

Fast saccades toward faces: Face detection in just 100 ms

Sébastien M. Crouzet

Centre de Recherche Cerveau and Cognition, UMR, CNRS,
Université Toulouse, Toulouse, France



Holle Kirchner

Centre de Recherche Cerveau and Cognition, UMR, CNRS,
Université Toulouse, Toulouse, France



Simon J. Thorpe

Centre de Recherche Cerveau and Cognition, UMR, CNRS,
Université Toulouse, Toulouse, France



Previous work has demonstrated that the human visual system can detect animals in complex natural scenes very efficiently and rapidly. In particular, using a saccadic choice task, H. Kirchner and S. J. Thorpe (2006) found that when two images are simultaneously flashed in the left and right visual fields, saccades toward the side with an animal can be initiated in as little as 120–130 ms. Here we show that saccades toward human faces are even faster, with the earliest reliable saccades occurring in just 100–110 ms, and mean reaction times of roughly 140 ms. Intriguingly, it appears that these very fast saccades are not completely under instructional control, because when faces were paired with photographs of vehicles, fast saccades were still biased toward faces even when the subject was targeting vehicles. Finally, we tested whether these very fast saccades might only occur in the simple case where the images are presented left and right of fixation by showing they also occur when the images are presented above and below fixation. Such results impose very serious constraints on the sorts of processing model that can be invoked and demonstrate that face-selective behavioral responses can be generated extremely rapidly.

Keywords: fast visual processing, face detection, visual processing time, saccade, eye movement

Citation: Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, 10(4):16, 1–17, <http://journalofvision.org/10/4/16/>, doi:10.1167/10.4.16.

Introduction

Measurements of processing speed in the visual system can be very useful for constraining models. For example, in a manual go/no-go task, subjects can reliably release a button when an animal is present in a natural scene from around 300 ms after stimulus onset (although mean reaction times are longer), and in the same situation, there is a differential EEG response between target and distractor trials that appears only 150 ms after stimulus onset (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Thorpe, Fize, & Marlot, 1996). These latencies are undoubtedly very short given the computational complexity of the task and have led to the suggestion that at least some sorts of high-level visual tasks can be performed on the basis of a single feed-forward sweep through the visual system (Serre, Oliva, & Poggio, 2007; Thorpe & Imbert, 1989; VanRullen & Thorpe, 2002). Nevertheless, there is evidence that processing involving feedback can occur very rapidly. For example, the question of whether a particular region of the image is foreground or background affects activity in areas such as V1 within a few tens of milliseconds of the start of the neural response (Qiu, Sugihara, & von der Heydt, 2007; Roelfsema, Tolboom, &

Khayat, 2007). As a consequence, even processing involving both a feed-forward and a feed-back pass could be possible very rapidly (Epshtein, Lifshitz, & Ullman, 2008). For this reason, precise measurements of processing time are likely to become even more important for distinguishing between processing that can be achieved with a single feed-forward pass and processing that leaves enough time for both bottom-up and top-down mechanisms to be involved.

While much of the evidence for ultra-rapid scene processing has come from using a manual go/no-go categorization task coupled with event-related potentials, it has become clear that there are limitations to this approach. One of the strongest arguments in favor of the idea that the differential EEG response at around 150 ms is indeed related to categorization comes from the fact that the differential activity can be modulated by changing the target category (VanRullen & Thorpe, 2001b). Thus, when subjects were required to switch the target category from “animal” to “means of transport” in different blocks, there were differences in the ERP response from around 150 ms that depended not on the physical characteristics of the stimulus but rather on the status (“target vs. distractor”) of the image. This effectively rules out the possibility that the differential activity simply results from

irrelevant low-level differences between the stimuli. However, other studies pointed out that a considerable part of the differential activity occurring at short latencies was not affected by the status of the stimulus (Johnson & Olshausen, 2003, 2005). This phenomenon was particularly marked in a study using human and animal faces as stimuli. Subjects were extremely good at responding selectively to either human or animal faces and could switch virtually effortlessly between the two different target categories from block to block (Rousselet, Mace, & Fabre-Thorpe, 2003), achieving accuracy levels of over 97% correct, irrespective of whether the target category was human or animal. Despite this, the analysis of the simultaneously recorded ERP signals failed to find any evidence that short latency responses (i.e., at latencies below about 180 ms) could be modulated by whether the particular stimulus was a target or not (Rousselet, Mace, Thorpe, & Fabre-Thorpe, 2007). There were strong differences in the ERP signals recorded with human and animal faces, but those differences were unaffected by whether the subject was treating the image as a target or not. If a dependence on task status is needed to be able to infer that a particular neural response is related to high-level processing, then it would be natural to conclude that no influence of such high-level factors is visible until around 180 ms in this task.

Does this mean that the differential brain responses seen earlier should be dismissed as simple artifacts due to irrelevant low differences between images? One type of result that argues strongly against this are recent studies using a saccadic choice task that show that useful behavioral responses can be generated well before the 150–180 ms latency value suggested by the task-dependent ERP effects. In the first such study, Kirchner and Thorpe reported that when two natural scenes are simultaneously flashed left and right of fixation, reliable saccades to images containing animals can be initiated as early as 120–130 ms after images onset (Kirchner & Thorpe, 2006). Given that motor preparation presumably needs at least 20 ms, this implies that the underlying visual processing may need only 100 ms, considerably earlier than the 150-ms latency of the first differential activity.

In the present study, we have used a similar saccadic choice task protocol to the one used previously by Kirchner and Thorpe, but with another type of highly significant visual stimulus—photographs of human faces. **Experiment 1** directly compared saccadic reaction times for three stimulus categories—animals, faces, and vehicles—and demonstrated a very impressive result for faces for which the fastest reliable saccades were seen at latencies as early as 100–110 ms. Then, **Experiment 2** revealed a search asymmetry between faces and vehicles: subjects found it much easier to saccade toward the faces than toward the vehicles. Interestingly, they were only able to fully overcome this bias for saccades with relatively long latencies. Finally, **Experiment 3** allowed us to test a simplistic model of the task based on a

comparison between activation in the left and right hemispheres. It showed that the fast detection of faces is not restricted to a horizontal display arrangement and must presumably rely on a more complex process than a simple comparison between activation levels in the two hemispheres.

The results, some of which have been presented previously (Crouzet, Kirchner, & Thorpe, 2008), fit with a large range of previous studies that have demonstrated that faces have a special status. More importantly, they impose some very serious temporal constraints on the underlying processing. Specifically, face-selective behavior can be initiated at latencies so short that there may not be enough time to complete the first wave of processing in the cortical areas of the ventral stream. If so, other possibilities including the involvement of subcortical pathways may be involved.

Experiment 1: Comparison between three target categories

Experiment 1 examined whether performance in the Saccadic Choice Task varies when different object categories are used as the target (faces, animals, or vehicles). Each of these target categories was displayed in combination with the same set of “neutral distractors”, corresponding to various natural scenes. Thus, any differences observed between the three conditions will reflect difference of processing time between the three categories of objects, rather than differences between the distractor stimuli.

Methods

Participants

Eight volunteers (7 men; mean age 24.5 years, ranging from 22 to 31 years) with normal or corrected-to-normal vision participated in a 2-AFC saccadic choice task. They all gave written informed consent to participate in the experiment.

Stimuli

One thousand photographs selected from the Corel Photolibrary database or downloaded from the Internet were used to set up four object categories of 250 natural scenes: faces, animals, vehicles, and neutral distractors. The neutral distractor category was composed of a range of images that all contained a salient object in the foreground. All the images were converted to grayscale and resized to 330 × 330 pixels. The global contrast of each image was reduced to 80% of the original image,

