This effect might thus be due to the very short presentation time (20 ms). Finally, the fastest trials from subject FAB (figures 2e and 3e) show that correct responses can be obtained on varied

images even with very short reaction times. Only one error is observed out of these 10 fastest trials in the animal task, and two in the transport task.

Of the 480 distractor images that were neither animals nor means of transport, 224 were correctly ignored by all sixteen subjects. Figure 4 provides examples of such stimuli and illustrates the very wide range of photographs presented to the subjects. Clearly, distractor images cannot be defined uniquely by a simple feature such as their configuration (for example, the presence of the sky), their overall luminance, contrast, or range of spatial frequencies. The strategy used by the subjects must therefore reflect the diversity of the images to be categorised.

Animal task

Transport task
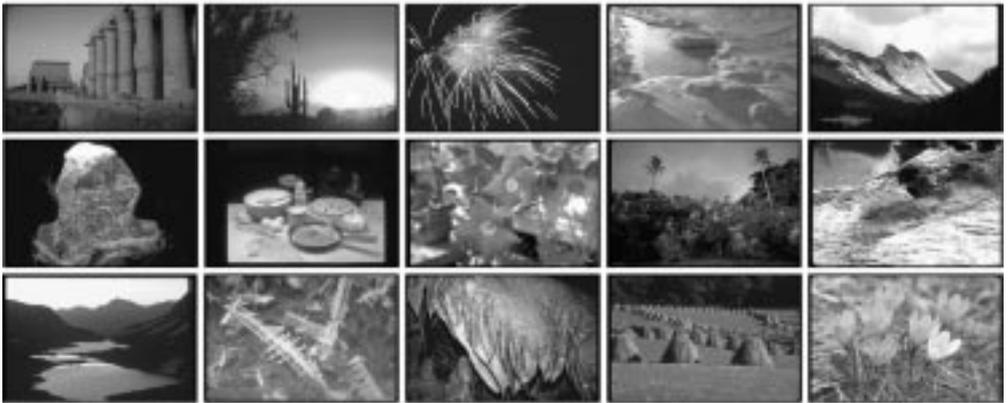
Figure 4. Examples of distractor images (nonanimal, nonvehicle) correctly categorised by all sixteen subjects in the two tasks. In the animal task, out of 240 distractor images, 116 were correctly categorised by all sixteen subjects, and 231 by more than half the subjects. In the transport task, out of 240 distractor images, 108 were correctly categorised by all sixteen subjects, and 236 by more than half the subjects. The great variety of distractor images and the performance obtained argue against the possibility that categorisation could be based on a limited set of simple low-level features.

3.2 Influence of the other categorisation task

Our experimental procedure also allows comparison between the false-positive trials on 'regular' distractors and on distractors that would be targets in the other task (ie means of transport in the animal task, animals in the means-of-transport task). There were far more errors for stimuli that changed status (900 for all subjects in both tasks) than for the ones that were always to be treated as a distractor (532) despite the fact that the probabilities of presentation were equal (figure 5). Yet there is no significant
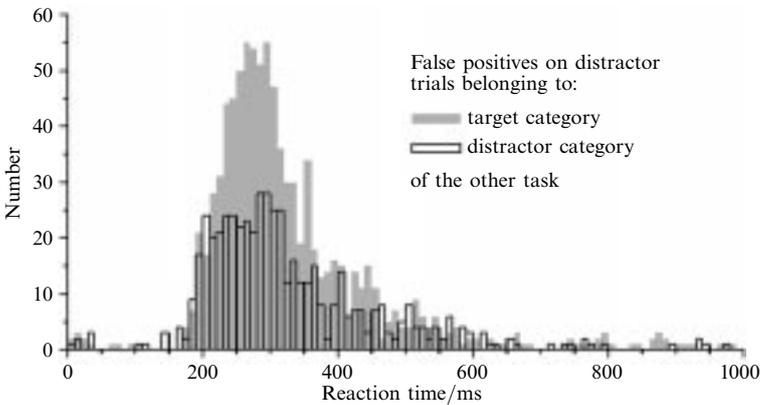
Figure 5. Histogram of reaction times for the false-positive trials in both tasks. Trials are separated according to the type that the image would have in the other task: distractor or target. Note that the probability of presentation is identical for both groups.

difference (ie $p > 0.1$ for every 10 ms period, using a χ^2 test with 1 degree of freedom) between the distributions of responses for these two categories of distractors before 230 – 240 ms (where significance level reaches $p < 0.02$), strengthening the idea that reaction times after this critical limit reflect the complex visual processing that is required for category discrimination.

