

A Limit to the Speed of Processing in Ultra-Rapid Visual Categorization of Novel Natural Scenes

Michèle Fabre-Thorpe, Arnaud Delorme, Catherine Marlot, and Simon Thorpe

Abstract

■ The processing required to decide whether a briefly flashed natural scene contains an animal can be achieved in 150 msec (Thorpe, Fize, & Marlot, 1996). Here we report that extensive training with a subset of photographs over a 3-week period failed to increase the speed of the processing underlying such rapid visual categorizations: Completely novel scenes could be categorized just as fast as highly familiar ones. Such data imply that the visual system processes new stimuli at a speed and

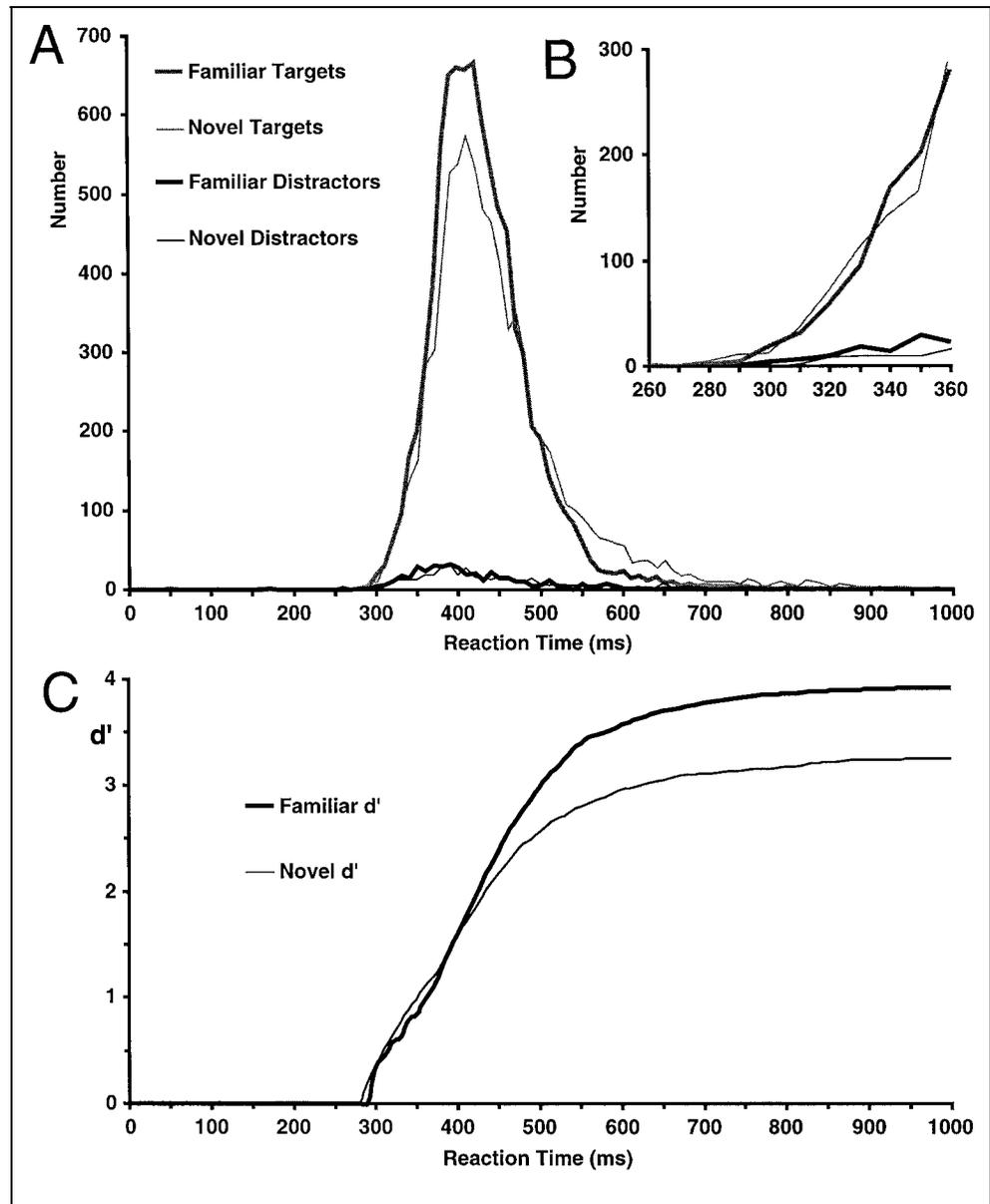
with a number of stages that cannot be compressed. This rapid processing mode was seen with a wide range of visual complex images challenging the idea that short reaction times can only be seen with simple visual stimuli and implying that highly automatic feed-forward mechanisms underlie a far greater proportion of the sophisticated image analysis needed for everyday vision than is generally assumed. ■

Both humans and monkeys are able to categorize natural images accurately and very rapidly (Fabre-Thorpe, Richard, & Thorpe, 1998; Thorpe, Fize, & Marlot, 1996). The nature of the underlying mechanisms is currently the subject of intense debate (Wallis & Bülthoff, 1999; Treisman & Kanwisher, 1998; Biederman & Kalocsai, 1997; Tanaka, 1997; Logothetis & Sheinberg, 1996), but it is generally believed that this sort of object recognition must involve bidirectional information processing, in which bottom-up information originating in the retina interacts with feedback mechanisms in a system preset by top-down knowledge (Rao & Ballard, 1999; Sastry, Shah, Singh, & Unnikrishnan, 1999; Mesulam, 1998; Przybyszewski, 1998; Vecera & O'Reilly, 1998; Frith & Dolan, 1997; Grossberg, Mingolla, & Ross, 1997; Humphreys, Riddoch, & Price, 1997; Kersten, 1997; Tallon-Baudry, Bertrand, Delpuech, & Permier, 1997; Ullman, 1996; Kosslyn, 1994). Indeed, recent neurophysiological studies have demonstrated the importance of lateral and feedback interactions in fundamental visual processes that include texture segmentation and figure-ground effects (Hupé et al., 1998; Lamme, Super, & Spekreijse, 1998) as well as visual attention (Desimone, 1998; Hillyard, Teder-Salejarvi, & Munte, 1998; Roelfsema, Lamme, & Spekreijse, 1998; Vidyasagar, 1998; Duncan, Humphreys, & Ward, 1997; Luck, Chelazzi, Hillyard, & Desimone, 1997).