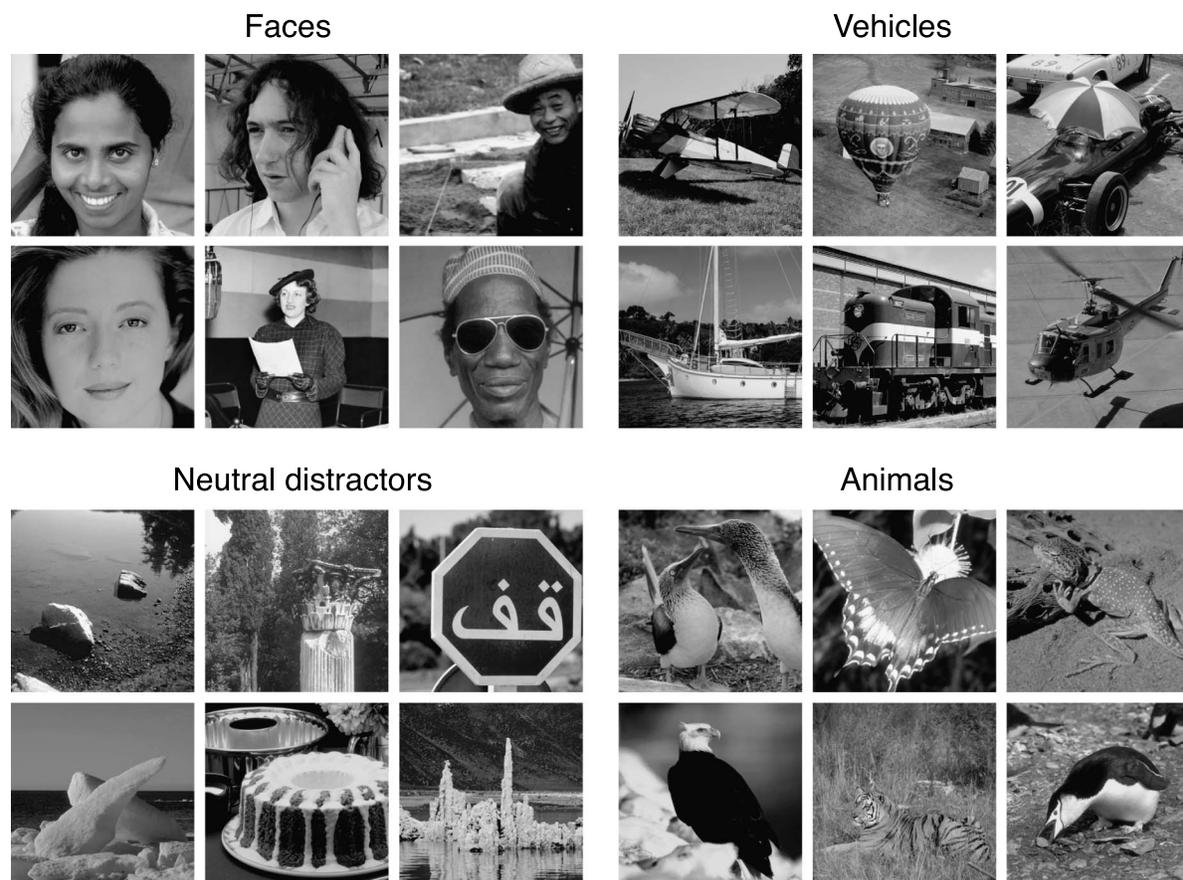


Figure 1. Examples of images used in this study.

allowing us to adjust the mean luminance of each image to a grayscale value of 128. (Guyonneau, Kirchner, & Thorpe, 2006). The complete set of images can be provided on request (Figure 1).

Apparatus

Participants viewed the stimuli in a dimly lit room with their heads on a chin rest to maintain the viewing distance at 60 cm. Stimuli were presented on an Iiyama Vision Master PRO 454 monitor with the screen resolution set to 800×600 pixels and a refresh rate of 100 Hz. The centers of the two images were always 8.6° from the fixation cross, resulting in a retinal size for each image of 14° by 14° . The experiment was run using the software Presentation 9.9 (Neurobehavioral Systems).

Protocol

The experiment was performed using a Saccadic Choice Task, a similar protocol to the one used by our team in previous studies (Guyonneau et al., 2006; Kirchner & Thorpe, 2006) with the exception that the natural scenes were displayed during 400 ms rather than flashed for 20 ms. The original reason for using such short presentation

times when using a conventional manual go/no-go protocol was to exclude the possibility of ocular exploration (Thorpe et al., 1996). However, in the present experiments, we specifically required the subjects to make saccades. Since we were recording the eye movements, the original argument for using flashed presentations was now obsolete. Preliminary experiments had already shown that, contrary to what might have been expected, this longer presentation durations significantly shortened the mean RT. Specifically, it appears that using longer presentation times reduces the number of late saccades, resulting in a leftward shift and a sharpening of the Saccadic Reaction Time (SRT) distribution. An explanation could be that the offset of the images in the original protocol perturbed the initiation of some saccades, resulting in a right-biased distribution. A second difference with respect to the previous studies was that the background screen was set at a grayscale value of 128 rather than black.

Observers had to keep their eyes on a black fixation cross that disappeared after a pseudo-random time interval (800–1600 ms), leaving a 200-ms time gap before the presentation of the images (Fischer & Weber, 1993; Kirchner & Thorpe, 2006). The use of such a gap allows saccades to be initiated more rapidly. Two natural scenes,

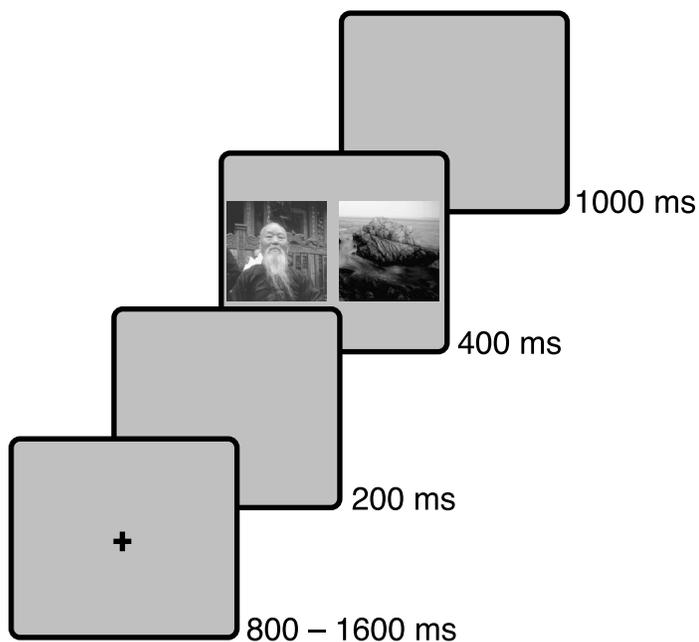


Figure 2. Protocol: The saccadic choice task. Observers had to fixate a cross in the center during a pseudo-random time (800–1600 ms). After a gap of 200 ms, 2 images were displayed left and right of fixation for 400 ms. Observers then had 1000 ms to prepare for the next trial.

one target and one distractor, were then displayed on each side of the screen for 400 ms (see Figure 2). The task was to make a saccade as quickly and as accurately as possible to the side where an object belonging to the target category has appeared.

Using a within-subject design, three object categories were tested here (faces, animals, and vehicles). Each subject saw each target image once during the experiment and thus performed 250 trials in each condition, divided into blocks of 50 trials. The “neutral distractor” images were the same in the three conditions, so each one was seen several times by each participant. The order of the three conditions was counterbalanced across participants. Each block was preceded by a training session of 50 trials with images not used in the experiment.

Response recording and detection

Eye position was recorded using horizontal EOG electrodes (1 kHz, low pass at 90 Hz, notch at 50 Hz, baseline correction [−400:0] ms; NuAmps, Neuroscan). Saccadic Reaction Time (SRT) was determined offline as the time difference between the onset of the images and the start of the saccade. Each trial was verified by the experimenter to make sure that only the largest inflection (if any) was taken as a real saccade (see Kirchner & Thorpe, 2006 for more detailed information about the procedure); 15.2% of trials (912 of 6,000) had to be

excluded because of a noisy eye signal, but this percentage was evenly spread across conditions (face task = 16%, animal task = 14.6%, vehicle task = 15%).

Minimum reaction times

To determine a value for the minimum SRT, we divided the saccade latency distribution of each condition into 10-ms time bins (e.g., the 120-ms bin contained latencies from 115 to 124 ms) and searched for bins containing significantly more correct than erroneous responses using a χ^2 test with a criterion of $p < 0.05$. If 5 consecutive bins reached this criterion, the first was considered to correspond to the minimum reaction time.