There seems to be a strong influence of the other task. Subjects had difficulties withholding responding to images that would be targets in the other task, ie they cannot completely switch from one task to the other without a concurrent cost in accuracy. Although the alternation between the two tasks certainly leads to biases in the observed performances, it is important to stress that this procedure means that performance can be compared under similar conditions for the two tasks. Moreover, mixing targets from the other task among the other distractors ensures that the subjects cannot rely on systematic low-level differences between the target and distractor images.

3.3 Intracategory differences

The subdivision of the means-of-transport photographs into the two subcategories 'cars' and 'other vehicles' made before the start of the experiment allows us to investigate intracategory differences in processing. We can compare the car images and the other targets in the transport categorisation task. Cars appeared to lead to a significantly higher proportion of responses with reaction times in the period from 250 – 260 ms to 400 ms (ie a χ^2 test with 1 degree of freedom was significant at the $p < 0.01$ level for every 10 ms period in this interval; see figure 6). Nevertheless, the distributions for cars and other vehicles both significantly differ from the distribution of false positives as early as 240 – 250 ms (χ^2 test with 1 degree of freedom significant at the $p \leqslant 0.001$ level for more than thirty consecutive 10 ms periods). Therefore, both categories can be successfully (ie better than chance) discriminated from distractors in less than 250 ms. Although the categorisation of car images would give a better performance than the categorisation of other vehicles if the analysis was restricted to responses with reaction times below 400 ms, if we take into account all the responses made before the response time limit of 1 s, there is no such advantage: the percentage-correct scores for cars and other vehicles were, respectively, 96.02% and 96.95%, ie virtually equal. The difference observed between cars and other vehicles might result from the fact that cars were more common than other forms of transport. If subjects happened to realise this, then they might, voluntarily or not, bias or prime visual processing to favour features present in photographs of cars. Another related possible explanation of the discrepancy between these two subcategories could be the level
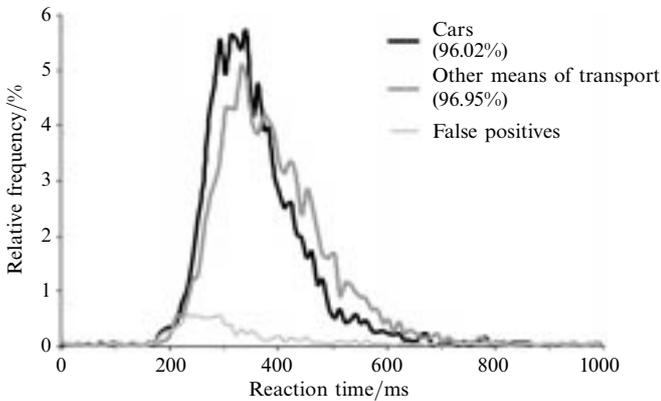
Figure 6. Histograms of reaction times in the vehicle categorisation task. Trials are separated according to the subcategory 'car' or 'other means of transport'. Percentage-correct scores are indicated for each subcategory. False positives are also shown (thin line).

of categorisation, ie basic level for cars versus superordinate level for other vehicles: if the subjects were actually searching for cars more actively than for other vehicles, performance would reflect a basic-level categorisation, even if subjects were instructed to categorise images at the superordinate level (with the target category 'means of transport'). Indeed, numerous experiments have shown an advantage for categorisation of images at the basic versus superordinate level (Rosch et al 1976), although other studies have reported that this advantage could decrease when objects are presented in whole scenes (Murphy and Wisniewski 1989), and ERP recordings have recently shown that this effect was only visible roughly 300 ms after stimulus presentation (Tanaka et al 1999).

4 Discussion

The results obtained in this experiment bring together evidence that human subjects can successfully differentiate complex visual categories in less than 250 ms. Even if median reaction times in the categorisation tasks presented in this study are around 350 ms, there is a significant bias towards correct responses from as early as 240–250 ms, for animal categorisation as well as means-of-transport categorisation. The difference in the distributions of false positives for 'regular' distractors and the distractor images belonging to the target category of the other task also begins shortly before 240 ms. A very plausible explanation of this effect is that enough visual processing is done already to allow distractor images to be rejected (or, equally, targets to be recognised), while the other task performed in alternation biases the responses and makes it more likely for the subject to respond on distractors that would be target trials in the other task. Finally, this remarkable speed of processing is not specific to a basic-level categorisation, because both the animals category and the means-of-transport category are clearly at the superordinate level.