On the other hand, it is likely that at least some forms of visual processing can be achieved on the basis of purely feed-forward mechanisms (Thorpe & Imbert, 1989), although it is a difficult experimental issue to determine the relative importance of bottom-up and top-down effects. One potential hallmark of relatively automatic feed-forward processing mechanisms is that their time course should be relatively fixed. The argument supposes that when processing involves extensive use of bidirectional interactions between bottom-up and top-down mechanisms one should predict that processing time will depend heavily on experience, and indeed, this is very often the case. Experiments on visual priming have demonstrated that, in many tasks, behavioral reaction times are reduced by experience with particular stimuli (Schacter & Buckner, 1998; Wiggs & Martin, 1998; Warren & Morton, 1982), and there is clear evidence that training can influence visual processing, even at relatively early stages in the visual system (Ahissar & Hochstein, 1997; Dolan et al., 1997; Karni & Bertini, 1997; Tovee, Rolls, & Ramachandran, 1996). Furthermore, there have been a number of recent reports showing that stimulus repetition can have pronounced effects on the event-related potentials (ERPs) to a variety of visual stimuli, and that in some cases these effects appear to occur at remarkably short latencies (Debruille, Guillem, & Renault, 1998; George, Jemel, Fiori, & Renault, 1997; Seeck et al., 1997; Begleiter, Porjesz, & Wang, 1995). Note that the existence of such familiarity effects is not enough to distinguish feed-forward and feedback mechanisms, because effects of familiarity can arise even in feed-forward models

Centre de Recherche Cerveau and Cognition (UMR 5549, CNRS-UPS), Faculté de Médecine de Rangueil, Toulouse, France

Figure 1. (A) Reaction time distributions for novel (thin lines) and familiar (thick lines) stimuli. The values give the number of trials in a 10-msec bin. Gray curves show the data for correct go responses to target stimuli, whereas the black lines provide data for incorrect go responses to distractors. (B) The inset shows a higher magnification view of the reaction time distributions for the critical region between 260 and 360 msec. Note that (i) the curves are effectively identical for novel and familiar targets, and (ii) that there is a strong bias in favor of responding to targets from very early on, thus ruling out the possibility that the subjects were making anticipatory responses. (C) A plot of how the sensitivity index d' improves as a function of time for familiar and novel stimuli. The d' was calculated from the formula $d' = z_n - z_s$, where z_n is chosen such that the area of the normal distribution above that value is equal to the false-alarm rate, and where z_s is chosen to match the hit rate. While it is clear that d' is higher for familiar stimuli if we take into account *all* the behavioral responses, no advantage is seen before 400 msec (36.1% of the responses).



(Mozer, 1991). On the other hand, if familiarity has no effect on processing speed, only preoptimized and largely hard-wired feedback mechanisms would be compatible.

In this study, we have looked at the effects of familiarity on processing speed and accuracy in a go/no-go visual categorization task. When natural images that they have never seen before are flashed for only 20 msec, humans can detect the presence of an animal with high accuracy (94% correct) and with a median reaction time of 445 msec (Thorpe et al., 1996). Moreover, the associated ERPs show differential cerebral activity between target and nontarget trials that develops from 150 msec after stimulus onset, suggesting that the visual processing in such a complex categorization could be mainly based on feed-forward mechanisms. In the present study, we investigated whether repetitive presentations

of the stimuli during 14 days spread over 3 weeks, and thus history and knowledge, would speed up the information processing in this fast go/no-go visual categorization task.

RESULTS

The task required the subjects to respond by releasing a button whenever a briefly flashed image contained a living animal. A set of 2000 color photographs was selected from a commercial photo library of which 50% were targets showing one or more animals in their natural environments. They were extremely varied and included mammals, birds, fish, reptiles, insects, arthropods, and crustaceans. The remaining pictures were distractors and also extremely varied, including not just pictures of natural scenes (landscapes, fruits, flowers,