Results and discussion

The principal finding from this first study was that subjects were fast and accurate in all three conditions (see

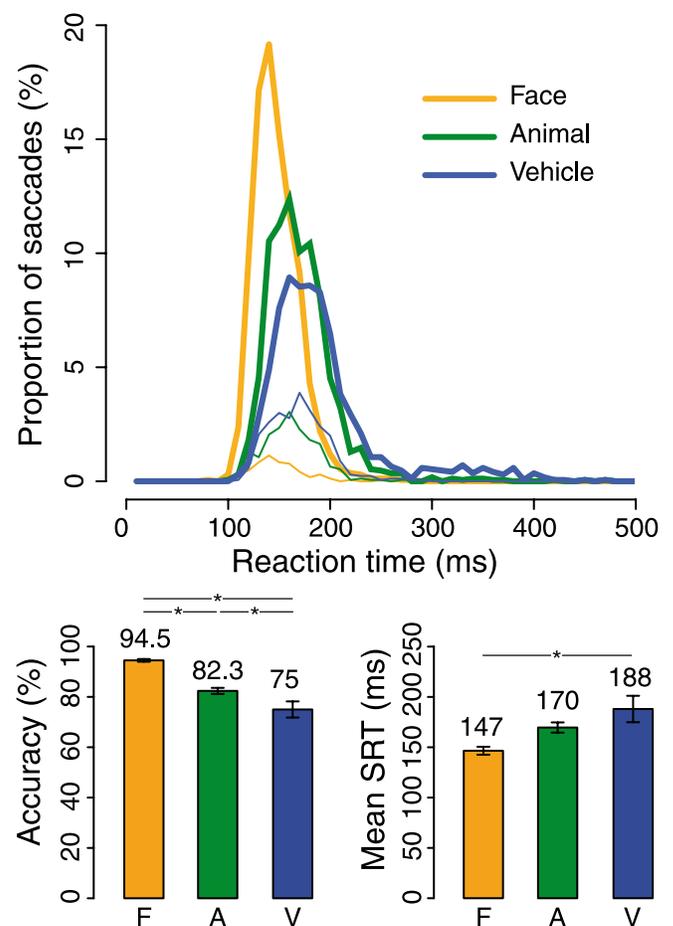


Figure 3. Experiment 1. (Top) Distributions of SRT for 3 different target categories: face, animal, vehicle. Correct responses are shown in thick lines, incorrect as thin lines. (Bottom) Mean accuracy and SRT in the 3 conditions. Errors bars are SEM.

Figure 3). Ultra-rapid processing of objects in the saccadic choice task is clearly not restricted to animals but can also be extended to vehicles and human faces. However, and contrary to what has been shown using manual response (Rousselet et al., 2003; VanRullen & Thorpe, 2001a), our results showed a clear ordering between categories for both SRT and accuracy. A one-factor ANOVA analysis showed that there was a global effect of the category used as target on both mean SRT ($F(2,14) = 10.622, p < 0.01$) and accuracy ($F(2,14) = 26.031, p < 0.001$). A post-hoc Tukey analysis for multiple comparisons showed a progressive increase on accuracy from the “vehicle” condition (75%) to the “animal” (82.4%) and “face” (94.5%) conditions. Mean SRT was only significantly different between the face (147 ms) and vehicle (188 ms) conditions.

Different processing times for different object categories

In addition to the very clear differences in mean reaction times seen for the three object categories, there were also very striking differences in the minimum reaction time values (i.e., the first bin of at least 5 consecutive bins in the reaction time distribution where there was a significantly higher proportion of correct than erroneous responses). In the case where the target category was “animal”, the value obtained replicated Kirchner and Thorpe’s (2006) study with a minimum SRT of 120 ms. Interestingly, this minimum SRT was clearly higher for vehicles (140 ms) and lower for faces (110 ms). Together, the results demonstrate a very clear advantage for the processing of faces over animals and vehicles when observers had to discriminate these object categories from “neutral distractors”.

Experiment 2: Faces vs. vehicles

Previous studies using a manual go/no-go protocol had shown that subjects can switch from one target category to another in different blocks with little cost in terms of either accuracy or reaction time. This was seen for both the situation where the target categories are animals and means of transport (VanRullen & Thorpe, 2001a), as well as with humans versus animals (Rousselet et al., 2003). In Experiment 2, we ask whether this ability to switch between target categories also exists in the case of the saccadic choice task. Specifically, we designed an experiment in which the subjects had to discriminate directly between two object categories: faces and vehicles. This design allowed us to directly compare processing times between these two categories of objects, but additionally, we can see if the task can be reversed under voluntary control. As an example, if a subject was instructed to treat

faces as targets and vehicles as distractors in a first block, subsequent blocks could require the reverse configuration, with vehicles as targets and faces as distractors. The results reveal a clear asymmetry, with saccading to faces being considerably faster and more accurate than to vehicles. Indeed, subjects had great difficulty in making fast saccades toward vehicle targets.

Methods

Participants

Eight volunteers (5 men; mean age 26.9 years, ranging from 23 to 34 years) with normal or corrected-to-normal vision participated in a 2-AFC saccadic choice task. They all gave written informed consent to participate in the experiment.

Stimuli

In order to have a more controlled set of stimuli, 200 photographs were selected from the Corel database and the Internet to generate two object categories, each with 100 images: faces and vehicles. The faces were 50% men and 50% women, while the vehicles were 50% cars and 50% trains. Each subcategory was divided equally into close-up and mid-distance views. Manipulations on the luminance and contrast of each image were the same as in Experiment 1.

Apparatus and protocol

The design was unchanged from Experiment 1 with the following exception: each subject saw each image four times, both as target and distractor and in both the left and right hemifields. Specifically, each participant performed 200 trials in each of the two conditions (face and vehicle). The order of the two conditions was counterbalanced across participants; 23% of trials (746 of 3,200) had to be excluded because of a noisy eye signal, but this percentage was evenly spread across conditions (face task = 22.3%, vehicle task = 24.4%).

Results and discussion

The first main result of Experiment 2 is that even with another object category as distractor, here vehicles, saccades toward faces can still be initiated very rapidly. Overall accuracy was 89.6%, with a mean SRT of 138 ms. Remarkably, the earliest reliable saccades appeared just 100–110 ms after scene onset. However, the situation was markedly different when the target category was vehicle. In this case, the values for mean SRT (167 ms) and accuracy (71%) were both significantly poorer than with faces ($F(1,7) = 84.723, p < 0.001$ and $F(1,7) = 44.867, p < 0.001$, respectively; Figure 4).

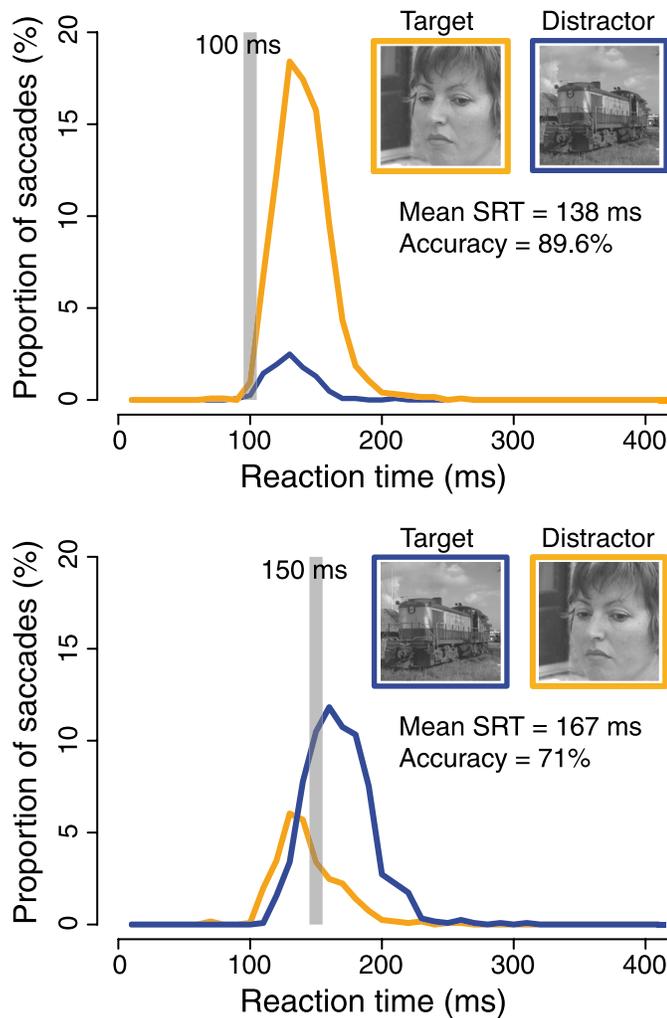


Figure 4. (Top) Distribution of SRT over all subjects when the task is to saccade toward faces (responses toward faces in orange, vehicles in blue). (Bottom) Distribution of SRT for all subjects when the task is to saccade toward vehicles. The gray vertical bar indicates the bin where correct responses start to significantly outnumber errors.

Even more striking was the distribution of response in the 100–140 ms time window. There was a tendency of saccades initiated in this time range to go toward the side with the faces, even if the task is to go to vehicles. Thus, saccades initiated before 140 ms seemed to be hard to control.