To our knowledge, no other study on the categorisation of natural scenes has ever reported reaction times as short as the ones presented here. The difference between reactions times reported here and those in previous similar experiments (eg Thorpe et al 1996) almost certainly arises from the improvements of the experimental procedure that we described in section 2.1. First, in the present experiment, the dimensions of the images were twice those used previously, thus increasing the amount of energy in the image. Second, we used a touch-sensitive plate which means that the subjects were not required to apply pressure to keep the button pressed, as was the case previously. This certainly means that the motor reaction times can be significantly improved. Even simple reaction times to sinusoidal gratings can be longer than 350 ms depending

on the contrast and spatial frequency of the stimulus, and are rarely found to be shorter than 200 ms (Breitmeyer 1975; Ejima and Ohtani 1987). With some subjects, the detection of a small flashed (50 ms) spot can take as long as 280 ms on average, even when the intensity of the stimulus is 100 times greater than the threshold (Lennie 1981). This reaction time can be reduced to 220–230 ms under gap conditions where the fixation point is turned off shortly before the target spot appears (Fischer and Rogal 1986). It is commonly said that minimal visual reaction times are not shorter than 180 ms, and an average estimated value of 250 ms fits the data from various studies fairly well (Luce 1986). These studies, however, cannot easily be compared to ours because of the simple kind of stimulation that they used. An experiment with natural images as the stimuli triggering detection would allow us to compare our reaction times with simple reaction times obtained under the same conditions.

Such an experiment had been conducted in the laboratory previously to the one reported here. Stimuli were natural images of animals or distractors that could include means of transport, and were presented at half the size used here. The device recording the subjects' response was the same type of touch-sensitive plate that we used. Five subjects from the present study (CHA, LAG, MAR, MAS, PAQ) were asked to perform 3 series of 100 trials where they had to respond to each image whatever the category involved. They performed these series in alternation with 3 series of 100 trials of the animal/nonanimal categorisation task. Their mean reaction time on the latter task was 411 ms (with 92.3% correct responses), and 227 ms on the simple detection task. However, a considerable proportion of the trials in the simple detection series were anticipatory responses, as supported by the fact that in 4% of the trials the response occured before stimulus presentation, and in 2% it occured within 100 ms. Even more strikingly, the two fastest subjects in this task (both with mean reaction times under 200 ms) averaged 12% of anticipatory responses. These anticipations could not be avoided even with randomised (2200 ± 800 ms) interstimulus intervals. Therefore, the reported average reaction time of 227 ms is likely to be underestimated. When comparing this simple reaction time to the ones obtained in our categorisation study, it appears that the average 'cost in reaction time' of processing a complex visual scene is probably around 100–120 ms, and in any case no more than 150 ms. Furthermore, we have shown that the animal and means-of-transport categorisation tasks could be performed accurately (ie above chance) with reaction times below 250 ms, a value surprisingly close to the simple reaction time of 227 ms. Thus, the complex processing required to detect objects belonging to a target visual category can be done almost as fast as a simple perceptual signal detection task. A very tempting interpretation of these results is that categorisation does not occur after perception, but these two mechanisms are performed in parallel, with some or all of the involved processing being shared or common. This theory has already been proposed by Schyns and Oliva (1999) who showed that the categorisation task being performed could strongly modify the underlying perceptual mechanisms. Here we strengthen this argument by demonstrating that reaction times in a high-level visual-categorisation task can be very close to simple visual reaction times.

It is important to stress once again that all reaction times mentioned above include not just visual processing, but also the decision process and the production of the motor output. The time actually required for visual processing by itself might thus be considerably shorter. If we subtract 80–100 ms, eg the minimum time required to generate a 'reaching command' (Kalaska and Crammond 1992), it would appear that the necessary visual mechanisms involved in the categorisation tasks presented here can take no longer than 150 ms. Of course, visual processing does not abruptly stop after 150 ms, but this time of processing appears sufficient for the visual system to successfully discriminate between visual categories. The suggestion that 150 ms is

sufficient for categorisation of previously unseen complex natural scenes is supported by the existence of differential category-specific ERPs in the animal categorisation task that starts at around that latency (Thorpe et al 1996; Fabre-Thorpe et al 2001). Recently, we found that category-specific activation can also be recorded for the vehicle categorisation task at about the same latency (VanRullen and Thorpe, in press), thus confirming the view that processing for both biological and artifactual categories takes roughly the same time.