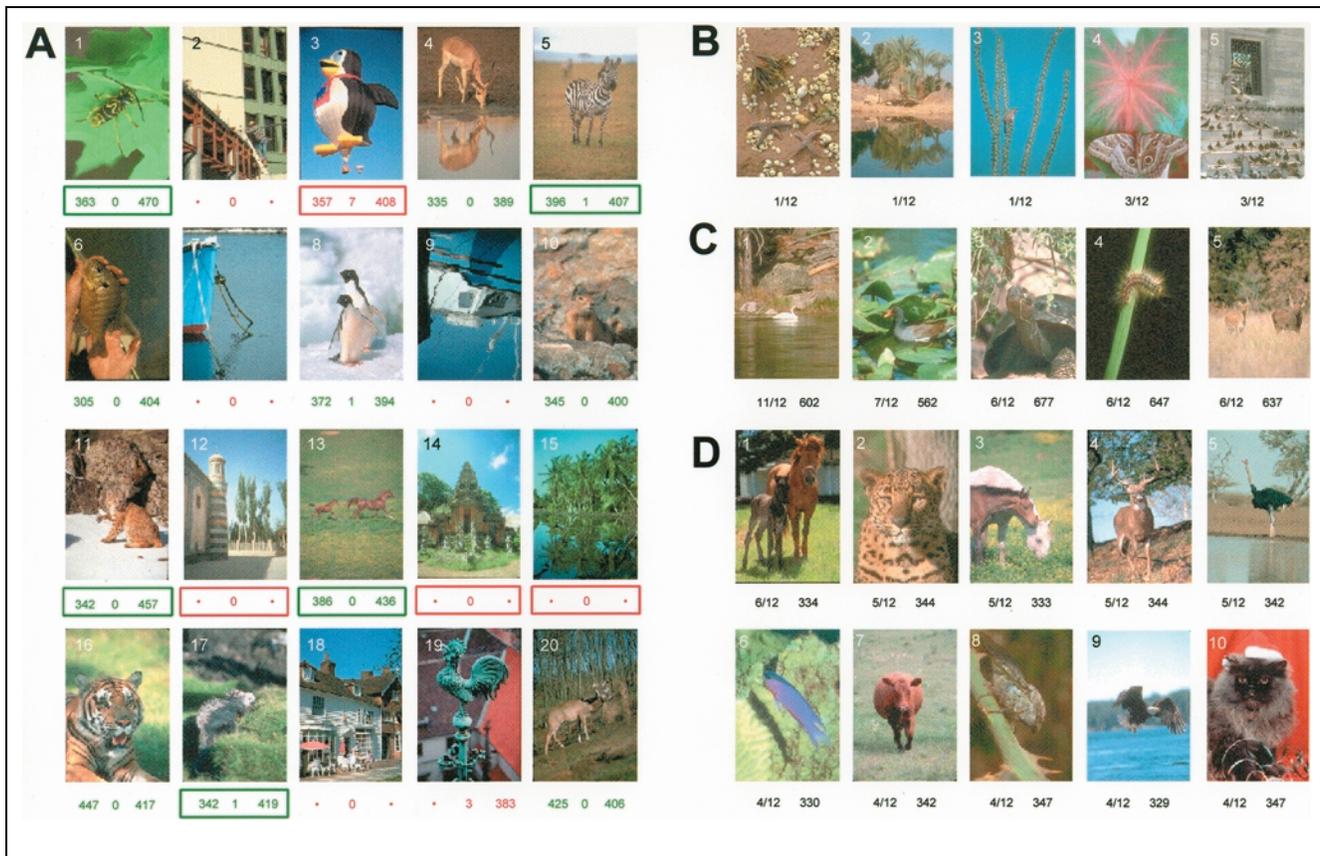


Figure 2. (A) A representative sequence of 20 images taken from one subject (RvR). Targets are shown with green text, distractors in red, and a box surrounds the text for images that were familiar to RvR. The left number gives the reaction time for RvR on trials where he responded. The center number gives the number of errors made by the 12 subjects for whom the stimulus was novel (for the other two subjects, the image was familiar), and the right-hand number gives the mean reaction time for those 12 subjects. Note that all 12 subjects to whom they were presented as novel successfully detected the targets in images 1, 4, 6, 10, 11, 13, 16, and 20. The only targets that posed a problem were images 5, 8, and 17 that were each missed by one of the 12 subjects. Similarly, the distractor stimuli in images 2, 4, 6, 10, 11, 13, 16, and 20 were successfully ignored by all the subjects. Only images 3 and 19, which were “trick” distractors provoked errors — 7 out of 12 in the case of image 3, and 3 out of 12 for image 19. (B) The five images that resulted in the highest number of errors. Below each image is indicated the number of subjects (out of 12) that correctly detected the target. (C) The five images that were correctly detected by all the subjects but that resulted in the highest number of particularly long reaction times (over 505 msec). Below each image is indicated: on the left, the number of subjects (out of 12) with reaction times above 505 msec, and on the right, the mean reaction time for these subjects. (D) A set of images that produced the highest proportion of very short (< 360 ms) reaction times in the 12 subjects for whom they were completely novel. Below each image is indicated: on the left, the number of subjects (out of 12) with reaction times below 360 msec, and on the right the mean reaction time for these subjects. Images 1–5 show the subset of images for which five or six subjects (out of 12) had very short RTs; images 6–10 are a representative subset of the 18 images for which four subjects (out of 12) had very short RTs.

etc.), but also a range of man-made environments. As a test of the sophistication of the decision strategies used by the subjects, human beings together with a number of “trick” images of animal-like objects (paintings, embroideries, statues, balloons, etc.) were used as distractors. During the first part of the experiment, each subject was trained on 14 separate days spread over 3 weeks with a subset of 200 randomly selected images. Performance improved significantly during this period, both in terms of accuracy (94% on day 1, 97% on day 14) and reaction time, which decreased from an initial mean value of 480 msec (median RT = 467 msec) to 436 msec (median RT = 420 msec) on day 14. On the two subsequent days, the familiar stimuli were randomly mixed with 1200 completely novel ones and testing was combined with ERP recordings. Behavioral and

electrophysiological responses with novel and familiar stimuli were directly compared.

The behavioral results showed that accuracy with familiar stimuli was higher than for the novel ones (96.9% correct vs. 94.7%) This improvement was mainly the result of a higher success rate with some of the more difficult targets (1.2% of the familiar targets were missed compared with 6.3% of the novel targets), and mean reaction time was also significantly shorter for familiar than for novel targets (424 msec for familiar vs. 444 msec for novel, $p < .0001$ paired t test, $df = 13$). However, analysis of the reaction time distributions plotted in Figure 1 reveals that this apparent increase in processing speed is virtually entirely due to the elimination of long reaction time responses to some of the familiar targets that were initially difficult to

Table 1. Distribution of Slow and Fast Behavioral Responses to Novel Targets

Number of subjects out of 12	Image distribution		
	Observed RT < 361 msec	Expected with $p = .118$	Observed RT > 505 msec
0–3	488	488	463
4	18	18	27
5	4	4	10
6	1	1	9
7	0	0	1
8–10	0	0	0
11	0	0	1
Total	511	511	511