An interesting observation is that if we divide the data according to the position of the target (left or right, Figure 5), there is a clear tendency for subjects to be faster and more accurate when the target is on the left. This tendency can be observed in the face task (left: 135 ms and 95.8%; right: 143 ms and 83.3%) and in the vehicle task (left: 165 ms and 77.2%; right: 170 ms and 64.8%) and is significant for mean RT ($F(1,7) = 6.4953, p < 0.05$) although not for accuracy ($F(1,7) = 4.0321, p = 0.084$). Thus, if the target is in the left hemifield, participants

produced fewer errors and their correct responses had a shorter mean SRT. When the target was in the right hemifield, participants made more early errors. Furthermore, many of the errors in the face task were made when the target is on the right and the distractor on the left, especially on fast saccades. For example, the specific pattern observed in Figure 4 with more responses toward faces than vehicles in the 100–140 ms time window is largely the result of the situation when the face is on the left. A similar tendency to produce more saccades on the left has also been reported when people look at chimeric faces (Butler et al., 2005). These left hemifield biases could be related to the well-known fact that neural responses in the right hemisphere are reliably stronger than on the left, a result that has been repeatedly seen in both fMRI studies (e.g., Hemond, Kanwisher, & Op de Beeck, 2007; Kanwisher, 2000) and ERPs (e.g., Jacques & Rossion, 2009; Rousselet, Mace, & Fabre-Thorpe, 2004). Indeed, the existence of a left hemifield advantage when saccading to faces supports the hypothesis that the saccadic choice task really does involve face processing mechanisms. If it were simply a bias toward making saccades toward the left, the same biases would be expected with any sort of target.

Additionally, half the stimuli were close-up views (CV), and the other half mid-distance views (MV). This allowed a post-hoc analysis of the effect of object size to be performed. It showed that there was absolutely no effect of the size of the target face on either SRT or on accuracy. This effect is null for the “face” (CV: 168 ms and 70%; MV: 165 ms and 72%) as well as for the “vehicle” conditions (CV: 138 ms and 90%; MV: 137 ms and 89%). Further studies to look at the effect of object size on discrimination performance in this task in a more systematic way would be of considerable interest.

In summary, performance in this task was remarkably fast and efficient. Subjects were able to move their eyes selectively to the side where the scene contains a face category target in as little as 100–110 ms. It is clear that we would not argue from this result that a face can be “recognized” in just 100 ms. Nevertheless, the result demonstrates that the visual system needs only around 100 ms to initiate an eye movement toward a face. Furthermore, the fact that subjects had such difficulty in reversing the task and saccading toward the vehicle suggests that this attractivity might be effectively hard-wired.

Experiment 3: Horizontal versus vertical positioning

One factor that might contribute to this remarkable level of performance may lie in the design of the protocol, in which the two images are displayed to the left and right of fixation. As a result, the two images will effectively be

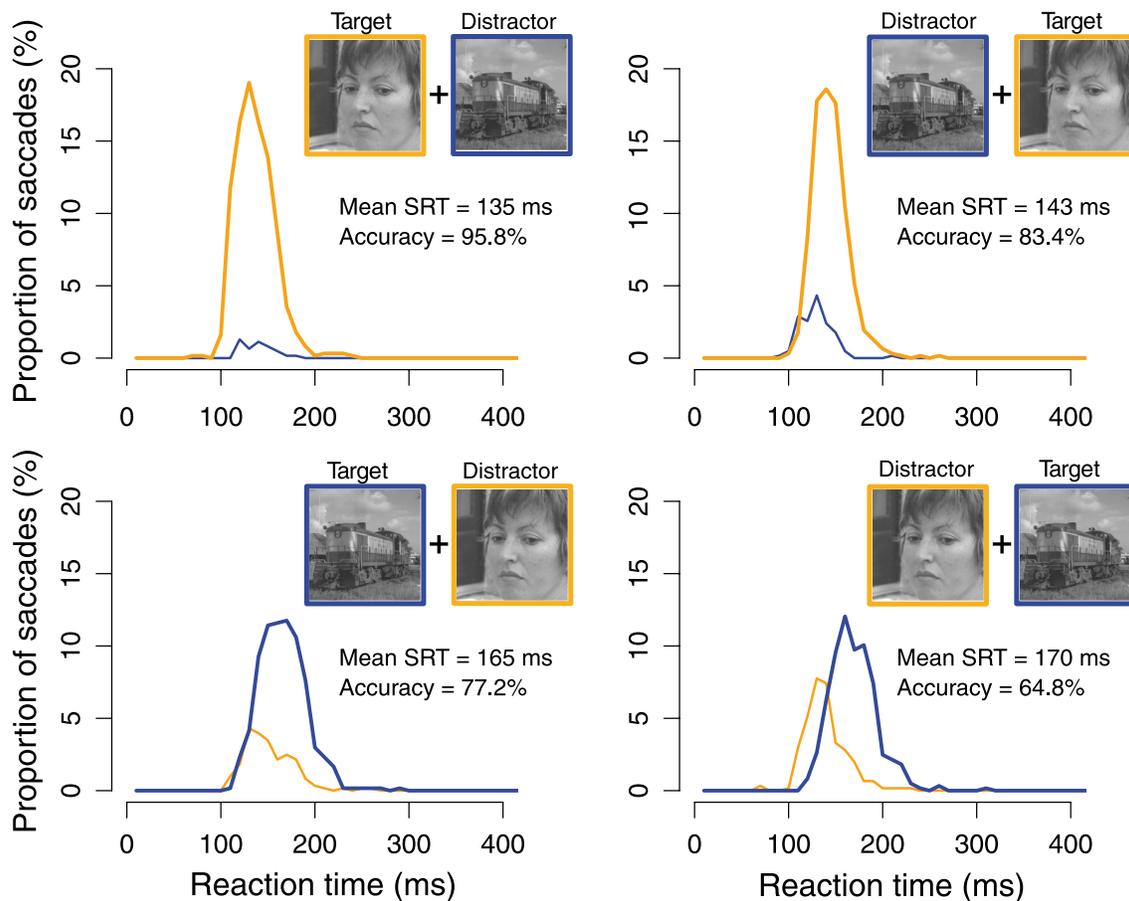


Figure 5. Distributions of SRT over all subjects when the task is to saccade toward faces (top row) or vehicles (bottom row) and when the target is on the left (left column) or on the right (right column). Correct responses are in thick lines, incorrect are in thin lines.

processed separately by the two hemispheres (at least initially) and this may provide a situation that is particularly favorable. Potentially, the task could be performed by comparing the activation in the two halves of the brain, and initiating the saccade to the side that has the strongest (or earliest) activation. This hypothesis was tested in [Experiment 3](#) in which subjects were asked to perform the same task with either the images displayed horizontally (to the left and right of fixation) or vertically (above and below the fixation point). [Experiment 3](#) also examined the difference between a saccadic choice task in which two images are presented at the same time and subjects required to saccade to a target, and simple detection task in which only one image is presented on each trial. As in [Experiment 2](#), subjects were required to perform the task with faces and vehicles images, varying the target category in different blocks.

Methods

Participants

Four volunteers (2 men; mean age 32.7 years, ranging from 23 to 50 years) with normal or corrected-to-normal

vision participated in a 2-AFC saccadic choice task. They all gave written informed consent to participate in the experiment.

Stimuli

Unchanged from [Experiment 2](#) ([Figure 6](#)).

Protocol

The protocol was unchanged from [Experiments 1](#) and [2](#) with the following exceptions. The experiment was divided in two sessions, each comprising 12 blocks of 50 trials. The first two blocks and the last two blocks in each session used a Simple Detection Task with only one image on each trial and no distractor, where the two categories of target (faces and vehicles) were mixed in the same block. Within each two-block group, one block used the images in the horizontal arrangement, while the other used a vertical one. For this Simple Detection Task, subjects were instructed to saccade as fast as possible on the side where there was an image, independently of the category of the image. It is important to notice that in the case of this Detection Task, no classification was needed.