But the main finding of this study is that the previously obtained results on the speed of visual processing can be generalised from natural to artificial categories. We have shown that the categorisation of means of transport is neither slower in reaction times nor less accurate than the categorisation of animals—a natural, highly biologically relevant category. It is very unlikely that the length of the decision process or the motor command would be significantly different between these two categorisations. Therefore we can easily conclude that visual processing is as fast for man-made categories as it is for natural ones. This clearly rules out the possibility that the ultra-rapid visual categorisation of animals reported previously could be a special case. Rather, it seems likely that similarly rapid go/no-go visual categorisation could be performed for a wide range of categories. Clearly, further experiments will be needed to test the precise limits of this ability. In particular, it will be important to test whether more specific basic-level categories (such as cars) are really detected earlier than more global categories (such as means of transport). The surprisingly good performance and very short reaction times obtained here cast doubt on the intuitive idea that visual processing would require a basic-level identification of the stimulus before its potential superordinate level categorisation (Rosch et al 1976). Indeed, this implies that basic-level categorisation should be performed faster than the kind of categorisation that we report here. While the organisation of the semantic system can certainly explain why naming tasks are performed slower at the superordinate level (Humphreys et al 1999), it seems unlikely however that any visual-processing task requiring a high-level analysis of the visual scene could be performed with reaction times much shorter than the ones that we obtained here for a superordinate categorisation.

A further point that is worth investigating is whether the observed performance can be explained by a simple low-level visual feature or property. For example, it has been shown that pigeons strongly rely on texture information to categorise natural scenes (Troje et al 1999). But that does not mean that texture information is sufficient to perform the task. If a single low-level property was sufficient to characterise distractor images in the two tasks (eg the presence of the sky), and categorisation was based on this property, then including animals in the distractors for the 'means-of-transport' categorisation (and vice versa) should severely impair performance (because this property would not be sufficient to discriminate between animals and vehicles). However, the performances obtained by the subjects in our experiment are if anything better than the ones reported by Thorpe et al (1996) for example. Therefore, distractor images cannot be described by a simple feature. Furthermore, the very wide range of pictures used in each target category effectively rules out the possibility of a single low-level visual dimension accounting for all the variance in these images. In his study involving a 'tree' versus 'non-tree' categorisation, Vogels (1999a, 1999b) systematically investigated the contribution of low-level features such as texture, colour, and size to the categorisation performance. None of these features was found to account for the subjects' performance. Recently, another experiment involving the same sort of rapid categorisation task (animal versus nonanimal, food versus non-food) used here demonstrated that fast visual processing in man and monkeys does not rely on colour cues (Delorme et al 2000). Finally, a recent categorisation study using the same experimental procedure used here showed that the categorisation of simple forms on the basis of a single low-level

feature such as shape could be done 50 ms faster (for mean reaction times) than the animal/nonanimal task (Aubertin et al 1999). If the latter also relied on a single low-level dimension, one would certainly not expect such a discrepancy.

Another related issue is the influence of context on categorisation performance. Although contextual effects cannot be simply ruled out, and certainly take place for some images, these effects alone cannot account for the obtained performance. For example, the target images in figures 2d and 3d were missed by the most subjects, although they present prototypical backgrounds where the presence of a target would be expected (nature scenes for the 'animal' task, street and city scenes for the 'means-of-transport' task). Similarly, many of the distractor images presented in figure 4 represent typical contexts for the presence of an animal or a vehicle. Nevertheless, none of the subjects responded to any of these distractors. Moreover, even if subjects were using the scene context as a cue for categorisation, this would not make the task intrinsically easier: a street scene certainly makes the presence of a vehicle more likely, but categorising a photograph as a street scene also constitutes a rather challenging task for the visual system.

The fact that the visual mechanisms described here are too complex to rely on a simple stimulus-feature estimation does not mean, of course, that it cannot be based on a combination of many such features. A recent neural-network model (Campbell et al 1997) demonstrated that it is possible to classify complex outdoor scenes on the basis of a set of low-level features.

Finally, the implications of the present findings for current computational theories of visual processing need to be stressed. We have demonstrated that, even in the case of a demanding visual-categorisation task requiring visual processing that would severely challenge the most sophisticated artificial vision systems, the entire input – output visuo-motor sequence can be completed in under 250 ms. Given the delays in the motor pathways, this means that, at least for those images that were correctly categorised in around 250 ms, the underlying visual processing has been done in not much more than 150 ms, a value that fits well with the latencies for differential ERP activity seen in the same sort of task (Thorpe et al 1996). Few if any of the currently available models of visual processing are compatible with such constraints. Given the number of stages involved in visual processing and the mean firing frequencies of neurons in the visual system, this very short time window implies that most of the processing has to be done in a feed-forward way, with probably no more than one spike emitted by each neuron (Thorpe and Imbert 1989). Although these limitations seem to strongly constrain neural-network models of visual processing rather than to provide a way of improving them, recent simulations have demonstrated the power of neural coding schemes based on the temporal information transmitted by single spikes (Delorme et al 1999). For example, one such model was able to detect human faces in natural images with a much higher level of performance than classical models of face processing (VanRullen et al 1998). We believe that temporal constraints on visual processing, as highlighted by psychological studies on human subjects as well as electrophysiological recordings in animals, constitute the basis of a vivid framework in which models can be developed that could eventually approach human performance.