The analysis was restricted to the 511 novel targets that were correctly identified by all 12 subjects so that each target had induced the same number of behavioral go responses (a total of $12 \times 511 = 6132$ behavioral responses). Of these, 724 were triggered with RTs of 360 msec or less (a limit below which familiarity effects could not be seen); they were compared to the similar number of responses (722) that were triggered with abnormally long RTs (505 msec or more). The table shows how many of the 511 targets were detected by k subjects (from 0 to 11) with either a short RT (left-hand column) or a long RT (right-hand column). For comparison, the central column gives the expected number of images for which k of the 12 subjects would have had short RTs if the reaction times were totally random [binomial distribution with $p(\text{RT} = 360 \text{ msec})$ equal to $724/6132 = 0.118$]. The distribution seen for short RTs (left-hand column) fits exactly the prediction of a random distribution, which means that there was no evidence to support the view that certain images were associated with particularly fast responses. In contrast, the distribution of images with long RTs is heavily skewed (right-hand column). For example, one image (see Figure 2C, image 1) produced RTs of over 505 msec in 11 of the 12 subjects. This is strong evidence that even when we only take into consideration images that were perfectly identified, a particular subset of these targets was especially hard to detect. Note that this analysis does not include even more difficult targets that were missed by one or more subjects (see Figure 2B).

detect. In contrast, the initial parts of the two distributions that correspond to the earliest reaction times can be superimposed, showing that processing was no faster for the highly familiar images. In fact, the tenth percentile point of the reaction time distribution is identical for novel and familiar natural scenes; 10.5% of responses to novel targets had reaction times of 360 msec or less, whereas the corresponding figure for familiar targets was 10.4%. Furthermore, by using a measure of d' to assess sensitivity, we found that only after 400 msec was there any evidence that performance was actually better for familiar stimuli (Figure 2C). The reason is that while the probability of a hit (correct detection of a target) earlier than 400 msec was higher for familiar targets than novel ones (.372 vs. .312), so was the false alarm rate (.027 for familiar distractors, and .019 for novel ones). Thus, over a third of the images used in the present study was processed

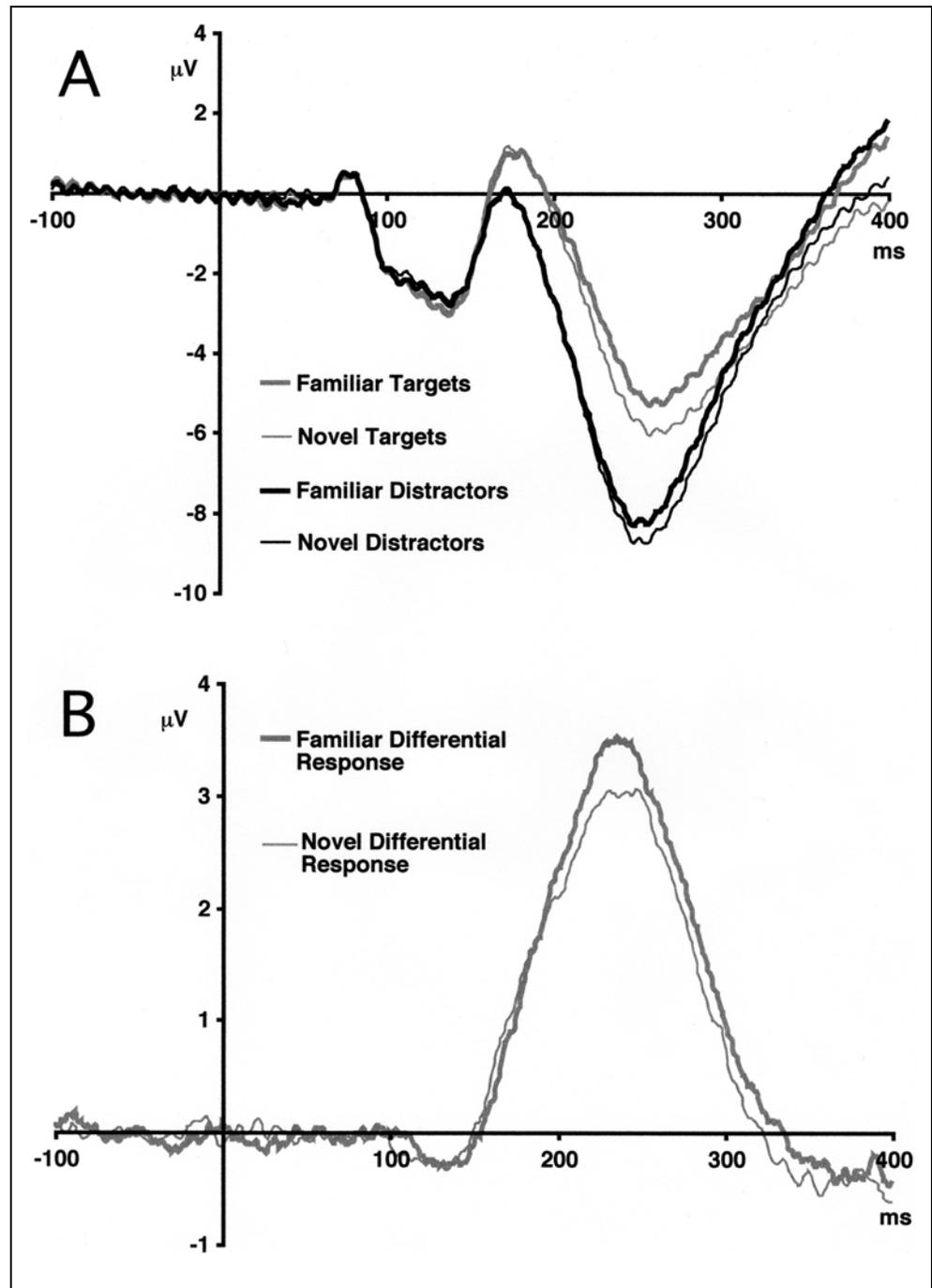
so quickly that an improvement in performance with familiarity could not be seen.