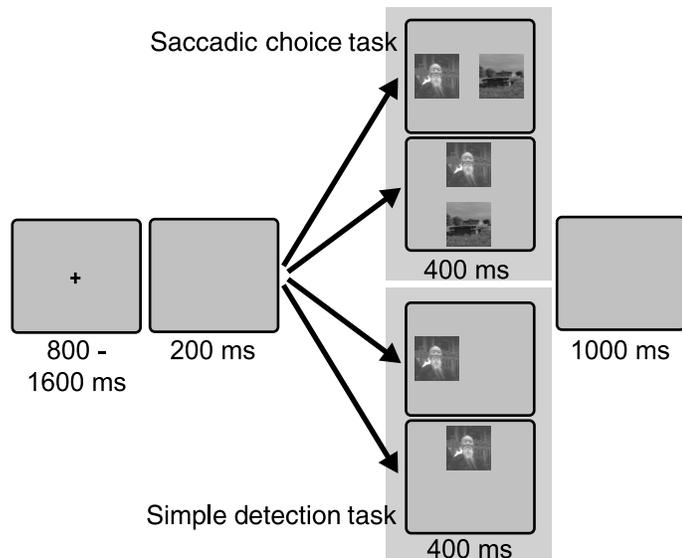


Figure 6. Design of [Experiment 3](#). The protocol was similar to the one used in [Experiments 1](#) and [2](#). After the 200-ms gap, and following a block design, participants had to perform a task in which either one image was presented (two screens on the bottom of the figure—Simple Saccadic Detection Task) or two images are presented simultaneously (two screens on the top—Saccadic Choice Task). In both cases, the images can be displayed horizontally or vertically.

In the middle part of each session, the subjects performed the Saccadic Choice Task with either four blocks with Faces as targets followed by four blocks with Vehicles as targets, or the contrary. In addition, the arrangement of the stimuli was varied with two blocks of vertically arranged stimuli alternating with two blocks using the horizontal arrangement. All the different orders were counterbalanced across the four subjects and the two sessions. Image size

was still set at 330×330 but the resolution of the screen was increased to 1024×768 to allow the images to be displayed vertically. As a consequence, the retinal size of the images was 11° by 11° , and the center of images was 6.8° from the center of the screen.

Eye movement recording

Unlike [Experiments 1](#) and [2](#), in this experiment the eye movements were monitored using a Chronos Eye Tracker (Chronos Vision, Berlin, Germany). This infrared tracking system samples eye position at 200 Hz binocularly. Saccade detection was performed offline, based on a velocity criterion and all the saccades were verified by the experimenter. Only the first saccades to end beyond 4° of eccentricity (corresponding to a minimum of 25% of the width of the image) were included in the analysis; 10.9% of trials had to be excluded using this criterion or because of a poor detection of the pupil. Before each block, an 8-point calibration procedure was performed.

Results and discussion

Vertical vs. horizontal display

As can be seen from [Table 1](#), performance when the images were arranged vertically remained very good and was quite similar to the results obtained with the horizontal arrangement. Indeed, there was no overall effect of the arrangement of the stimuli (horizontal or vertical) on accuracy (as demonstrated by a two-way ANOVA). In contrast, the mean RTs for horizontal saccades were significantly shorter than vertical ones (167 ms and 178 ms, respectively; $F(1,9) = 9.1069$, $p < 0.05$). As might be expected from the results of [Experiments 1](#) and [2](#), the nature of the target (face or vehicle) still has a strong

Target category	Target location	Saccadic choice task			Simple detection task	
		Mean SRT (ms)	Accuracy (%)	Min. SRT (ms)	Mean SRT (ms)	Min. SRT (ms)
Face	Left	150 ± 14	95.5 ± 2		119 ± 12	
	Right	159 ± 13	84.3 ± 5.9		133 ± 14	
	Horizontal display	154 ± 13	89.8 ± 3	100	126 ± 12	80
	Bottom	165 ± 17	85.4 ± 6.5		146 ± 11	
	Top	168 ± 11	86.6 ± 5.8		135 ± 10	
Vehicle	Vertical display	166 ± 14	86.3 ± 4.2	110	139 ± 10	90
	Left	176 ± 17	83 ± 9		125 ± 15	
	Right	185 ± 19	68.8 ± 9.1		147 ± 16	
	Horizontal display	180 ± 17	75.8 ± 6.8	170	136 ± 15	80
	Bottom	190 ± 20	67.1 ± 8		140 ± 13	
	Top	187 ± 13	74.3 ± 5.8		138 ± 10	
	Vertical display	189 ± 16	71 ± 4.7	190	139 ± 11	90

Table 1. Results for [Experiment 3](#). Mean SRT and accuracy are presented for both the Saccadic Choice Task (with two simultaneously presented images) and the Simple Detection Task (a single image presented).

effect both on mean RT and accuracy ($F(1,9) = 48.7835$, $p < 0.001$; $F(1,9) = 31.5336$, $p < 0.001$, respectively). Thus, these results showed a slight difference between a horizontal and a vertical display.

However, a more detailed analysis that divided the results according to the precise location of the target (Left, Right, Top, Bottom) revealed that the difference between the results with horizontal and vertical displays was essentially due to the strong advantage when the target is presented on the left, an effect already seen in [Experiment 2](#). Thus, it seems that in these experiments, saccading to the right, the top, or the bottom of the screen was roughly equivalent, but that saccading to the left was faster.

The critical point here was that subjects were still able to produce very fast responses in the saccadic choice task even when the stimuli were positioned vertically. This effectively rules out a simple “hemisphere comparison” hypothesis in which the eyes simply move because there is an imbalance of activity between the left and right hemispheres. This is because, presumably, when the images are positioned vertically the amount of activation in the two hemispheres will be roughly balanced. However, it is worth noting that there is also evidence that the processing of the upper and lower visual fields involves anatomically separate areas in extrastriate cortex. As a consequence, it may still be possible to envisage a competitive mechanism in which global activation levels in two separate brain structures are compared. Further experiments would be needed to test the limits of this sort of ability by, for example, presenting both the target and distractor stimuli in the same hemifield, or even in the same quadrant.

Saccadic choice task vs. simple saccadic detection task

A second major result emerging from [Experiment 3](#) was the relatively small difference between SRT distributions in the simple saccadic detection task and the saccadic choice task, at least with faces as targets. The fact that there are essentially no errors in the simple saccadic detection task means that it is not useful to compare accuracy levels. Mean SRT values were 159 ms for faces versus 183 ms for vehicles in the saccadic choice task. The corresponding values were 131 ms and 137 ms in the simple saccadic detection task. A two-way ANOVA showed that there is still a significant effect of the category of the target ($F(1,9) = 22.9889$, $p < 0.001$), and a clear advantage for simple detection over the saccadic choice task ($F(1,9) = 144.6105$, $p < 0.001$). Furthermore, the interaction between the task and the category of the target is significant ($F(1,9) = 9.0315$, $p < 0.05$), meaning that the difference between saccadic choice and simple detection is much larger for vehicles than for face targets. This is also clear from looking at the minimum SRT because when the target was a face, the difference between

minimum SRTs for the choice task and the simple detection task was only 20 ms. In comparison, when the target was a vehicle, this additional time cost was 80 ms.

General discussion

The present study used a saccadic choice task to investigate the time course of the processing involved in ultra-rapid detection of objects in natural scenes. The experiments follow on from an earlier study that had shown that subjects can make rapid and reliable saccades toward animal targets when two images are simultaneously flashed left and right of fixation (Kirchner & Thorpe, 2006). [Experiment 1](#) showed that these very rapid saccadic responses can be initiated even more rapidly when human faces are the target category, with the fastest saccades being initiated from around 110 ms following the onset of the stimuli. Then, [Experiment 2](#) showed that this strong bias toward saccading toward faces is very difficult to suppress, because even when subjects are actively trying to saccade toward vehicles, they still show a very clear tendency for fast saccades to be directed toward faces. Finally, [Experiment 3](#) showed that this ability to initiate very fast saccades toward face targets is not restricted to the specific left/right design and thus cannot be explained by a simple comparison between activation levels in the two hemispheres.

Ultra-rapid processing of faces

The values reported here for saccadic reaction times with faces are remarkably short. In [Experiment 2](#), in which faces were paired with photographs of cars and trains, the mean onset latency for saccades toward faces was a mere 138 ms. Virtually all the saccades were initiated in under 200 ms, but despite this, accuracy was a very respectable 89.6%. By examining the distribution of correct and erroneous saccades in each 10-ms bin of the reaction time distribution, we were able to show that the minimum reaction time in this task was only 100–110 ms. Such values put very severe temporal constraints on the underlying visual processing, especially when one takes into account the fact that the latencies obtained from the EOG and eye tracker data used in this study include the time needed to initiate the eye movement. Most neurophysiologists would allow roughly 20 ms for the activation of the brain stem structures involved in oculomotor control and the muscles of the eye itself. If true, then it appears that information about the presence of a face in an image may be available as little as 80 ms after the onset of the image. Such values are considerably shorter than

previous behavioral estimates of processing times in the human visual system (Thorpe et al., 1996) and are even a lot shorter than the 150-ms differential ERP response between targets and distractors that has previously been used as a measure of processing time.