References

Aubertin A, Fabre-Thorpe M, Fabre N, Geraud G, 1999 "Fast visual categorization and speed of processing in migraine" *Comptes Rendus de l'Académie des Sciences, Série III* **322** 695 – 704

Breitmeyer B G, 1975 "Simple reaction time as a measure of the temporal response properties of transient and sustained channels" *Vision Research* **15** 1411 – 1412

Campbell N W, Thomas B T, Troscianko T, 1997 "Automatic segmentation and classification of outdoor images using neural networks" *International Journal of Neural Systems* **8** 137 – 144

Delorme A, Richard G, Fabre-Thorpe M, 2000 "Rapid categorisation of natural scenes is colour blind: A study in monkeys and humans" *Vision Research* **40** 2187 – 2200

Delorme A, VanRullen R, Gautrais J, Thorpe S J, 1999 "SpikeNET: a simulator for modeling large networks of integrate and fire neurons" *Neurocomputing* **26 – 27** 989 – 996

Ejima Y, Ohtani Y, 1987 "Simple reaction time to sinusoidal grating and perceptual integration time: contributions of perceptual and response processes" *Vision Research* **27** 269 – 276

Fabre-Thorpe M, Delorme A, Marlot C, Thorpe S, 2001 "A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes" *Journal of Cognitive Neuroscience* **13** 171 – 180

Fabre-Thorpe M, Fize D, Richard G, Thorpe S J, 1998a "Rapid categorization of extrafoveal natural images: Implications for biological models", in *Computational Neuroscience: Trends in Research* Ed. J Bower (New York: Plenum Press) pp 7 – 12

Fabre-Thorpe M, Richard G, Thorpe S J, 1998b "Rapid categorization of natural images by rhesus monkeys" *Neuroreport* **9** 303 – 308

Fischer B, Rogal L, 1986 "Eye – hand-coordination in man: a reaction time study" *Biological Cybernetics* **55** 253 – 261

Humphreys G W, Price C J, Riddoch M J, 1999 "From objects to names: a cognitive neuroscience approach" *Psychological Research* **62** 118 – 130

Kalaska J F, Crammond D J, 1992 "Cerebral cortical mechanisms of reaching movements" *Science* **255** 1517 – 1523

Lennie P, 1981 "The physiological basis of variations in visual latency" *Vision Research* **21** 815 – 824

Luce R D, 1986 *Response Times* (Oxford: Oxford University Press)

Murphy G L, Wisniewski E J, 1989 "Categorizing objects in isolation and in scenes: what a superordinate is good for" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **15** 572 – 586

Oram M W, Perrett D I, 1992 "Time course of neural responses discriminating different views of the face and head" *Journal of Neurophysiology* **68** 70 – 84

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

Rosch E, Mervis C B, Gray W D, Johnson D M, Boyes-Braem P, 1976 "Basic objects in natural categories" *Cognitive Psychology* **8** 382 – 439

Schyns P G, Oliva A, 1999 "Dr Angry and Mr Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations" *Cognition* **69** 243 – 265

Tanaka J, Luu P, Weisbrod M, Kiefer M, 1999 "Tracking the time course of object categorization using event-related potentials" *Neuroreport* **10** 829 – 835

Thorpe S J, Fize D, Marlot C, 1996 "Speed of processing in the human visual system" *Nature* **381** 520 – 522

Thorpe S J, Gegenfurtner K, Fabre-Thorpe M, Bülthoff H H, 1999 "Categorisation of complex natural images in extreme peripheral vision" *Perception* **28** Supplement, 61

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

Troje N F, Huber L, Loidolt M, Aust U, Fieder M, 1999 "Categorical learning in pigeons: the role of texture and shape in complex static stimuli" *Vision Research* **39** 353 – 366

VanRullen R, Gautrais J, Delorme A, Thorpe S J, 1998 "Face processing using one spike per neuron" *Biosystems* **48** 229 – 239

VanRullen R, Thorpe S J, 2001 "The time course of visual processing: from early perception to decision making" *Journal of Cognitive Neuroscience* in 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