Note that the existence of such short reaction times cannot be the result of subjects making random anticipations. If subjects were just guessing, they could not do better than 50% correct, since the probability of target and distractor trials was equal. However, the comparison of the reaction time distributions for correct “go” responses on targets and incorrect “go” responses on distractor trials clearly shows that even the shortest reaction times (those between 280 and 300 msec) are very strongly and statistically significantly biased in favor of targets. Thus, even such fast responses allow enough time for both visual processing and response execution.

The sequence shown in Figure 2A illustrates the wide range of images used in these experiments, and demonstrates the remarkable sophistication of the decision strategies required to perform the task. Despite this great variety, when presented as novel, 511 of the 700 targets were correctly detected by all 12 subjects. Similarly, 533 of the 700 distractors seen as novel were correctly ignored by all 12 subjects. Errors, whether for novel targets or distractors, tended to be concentrated on a relatively small subset of the images in that over half of them (53.7%) were produced by only 5.5% of the images. Indeed, behavioral responses to this particular subset of targets were also characterized by particularly long reaction times (mean RT, 557 msec). One may wonder what factors make certain target images difficult to detect. To investigate this issue we examined all the stimuli that produced either a high proportion of misses, or an abnormally large number of “go” responses with long reaction times. Figure 2B shows five of the most difficult images as defined by the error rates, and Figure 2C shows the five images that, although correctly categorized by all 12 subjects, had the highest proportion of long reaction times. Note that with 3 weeks of training these same targets could often be accurately detected (89% correct), with a reduced mean reaction time (470 msec) and that the elimination of these long latency responses accounts for much of the training-related changes in the RT distributions seen in Figure 1. Various reasons could account for the difficulty in spotting the animals in such images. They include small target size in the scene (B2, B3, C1), the presence of several small targets (B5), the unusual or ambiguous aspect of some animals (B1, C3), targets for which contours are difficult to extract (B3, C1, C2), and images in which the animal is not the main object (B2, B4, B5). Further analysis would be required to pinpoint all the factors that make certain images difficult to analyze.

Given that a small subset of the images was particularly difficult to categorize, it is natural to suppose that there might be another subset that was particularly easy to process. In such a case, the lack of improvement seen

Figure 3. (A) Grand average evoked potentials obtained from electrode Fz using 14 subjects. Separate curves are shown for familiar targets and distractors, and for novel targets and distractors. Note that familiarity has essentially no effect on either set of curves until at least 200 msec poststimulus. (B) Plots of the evolution of the difference between the response to targets and distractors separately for familiar and novel stimuli. Note that the onset and slope of the differential responses are essentially identical for novel and familiar stimuli. In addition, there is a small differential response of opposite polarity that precedes the 150-msec deflection. This difference is statistically significant for the time steps 113 to 136 msec (two-tailed t test, $t > 2.65$, $df = 13$, $p < .02$).



for the most rapidly categorized images following training could be explained by the existence of a floor effect: No further improvement was possible because the processing required was already so trivial. To test this possibility, we looked at the images for which the proportion of subjects that responded with particularly short reaction times was unusually high. Of the 700 novel targets, 23 images induced an RT < 360 msec for four or more subjects (all were correctly detected by all 12 subjects). Ten of them are illustrated in Figure 2D. Intriguingly, unlike the difficult images, there seems to be little to distinguish them from the rest of the set.

They are highly varied, including not only mammals, but also birds, fish, and insects. Furthermore, they include examples that, by almost any criterion, would be considered very challenging for current models of visual processing (noncanonical views, overlapping objects, low-contrast images, etc.). This makes it very difficult to argue that the lack of an improvement in rapid processing could be explained by supposing that these images were abnormally easy to process.

This clear contrast between “easy” and “hard” targets is made even more obvious in Table 1 that gives the numbers of images for which four or more subjects had

abnormally long or short reaction times. While the distribution for long RTs clearly points to a subset of particularly difficult images, the distribution for short RTs is exactly what would be expected on the basis of a purely random distribution.

Further evidence for the lack of an early effect of familiarity comes from the analysis of the differential ERPs between target and distractor trials that were recorded in parallel during the final 2 days of testing. As in the previous study (Thorpe et al., 1996), we found a very strong and robust difference in the frontally recorded evoked potentials (averaging together all the waveforms obtained either on correct target and distractor trials) that developed from 150 msec after stimulus onset. Indeed, the strong differential reaction at 150 msec is actually preceded by a deflection in the opposite direction that becomes statistically significant at around 120 msec. Differential event-related electrical and magnetic responses of this type have now been seen by a number of authors in relation to higher order visual processing (Eimer, 1998; Linkenkaer-Hansen et al., 1998; Schendan, Ganis, & Kutas, 1998; Bentin, Allison, Puce, Perez, & McCarthy, 1996). Surprisingly, there was no difference in either the onset latency of the differential activity or its slope for familiar and for novel images (see Figure 3), although it is clear that (as in a number of other studies, Buckner et al., 1998; Hertz, Porjesz, Begleiter, & Chorlian, 1994) the responses to familiar stimuli were in general somewhat smaller than with novel ones. On the other hand, there was evidence that the amplitude of the differential response between target and distractor trials was slightly larger in amplitude with familiar stimuli, but this effect was not evident until 30–50 msec after the start of the differential response.

DISCUSSION

The present results argue for the existence of a specific mode of visual processing that cannot be speeded up by extensive training; a mode that we term ultrarapid visual categorization. It may well be that this very rapid processing mode is used in many everyday visual tasks, but it is only when subjects are forced to respond very quickly (as in the current task) that such processing can be revealed at the behavioral level. Both the reaction time data and the differential ERP effects indicate that the underlying visual processing can be just as fast with completely novel images as it can be for images with which the subjects are highly familiar. The existence of such a floor effect is certainly a surprising finding, given that familiarity normally facilitates processing, as shown by the extensive literature on the effects of priming. Although we are not claiming that priming did not occur, since we did indeed see a clear 20-msec decrease in mean reaction time for familiar images that was associated with a 2.2% increase in accuracy, we would

claim that priming has no effect on either the shortest behavioral reaction times (i.e., those under 360 msec) or the onset latency of the differential category-specific ERP signal.