The mean SRT values seen here are substantially shorter than those reported in the previous study by Kirchner and Thorpe (2006) in which the median SRT when using photographs of animals as targets was 228 ms. However, there are a number of differences between the two experiments that could explain why the reaction times were so much shorter here. In [Experiment 1](#), we used a very similar situation with photographs of animals paired with complex natural scenes as distractors and obtained a mean SRT value of 170 ms. It seems likely that some of this reduction in reaction time results from the fact that in the current experiments, the two images remain present for 400 ms, instead of simply being flashed for 20 ms, as in the original study. It could be that removing the stimulus before the saccade has been initiated in the original design might interfere with saccade initiation. In contrast, by leaving the images on for 400 ms, the subject is in the very natural situation of initiating saccades toward a stimulus that is still present when the eyes arrive at their destination. Given that the aim of the experiment is to obtain the shortest realistic measurement of the time required to initiate a saccade toward a target, there seems to be little reason to continue with the original design, which only seems to introduce additional variability in the reaction time distribution.

Comparison with previous eye movement studies

The tendency of humans to look preferentially at faces when exploring visual scenes was already clear in the classic studies of Buswell (1935) and Yarbus (1967), and a recent study showed that similar biases also exist in chimpanzees (Kano & Tomonaga, 2009). However, it is important to realize that several factors may be involved in producing such biases (Henderson, 2003). For example, there may be a tendency to fixate faces for longer than other less interesting parts of the image. The most well widely used models of gaze control, such as Itti and Koch's (2000) saliency model rely on local variations in relatively low-level factors such as color, orientation, and luminance. While such models can account for a substantial proportion of real-world gaze patterns in humans (Parkhurst, Law, & Niebur, 2002; Peters, Iyer, Itti, & Koch, 2005; Tatler, Baddeley, & Gilchrist, 2005), there are a number of studies showing that such models cannot be considered complete (Birmingham, Bischof, & Kingstone, 2008; Cerf, Harel, Einhäuser, & Koch, 2008). Indeed, by changing the task requirements, it is possible to override these low-level biases (Einhauser, Rutishauser, & Koch, 2008).

A separate question concerns the issue of whether the very first saccades that are generated in response to a scene can be directed to important objects such as faces, and if they can, at what latency. The study by Kirchner and Thorpe had already demonstrated this for animal targets, and a recent study showed that these rapid saccades toward animals can be quite accurate in terms of localization (Drewes, Trommershaeuser, & Gegenfurtner, 2009). Another recent study by Fletcher-Watson, Findlay, Leekam, and Benson (2008) extended the use of the saccadic choice task to the detection of humans. Observers viewed two images presented at the same time on the left and right of a screen, only one of which contained a human. They reported that participants tend to saccade more on the side with the human, and that this bias was indeed seen for the very first saccades. The saccades were grouped into bins of 50 ms, and not surprisingly, the first two bins had very few responses. However, in the bin from 100 to 149 ms, 90% of the saccades were oriented toward the image containing a face. One particular feature of that study was that they included a task where the subjects had to saccade spontaneously to one of two images, with no task to perform. The fact that even here subjects showed a strong tendency to look toward the side with the human suggests that there may well be a built-in bias toward looking at humans.

Other data supporting the idea that faces can be processed very efficiently comes from the extensive literature on attentional capture (Bindemann, Burton, Hooge, Jenkins, & de Haan, 2005; Langton, Law, Burton, & Schweinberger, 2008; Ro, Russell, & Lavie, 2001; Theeuwes & Van der Stigchel, 2006; Vuilleumier, 2000) as well as recent studies reporting pop-out in displays containing large numbers of elements (Hershler & Hochstein, 2005, 2006; although see also Brown, Huey, & Findlay, 1997; Vanrullen, 2006). Additionally, the bias toward faces has also been seen using an anti-saccade protocol, which demonstrated that subjects have difficulty in looking away when a face is present (Gilchrist & Proske, 2006). Together, all these results point toward a real behavioral advantage for faces in a wide range of situations.

Underlying brain mechanisms

These behavioral effects are also reflected at the level of brain mechanisms. Numerous studies have suggested that faces may have a special computational status that would allow them to be processed more efficiently and faster than other classes of objects (Farah, Wilson, Drain, & Tanaka, 1998; Haxby, Hoffman, & Gobbini, 2000; Kanwisher, 2000; but see Tarr & Gauthier, 2000). Faces are known to activate spatially adjacent but distinct brain regions in both humans and monkeys (Freiwald, Tsao, & Livingstone, 2009; Sergent, Ohta, & MacDonald, 1992; Tsao & Livingstone, 2008). This has been shown in a

range of techniques including PET, fMRI, and single unit recording (Freiwald et al., 2009; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999; Kanwisher, McDermott, & Chun, 1997; Puce, Allison, Gore, & McCarthy, 1995). Nevertheless, there is still debate about the nature and degree of specialization (Cohen & Tong, 2001; Downing, Jiang, Shuman, & Kanwisher, 2001; Haxby et al., 2001).

The fact that the earliest reliable saccades toward faces can be seen as early as 100–110 ms after stimulus onset places particularly severe constraints on the underlying brain mechanisms. In the present context, it is particularly important to look at experimental evidence on the speed with which information about faces can be processed. Following the earliest reports of face-selective Event Related Potentials (Jeffreys, 1989), much attention has been paid to the N170 potential that seems to be particularly strongly associated with face processing (Bentin, Allison, Puce, Perez, & McCarthy, 1996; McCarthy, Puce, Belger, & Allison, 1999; for a recent review, see Rossion & Jacques, 2008). However, it seems likely that the N170 occurs too late to be directly involved in triggering the fastest saccades reported here. Nevertheless, there have been repeated reports of face-selective electrophysiological responses occurring at even earlier latencies. For example, Liu, Harris, and Kanwisher (2002) reported face-selective MEG activation at latencies of around 100 ms, and selective ERP responses to emotional faces have been reported with latencies of around 120 ms (Eimer & Holmes, 2002). There have even been reports of face-selective repetition-related effects at even shorter latencies, sometimes as early as 45–80 ms (George, Jemel, Fiori, & Renault, 1997; Mouchetant-Rostaing & Giard, 2003; Mouchetant-Rostaing, Giard, Bentin, Aguera, & Pernier, 2000) or even 30–60 ms (Braeutigam, Bailey, & Swithenby, 2001), but it has been unclear whether these very rapid differential effects are really related to face perception. Clearly, in the light of the present behavioral responses, it may be appropriate to reconsider the significance of these very early phenomena.

Another important source of information about processing speed is the results of single-cell recording studies in awake primates that have shown that face-selective neuronal responses can be seen from around 100 ms although the fastest single unit responses can be as early as 70 ms (Oram & Perrett, 1992). It is also important to determine precisely when information about object identity can be read out from the activity of a population of cells. This issue was addressed in a recent study that examined the responses of populations of single neurons in monkey inferotemporal cortex and found that decisions about both object category and identity could be made on single trials from around 100 ms using a temporal window of only 12.5 ms in duration (Hung, Kreiman, Poggio, & Dicarlo, 2005).

In contrast to work in monkeys, there have been relatively few single unit recording studies in humans (though see, for example, the work that has been done on recordings from the medial temporal lobe in epileptic

patients; Mormann et al., 2008). Most human data comes from ERP and MEG recordings that are less easy to relate directly to behavioral reaction times because their analysis requires pooling together responses from a very large number of trials. The problem is that while a significant effect may be detected in the pooled data at a given latency, this does not mean that enough information could be extracted in real time on a single trial as would be required to initiate a behavioral response. However, a recent study of local field potential recordings obtained from occipito-temporal cortex in human epileptic patients reported that even in humans, information about object category can reliably be derived on a single trial basis from only 100 ms after stimulus onset (Liu, Agam, Madsen, & Kreiman, 2009). The study also showed that face-selective intracerebral responses were remarkably invariant to changes in size, position, and viewpoint, even at such short latencies. While these neurophysiological studies provide clear evidence that information about the presence of (for example) a face can *potentially* be extracted from brain activity shortly after stimulus onset, the current behavioral data goes further by demonstrating that such information can indeed be used by the brain to control behavior. Furthermore, while information can be extracted from intracortical potentials in humans from 100 ms (Liu et al., 2009), this does not imply that the information is necessarily already visible in neuronal firing. For example, in monkey IT, the earliest face-selective firing has been reported at latencies of 70–90 ms (Oram & Perrett, 1992), but deflections in intracerebrally recorded potentials are seen from as early as 50 ms (Schroeder, Mehta, & Givre, 1998). Thus, it might be that face-selective firing in the brain regions studied by Liu et al. might not occur until appreciably later, perhaps 120–30 ms, which would be well after the earliest face-selective saccades reported here.

While it may seem natural to assume that the processing required to initiate these fast responses involves the ventral cortical processing stream, it is important to realize that there is little reason to exclude subcortical processing pathways. For example, there is certainly evidence that face information can be processed in subcortical structures such as the amygdala, and there is evidence that visual information can reach the amygdala via the superior colliculus and the pulvinar (Johnson, 2005). Much of the evidence for subcortical processing has come from work with emotional faces and fear-inducing stimuli (Ohman, Carlsson, Lundqvist, & Ingvar, 2007), but it seems clear that the fast saccadic responses that we describe here are not restricted to faces with emotional expressions. Nevertheless, there is no strong argument for excluding the possibility that at least some of the information needed for face detection might have a subcortical origin.

One further result from the monkey neurophysiology that seems particularly important in the present context is the finding that the onset latencies of neurons in

inferotemporal cortex can vary significantly depending on the stimulus. Kiani, Esteky, and Tanaka (2005) reported that onset latencies of responses to primate and human faces are roughly 20 ms earlier than to faces of other animals. This difference was seen even though the total amount of activity is similar for the two types of stimulus, demonstrating that it is not simply that the neurons respond better to primate and human faces.

This result suggests an interesting possibility that could go some way toward explaining the remarkably rapid saccadic responses reported here. Suppose that when a face and a vehicle are simultaneously presented in the left and right visual fields, the neurons in the ventral stream contralateral to the face fire 20 ms earlier than the neurons responding to the vehicle. This may produce an imbalance in the levels of activation, which could result in differential activation in areas involved in saccade initiation such as the Frontal Eye Field (FEF) and the Lateral Intraparietal Area (LIP). It is as if the face would have a higher salience than other stimuli, simply because of this difference in onset latency.

If there is indeed a difference in the onset latency of neuronal responses to different types of stimuli, with the shortest latencies being seen for faces, then this might well explain why such short latency behavioral responses can be seen to faces. However, such an explanation would lead to the following hypothesis. Suppose that it is not possible to alter the latency of the neural responses under top-down control. In this case, we might expect that even if the subject was trying to direct their eyes toward the other stimulus, the latency advantage for faces would still result in a bias toward faces, at least for the fastest responses. This is precisely what we observed in [Experiment 2](#), where we noted that when the subjects were instructed to saccade toward the vehicle, they nevertheless tend to saccade toward the face, at least when the saccades were initiated in under 140 ms.

A difference between manual and saccadic tasks?

The strong bias toward responding to faces reported here is not something that we have seen in previous studies using manual responses. These earlier studies using similar visual stimuli reported that subjects could easily shift between one target category and another from block to block. This was seen for both animals and means of transport (VanRullen & Thorpe, 2001a) and for animal and human faces (Rousselet et al., 2003). However, in these studies using a manual go/no-go task, even the very fastest responses are never seen earlier than about 250–300 ms following stimulus onset, considerably longer than the saccadic responses reported here. This raises the possibility that the extra processing time available in the manual task allows the subjects to generate a behavioral response that is fully under top-down control. In contrast,

eye movements may be generated without allowing enough time for complete modulation of the behavior. This is clear from the results of [Experiment 2](#) in which participants tried to selectively make saccades toward the vehicle. This pattern of results is precisely what would be expected if the visual system had an in-built bias in favor of faces, which would be evident even under conditions where the task requires the participants to saccade elsewhere.

The latency at which the saccades start to be modulated by the task requirements could be related to neurophysiological studies of attentional effects on neuronal responses. In studies of visual responses, it has repeatedly been noted that the initial transient part of the response tends to be relatively fixed and that task-related modulations only start to be clear after a further delay (Roelfsema et al., 2007; Treue, 2001). It is therefore possible that the difference in latency between the earliest saccades to faces (100–110 ms) and the earliest point at which subjects can start to reliably make saccades toward the vehicle target (around 140–150 ms) could reflect the duration of this initial period of the neural response, which appears to be relatively insensitive to top-down task modulation.

Once this top-down modulation has started, a decision mechanism that depended on the amount of cumulated activity originating from the two parts of the visual field would progressively become more and more strongly biased toward the intended target. As a consequence, saccades that are initiated at relatively long latencies could be reliably made in the direction of the target, even when the target is a vehicle. In the case of behavioral responses that have even longer latencies, such as a manual go/no-go response, the accumulation of activity will be sufficient to ensure that the response can be made to whatever target category is currently in use.

The fact that the effectiveness of top-down task-dependent modulation is not fixed, but varies depending on the latency at which the saccades are initiated, means that the saccadic choice task can be used to track the time course of top-down effects, something that may be difficult or even impossible using conventional manual reaction time methodologies. By the time a manual response is initiated, the brain has had plenty of time to complete a wide range of operations, including attentional modulation. In contrast, the very fastest saccadic responses appear to occur before these modulatory effects had time to go to completion. Until now, the only methods that have been able to investigate this early time window (100–150 ms) have been techniques such as EEG, MEG, and single-cell recording, but the behavioral significance of such effects have been difficult to establish.

What makes faces special?

It appears that faces may be in a class of their own in their ability to trigger very fast saccades, and a natural

question concerns the origin of this advantage. One possibility is that faces are special because we have a great deal of expertise in processing faces from an early age. Indeed, faces occupy a very important part of the visual environment of the newborn child (Sinha, Balas, & Ostrovsky, 2007), and this increased exposure could lead to the development of selective mechanisms using unsupervised learning (see, for example, Masquelier & Thorpe, 2007). On the other hand, there is considerable evidence that humans have an innately specified bias in favor of face stimuli (Johnson, Dziurawiec, Ellis, & Morton, 1991; Johnson & Mareschal, 2001), and recent work suggests that this preference carries across to photographs of chimpanzees with which we have had far less experience (Tauber, 2009). Another recent study showed that in a change-blindness paradigm, we are much more likely to notice a change involving an animal than one involving a vehicle, even when the stimuli have been matched for size within the image (New, Cosmides, & Tooby, 2007). The authors interpreted their results as favoring an ancestral priority for animals, including humans. Irrespective of the origin of this face bias, it will undoubtedly have a major impact on the way in which humans explore the visual environment.

Selectivity mechanisms

The final point concerns the mechanisms that can allow face-selective responses to be produced so rapidly. Our ability to initiate directed saccades toward faces as early as 100–110 ms after stimulus onset clearly leaves little time for anything other than a feed-forward pass. This point is strengthened by the fact that reaction times in the choice task are only marginally longer than in a simple detection task (see [Experiment 3](#)), meaning that the processing overhead associated with detecting the presence of a face in an image can be no more than a few tens of milliseconds. It has been known for some time that a single wave of spikes can be sufficient not only for face detection (VanRullen, Gautrais, Delorme, & Thorpe, 1998) but also for face identification (Delorme & Thorpe, 2001). There is further recent evidence that purely feed-forward hierarchical processing mechanisms may be sufficient to account for at least some forms of rapid categorization (Serre et al., 2007). In the field of computer vision, considerable progress has been made in developing algorithms for detecting and localizing faces in natural images (Hjelmas & Low, 2001; Viola & Jones, 2004), some of which have found their way into consumer products such as digital cameras.

Indeed, the shortest saccadic reaction times that we report here appear to be so fast that there may even not be enough time to complete the first feed-forward pass through the ventral processing stream. Remember that since we need to allow around 20 ms for response initiation, it seems inevitable that face-selective mechanisms must

become activated from as early as 80 ms following stimulus onset. Even in monkeys, this would correspond to the earliest responses in inferotemporal neurons, but it needs to be remembered that monkeys are known to be substantially faster than humans on virtually all behavioral tasks, probably because their smaller heads lead to a reduction in conduction delays.