It appears, therefore, that even extensive training with particular stimuli cannot result in significantly faster processing than that seen with totally novel images. This appears to be in contradiction with the generally accepted view that visual processing involves extensive bidirectional mechanisms based on contextual knowledge. Rather, it implies that even a challenging visual task of the sort used here can be performed using highly automatic and data-driven routines, probably involving extensive use of feed-forward mechanisms. The processing complexity needed for the classification of some of the stimuli (humans from other animals, or living animals from artificial ones) also implies that these feed-forward mechanisms can be much more sophisticated than previously thought.

There is, in fact, a range of arguments in favor of the view that some types of processing can indeed be achieved using feed-forward mechanisms (Thorpe et al., 1989). One argument comes from studies that have demonstrated that many neurons are already highly selective even at the very beginning of the visual response, a result that suggests that iterative feedback mechanisms are not required to build up the response selectivity (Rolls & Tovee, 1994; Celebrini, Thorpe, Trotter, & Imbert, 1993; Oram & Perrett, 1992). For example, this has been shown for face-selective neurons in the monkey temporal lobe that start firing 100 msec or less after the onset of a visual stimulus. Such data need to be considered in relation to the remarkably short behavioral reaction times that we have seen in the present task and that cannot be explained by random anticipations. In humans, such early responses start at around 280 msec, but in monkeys, the input–output sequence linking the retina to the hand muscles can take as little as 180 msec (Fabre-Thorpe et al., 1998). This is only 80 or so milliseconds after the onset of firing of neurons in the inferotemporal cortex (Perrett, Rolls, & Caan, 1982), a structure that is almost certainly needed to perform this sort of task. However, the inferotemporal cortex has no direct outputs to the motor system, and so information probably needs to be sent via relays in the prefrontal, premotor, and motor cortex before being sent down the pyramidal tract to the motoneurons in the spinal cord. This leaves very little time to do anything other than a feed-forward pass.

A second argument comes from the finding that, in this same rapid visual categorization task, the removal of chromatic information from such images has essentially no effect on performance, either in monkeys, or in humans with fast reaction times (Delorme, Richard, & Fabre-Thorpe, 1999). One interpretation of this result depends on the fact that luminance-based information

in the magnocellular pathways reaches visual cortex slightly before the chromatic input from the parvocellular pathways (Nowak & Bullier, 1997). In this case, any processing or behavior relying on just the initial part of the response in each cortical area and using feed-forward processing would not be influenced by chromatic information. Rather it would be dominated by relatively coarse luminance-based magnocellular information as has been suggested by experiments on coarse to fine visual processing (Schyns & Oliva, 1994; Parker, Lishman, & Hughes, 1992).

A third argument comes from the existence of neural network models using only feed-forward mechanisms that are able to perform nontrivial visual processing tasks (Rolls & Treves, 1998; Mel, 1997; Wallis & Rolls, 1997). For example, Van Rullen et al. recently demonstrated that a simple multilayer feed-forward network of asynchronously spiking neurons can be trained to locate faces in natural images (Van Rullen, Gautrais, Delorme, & Thorpe, 1998).

Finally, recent fMRI and ERP data have demonstrated that even visual stimuli that have been masked so that the subject is unaware of their presence are capable of significantly modulating activity in the motor system (Dehaene et al., 1998). There is therefore reason to believe that sophisticated visual processing is indeed possible using rapid feed-forward mechanisms, and it makes sense from an evolutionary point of view to use feed-forward mechanisms whenever possible.

While few would argue with the suggestion that at least some information can be extracted from an image on the basis of the first wave of information passing through the visual system, the present data suggest that the complexity of the analysis, which can be achieved using such mechanisms, is substantially greater than has been suspected in the past. One particularly important issue concerns the question of image segmentation, since it is generally assumed that cluttered natural images of the type used in the present experiments must first be correctly segmented before object recognition is possible. Recent neurophysiological experiments on texture segmentation and figure-ground effects imply that this is a time-consuming process that involves feedback circuits (Lamme et al., 1998), so it is difficult to see how such a complex process could be achieved in a feed-forward network. On the other hand, the present results suggest that the visual processing needed to detect the presence of an animal in a natural scene (and trigger a behavioral response) may be possible without having first to complete these relatively time-consuming low-level processes. It could be that detecting an animal in a natural scene can be done with a massively parallel search for a large number of component features, some of which may be sufficient on their own to indicate—with a high probability—the presence of an animal. For example, the presence of an

eye, a fin, a claw, or teeth would all be sufficient to trigger a behavioral response in an animal-detection task, even under conditions where the image has not yet been completely segmented. In this respect, it may be worth noting that top-down modulation of visual processing may well play a vital role in presetting the visual system in such a way that neuronal mechanisms sensitive to components of animals are primed. Note, however, that this must be done before the start of a trial.

Whereas all these arguments support the view that feed-forward processing could play a critical role in the present task, it is certainly not possible to rule out the use of feedback loops. Certain critical steps in visual processing may involve such mechanisms, but if this is the case, the current evidence provides a number of serious constraints. First, their number must be very restricted in order to be compatible with the short behavioral reaction times observed both in the present study and in previous ones (Fabre-Thorpe et al., 1998; Thorpe et al., 1996). Second, the use of feedback must presumably be restricted to relatively hard-wired and preoptimized processing that does not depend on specific knowledge about particular images.

The present data support the idea that categorization can involve two different types of mechanism that have different time courses. Ultrarapid visual categorization would be the result of an initial wave of fast feed-forward visual processing that can quickly trigger behavioral responses. In our task, many targets seem to be detected that way, but to deal with particularly difficult stimuli, additional time-consuming processing would be necessary. In everyday life, when constraints in response speed are not so high and accuracy is at a premium, most responses will only be triggered once analysis is sufficiently detailed. This additional processing would almost certainly involve extensive top-down and interactive mechanisms, and would be expected to be highly subject to the effects of training. Indeed, this is precisely what was observed here. Extensive training certainly did improve performance, but only on the (relatively limited) subset of targets that were initially difficult to detect. It seems likely that such images may be impossible to categorize on the basis of the first wave of visual information and require a longer processing phase with more detailed or more specific analysis. Such two-stage visual processing has recently been shown for the analysis of faces (Sugase, Yamane, Ueno, & Kawano, 1999) and is thought to require a number of iterative processes (Oram & Richmond, 1999).

It should be stressed that even if very rapid and automatic unidirectional mechanisms are often sufficient to trigger a response in our task, there is certainly no reason to believe that this would signal the end of visual processing. Indeed, there is actually no good reason to suppose that the ability of subjects

to perform this challenging task involves conscious visual perception at all. When presented with a flashed photograph of the type used in the current experiments, subjects report that the first thing they perceive is a fully segmented scene in which the relations of all the objects in the image are clear. However, this end result may involve a much more complex process than that required to trigger (or not) the behavioral response in the task used here. Furthermore, the rapid detection of animal features using fast feed-forward mechanisms could be used to help the segregation process occurring at lower levels in the visual system via feedback connections. The precise limits of what can be achieved by the first feed-forward pass through the visual system is clearly something that will need a great deal of further research, but it is already clear that the power of this type of processing has been seriously underestimated in the past.

METHODS

Subjects

Fourteen subjects participated in the study (seven males, seven females, mean age 30.5 years, range 21–46). They were divided into seven pairs (one male, one female). A total of 1400 images were used in this experiment, and each pair of subjects was assigned a different subset of 200 images on which performance was tested Monday to Friday for 3 weeks. During the testing session, the subjects were seated 1 m from a color monitor in a dimly lit room. They were required to start each trial by pressing a button. After a short delay, an eight-bit color photograph (256 pixels wide by 384 pixels high) was flashed for 20 msec using a programmable graphics board (VSG 2.1, Cambridge Research Systems) and the subject was required to respond by releasing the button if the image contained an animal. On distractor trials, the subjects had to keep pressing the button for at least 1 sec. Trials were grouped in blocks of 100, with 50 targets and 50 distractors shuffled at random before the start of each block. The 20-msec presentation prevented any exploratory eye movement.

Behavioral Procedure and ERP Recordings

On the fourteenth and fifteenth days of testing (Thursday and Friday of week 3) testing was associated with EEG recording. The training session with the 200 familiar images was immediately followed by a testing session in which the familiar images were mixed three times with 600 of the remaining 1200 novel images. Each day's testing involved 12 blocks of 100 trials, each of which contained 50 of the familiar stimuli together with 50 of the novel images, chosen at random. Over the course of the final two days, each subject was tested on a total of

2400 trials, 1200 novel stimuli, and six repetitions of each of the 200 familiar ones. By assigning different sets of 200 stimuli for training with each of the seven pairs of subjects, we ensured that all 1400 images were used equally often as familiar and novel stimuli. The brain electrical activity was recorded using a 32-channel SynAmps amplification system (Neuroscan Inc.) using an sampling rate of 1 kHz and linked ears as the reference. Potentials on each trial were baseline corrected using the signal during the 100 msec that preceded the onset of the stimulus. Trials with artifacts related to ocular movements and blinks were excluded from the analysis, as were trials contaminated by strong alpha frequency activity. As in the previous study by Thorpe et al., comparisons between responses on target and distractor trials were made using paired *t* test.

Acknowledgments

This work was supported by the CNRS, the Cognisciences Program, and the Midi-Pyrénées Region. The experimental procedures used were authorized by the local ethical committee (CCPPRB No. 9614003).

REFERENCES

- Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*, 401–406.
- Begleiter, H., Porjesz, B., & Wang, W. (1995). Event-related brain potentials differentiate priming and recognition to familiar and unfamiliar faces. *Electroencephalography and Clinical Neurophysiology*, *94*, 41–49.
- Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*, 551–565.
- Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *352*, 1203–1219.
- Buckner, R. L., Goodman, J., Burock, M., Rotte, M., Koutstaal, W., Schacter, D., Rosen, B., & Dale, A. M. (1998). Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. *Neuron*, *20*, 285–296.
- Celebrini, S., Thorpe, S., Trotter, Y., & Imbert, M. (1993). Dynamics of orientation coding in area V1 of the awake monkey. *Visual Neuroscience*, *10*, 811–825.
- Debruille, J. B., Guillem, F., & Renault, B. (1998). ERPs and chronometry of face recognition: Following-up Seeck et al. and George et al. *NeuroReport*, *9*, 3349–3353.
- Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P. F., & Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*, 597–600.
- Delorme, A., Richard, G., & Fabre-Thorpe, M. (1999). Rapid processing of complex natural scenes: A role for the magnocellular visual pathways? *Neurocomputing*, *26–27*, 663–670.
- Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *353*, 1245–1255.

- Dolan, R. J., Fink, G. R., Rolls, E., Booth, M., Holmes, A., Frackowiak, R. S. J., & Friston, K. J. (1997). How the brain learns to see objects and faces in an impoverished context. *Nature*, *389*, 596–599.
- Duncan, J., Humphreys, G., & Ward, R. (1997). Competitive brain activity in visual attention. *Current Opinion in Neurobiology*, *7*, 255–261.
- Eimer, M. (1998). Does the face-specific N170 component reflect the activity of a specialized eye processor? *NeuroReport*, *9*, 2945–2948.
- Fabre-Thorpe, M., Richard, G., & Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. *NeuroReport*, *9*, 303–308.
- Frith, C., & Dolan, R. J. (1997). Brain mechanisms associated with top-down processes in perception. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *352*, 1221–1230.
- George, N., Jemel, B., Fiori, N., & Renault, B. (1997). Face and shape repetition effects in humans: A spatio-temporal ERP study. *NeuroReport*, *8*, 1417–1423.
- Grossberg, S., Mingolla, E., & Ross, W. D. (1997). Visual brain and visual perception: How does the cortex do perceptual grouping? *Trends in Neurosciences*, *20*, 106–111.
- Hertz, S., Porjesz, B., Begleiter, H., & Chorlian, D. (1994). Event-related potentials to faces: The effects of priming and recognition. *Electroencephalography and Clinical Neurophysiology*, *92*, 342–351.
- Hillyard, S. A., Teder-Salejari, W. A., & Munte, T. F. (1998). Temporal dynamics of early perceptual processing. *Current Opinion in Neurobiology*, *8*, 202–210.
- Humphreys, G. W., Riddoch, M. J., & Price, C. J. (1997). Top-down processes in object identification: Evidence from experimental psychology, neuropsychology and functional anatomy. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *352*, 1275–1282.
- Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, *394*, 784–787.
- Karni, A., & Bertini, G. (1997). Learning perceptual skills: Behavioral probes into adult cortical plasticity. *Current Opinion in Neurobiology*, *7*, 530–535.
- Kersten, D. (1997). Perceptual categories for spatial layout. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *352*, 1155–1163.
- Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge: MIT Press.
- Lamme, V. A., Super, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology*, *8*, 529–535.
- Linkenkaer-Hansen, K., Palva, J. M., Sams, M., Hietanen, J. K., Aronen, H. J., & Ilmoniemi, R. J. (1998). Face-selective processing in human extrastriate cortex around 120 ms after stimulus onset revealed by magneto- and electroencephalography. *Neuroscience Letters*, *253*, 147–150.
- Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*, 577–621.
- Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology*, *77*, 24–42.
- Mel, B. W. (1997). SEEMORE: Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Computing*, *9*, 777–804.
- Mesulam, M. M. (1998). From sensation to cognition. *Brain*, *121*, 1013–1052.
- Mozer, M. C. (1991). *The perception of multiple objects: A connectionist approach*. Cambridge: MIT Press.
- Nowak, L. G., & Bullier, J. (1997). The timing of information transfer in the visual system. In J. Kaas, K. Rockland, & A. Peters (Eds.), *Extrastriate cortex in primates* (vol. 12, pp. 205–241). New York: Plenum.
- 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.
- Oram, M. W., & Richmond, B. J. (1999). I see a face—a happy face. *Nature Neuroscience*, *2*, 856–858.
- 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*, 329–342.
- Przybylski, A. W. (1998). Vision: Does top-down processing help us to see? *Current Biology*, *8*, R135–R139.
- Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extraclassical receptive-field effects. *Nature Neuroscience*, *2*, 79–87.
- Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, *395*, 376–381.
- Rolls, E. T., & Tovee, M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *257*, 9–15.
- Rolls, E. T., & Treves, A. (1998). *Neural networks and brain function*. New York: Oxford University Press.
- Sastry, P. S., Shah, S., Singh, S., & Unnikrishnan, K. P. (1999). Role of feedback in mammalian vision: A new hypothesis and a computational model. *Vision Research*, *39*, 131–148.
- Schacter, D. L., & Buckner, R. L. (1998). Priming and the brain. *Neuron*, *20*, 185–195.
- Schendan, H. E., Ganis, G., & Kutas, M. (1998). Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms. *Psychophysiology*, *35*, 240–251.
- Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time and scale dependent scene recognition. *Psychological Science*, *5*, 195–200.
- Seeck, M., Michel, C. M., Mainwaring, N., Cosgrove, R., Blume, H., Ives, J., Landis, T., & Schomer, D. L. (1997). Evidence for rapid face recognition from human scalp and intracranial electrodes. *NeuroReport*, *8*, 2749–2754.
- Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, *400*, 869–873.
- Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1997). Oscillatory gamma-band (30–70 Hz) activity induced by a visual search task in humans. *Journal of Neuroscience*, *17*, 722–734.
- Tanaka, K. (1997). Mechanisms of visual object recognition: Monkey and human studies. *Current Opinion in Neurobiology*, *7*, 523–529.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.
- Thorpe, S. J., & Imbert, M. (1989). Biological constraints on connectionist models. In R. Pfeifer, Z. Schreier, F. Fogelman-Soulie, & L. Steels (Eds.), *Connectionism in perspective* (pp. 63–92). Amsterdam: Elsevier.
- Tovee, M. J., Rolls, E. T., & Ramachandran, V. S. (1996). Rapid visual learning in neurones of the primate temporal visual cortex. *NeuroReport*, *7*, 2757–2760.
- Treisman, A. M., & Kanwisher, N. G. (1998). Perceiving visually presented objects: Recognition, awareness, and modularity. *Current Opinion in Neurobiology*, *8*, 218–226.

- Ullman, S. (1996). *High-level vision: Object recognition and visual cognition*. Cambridge: MIT Press.
- Van Rullen, R., Gautrais, J., Delorme, A., & Thorpe, S. (1998). Face processing using one spike per neurone. *Biosystems*, *48*, 229–239.
- Vecera, S. P., & O'Reilly, R. C. (1998). Figure-ground organization and object recognition processes: An interactive account. *Journal of Experimental Psychology, Human Perception and Performance*, *24*, 441–462.
- Vidyasagar, T. R. (1998). Gating of neuronal responses in macaque primary visual cortex by an attentional spotlight. *NeuroReport*, *9*, 1947–1952.
- Wallis, G., & Bülthoff, H. (1999). Learning to recognize objects. *Trends in Cognitive Sciences*, *3*, 22–31.
- Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology*, *51*, 167–194.
- Warren, C., & Morton, J. (1982). The effects of priming on picture recognition. *British Journal of Psychology*, *73*, 117–129.
- Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, *8*, 227–233.