Given these constraints, it seems likely that the visual system will effectively try to make use of any available cues to the presence of a face in the image, especially those that can be extracted early on during visual processing. Some recent research suggests that this may indeed be the case. For example, Dakin and Watt (2009) have shown how the horizontal orientation structure in the human face provides a form of “bar code” that can be used for judgments of face identity. Further evidence for the use of very low level heuristics comes from another study from our laboratory, that also made use of the saccadic choice task, two images with varying contrast were presented on the left and right, and participants were required to saccade to the one with the highest contrast (Honey, Kirchner, & VanRullen, 2008). One of the images was a face, and the other one a vehicle—as in the experiments reported here. They found a very strong face bias in that even when the images had the same contrast, 70% of the saccades were made toward the faces. They then showed that at least some of this bias was still present even when the image was completely phase scrambled in the Fourier domain—that is, the images retained the same spatial frequency and orientation structure as the original images. This result supports the idea that rudimentary and quickly accessible information could explain a part of the bias we observed. Interestingly, a similar bias could also explain the tendency of faces to pop-out in multiple search arrays (Vanrullen, 2006). The underlying idea is that the visual system could use a simple heuristic characteristic of faces (for example, a pattern of high energy on both horizontal high and vertical low spatial frequencies), can be computed very rapidly, and used to generate useful selectivity early on in visual processing.

Conclusions

We have shown using a Saccadic Choice Task that humans can make very rapid saccades toward images containing faces. The earliest reliable saccades can be seen from as little as 100 to 110 ms after stimulus onset. While early face-selective electrical activity has been reported in a number of previous studies, this is the first time that it has been clear that such early selective responses could have a clear impact on behavior. The study also shows that the Saccadic Choice Task provides an experimental tool for studying very early processing in

the visual system, in a time window previously only accessible using electrophysiological techniques.

Acknowledgments

H. K. was supported by a European grant “Decisions In Motion” and by the ANR project “Hearing in Time”. S. M. C. is supported by a grant from the Délégation Générale pour l’Armement. The research was also financed by the CNRS and by the ANR “Natstats” Project.

Commercial relationships: none.

Corresponding author: Simon Thorpe.

Email: simon.thorpe@cerco.ups-tlse.fr.

Address: Centre de Recherche Cerveau and Cognition, CNRS, University Toulouse 3, Toulouse, France.

References

- Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neurosciences*, *8*, 551–565.
- Bindemann, M., Burton, A. M., Hooge, I. T., Jenkins, R., & de Haan, E. H. (2005). Faces retain attention. *Psychonomic Bulletin & Review*, *12*, 1048–1053. [PubMed]
- Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Gaze selection in complex social scenes. *Visual Cognition*, *16*, 341–355.
- Braeutigam, S., Bailey, A. J., & Swithenby, S. J. (2001). Task-dependent early latency (30–60 ms) visual processing of human faces and other objects. *Neuroreport*, *12*, 1531–1536. [PubMed]
- Brown, V., Huey, D., & Findlay, J. M. (1997). Face detection in peripheral vision: Do faces pop out? *Perception*, *26*, 1555–1570. [PubMed]
- Buswell, G. T. (1935). *How people look at pictures: A study of the psychology of perception in art*. Chicago: University of Chicago Press.
- Butler, S., Gilchrist, I. D., Burt, D. M., Perrett, D. I., Jones, E., & Harvey, M. (2005). Are the perceptual biases found in chimeric face processing reflected in eye-movement patterns? *Neuropsychologia*, *43*, 52–59. [PubMed] [Article]
- Cerf, M., Harel, J., Einhäuser, W., & Koch, C. (2008). Predicting human gaze using low-level saliency combined with face detection. *Advances in neural information processing systems* (vol. 20, pp. 241–248). Cambridge, MA: MIT Press.
- Cohen, J. D., & Tong, F. (2001). Neuroscience. The face of controversy. *Science*, *293*, 2405–2407. [PubMed]
- Crouzet, S., Kirchner, H., & Thorpe, S. J. (2008). Saccading towards faces in 100 ms: What’s the secret? *Perception*, *37*(Supplement), 119–120.
- Dakin, S. C., & Watt, R. J. (2009). Biological “bar codes” in human faces. *Journal of Vision*, *9*(4):2, 1–10, <http://journalofvision.org/9/4/2/>, doi:10.1167/9.4.2. [PubMed] [Article]
- Delorme, A., & Thorpe, S. J. (2001). Face identification using one spike per neuron: Resistance to image degradations. *Neural Networks*, *14*, 795–803. [PubMed] [Article]
- Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*, 2470–2473. [PubMed]
- Drewes, J., Trommershaeuser, J., & Gegenfurtner, K. R. (2009). The effect of context on rapid animal detection [Abstract]. *Journal of Vision*, *9*(8):1177, 1177a, <http://journalofvision.org/9/8/1177/>, doi:10.1167/9.8.1177.
- Eimer, M., & Holmes, A. (2002). An ERP study on the time course of emotional face processing. *Neuroreport*, *13*, 427–431. [PubMed]
- Einhauser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, *8*(2):2, 1–19, <http://journalofvision.org/8/2/2/>, doi:10.1167/8.2.2. [PubMed] [Article]
- Epshtein, B., Lifshitz, I., & Ullman, S. (2008). Image interpretation by a single bottom-up top-down cycle. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 14298–14303. [PubMed] [Article]
- 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. [PubMed]
- Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is “special” about face perception? *Psychological Review*, *10*, 482–498. [PubMed]
- Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences*, *16*, 553–610.
- Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*, *37*, 571–583. [PubMed]
- Freiwald, W. A., Tsao, D. Y., & Livingstone, M. S. (2009). A face feature space in the macaque temporal

- lobe. *Nature Neuroscience*, *12*, 1187–1196. [PubMed] [Article]
- 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. [PubMed]
- Gilchrist, I. D., & Proske, H. (2006). Anti-saccades away from faces: Evidence for an influence of high-level visual processes on saccade programming. *Experimental Brain Research*, *173*, 708–712. [PubMed]
- Guyonneau, R., Kirchner, H., & Thorpe, S. J. (2006). Animals roll around the clock: The rotation invariance of ultrarapid visual processing. *Journal of Vision*, *6*(10):1, 1008–1017, <http://journalofvision.org/6/10/1/>, doi:10.1167/6.10.1. [PubMed] [Article]
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–2430. [PubMed]
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*, 223–233. [PubMed]
- Hemond, C. C., Kanwisher, N. G., & Op de Beeck, H. P. (2007). A preference for contralateral stimuli in human object- and face-selective cortex. *PLoS One*, *2*, e574. [PubMed] [Article]
- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*, 498–504. [PubMed] [Article]
- Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Research*, *45*, 1707–1724. [PubMed] [Article]
- Hershler, O., & Hochstein, S. (2006). With a careful look: Still no low-level confound to face pop-out. *Vision Research*, *46*, 3028–3035. [PubMed] [Article]
- Hjelmas, E., & Low, B. K. (2001). Face detection: A survey. *Computer Vision and Image Understanding*, *83*, 236–274.
- Honey, C., Kirchner, H., & VanRullen, R. (2008). Faces in the cloud: Fourier power spectrum biases ultrarapid face detection. *Journal of Vision*, *8*(12):9, 1–13, <http://journalofvision.org/8/12/9/>, doi:10.1167/8.12.9. [PubMed] [Article]
- Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, *310*, 863–866. [PubMed]
- Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 9379–9384. [PubMed] [Article]
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506. [PubMed] [Article]
- Jacques, C., & Rossion, B. (2009). The initial representation of individual faces in the right occipito-temporal cortex is holistic: Electrophysiological evidence from the composite face illusion. *Journal of Vision*, *9*(6):8, 1–16, <http://journalofvision.org/9/6/8/>, doi:10.1167/9.6.8. [PubMed] [Article]
- Jeffreys, D. A. (1989). A face-responsive potential recorded from the human scalp. *Experimental Brain Research*, *78*, 193–202. [PubMed]
- Johnson, J. S., & Olshausen, B. A. (2003). Time course of neural signatures of object recognition. *Journal of Vision*, *3*(7):4, 499–512, <http://journalofvision.org/3/7/4/>, doi:10.1167/3.7.4. [PubMed] [Article]
- Johnson, J. S., & Olshausen, B. A. (2005). The earliest EEG signatures of object recognition in a cued-target task are postsensory. *Journal of Vision*, *5*(4):2, 299–312, <http://journalofvision.org/5/4/2/>, doi:10.1167/5.4.2. [PubMed] [Article]
- Johnson, M. H. (2005). Subcortical face processing. *Nature Reviews. Neuroscience*, *6*, 787–798. [PubMed]
- Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*, 1–19. [PubMed] [Article]
- Johnson, M. H., & Mareschal, D. (2001). Cognitive and perceptual development during infancy. *Current Opinion in Neurobiology*, *11*, 213–218. [PubMed] [Article]
- Kano, F., & Tomonaga, M. (2009). How chimpanzees look at pictures: A comparative eye-tracking study. *Proceedings of the Royal Society B: Biological Sciences*, *276*, 1949–1955. [PubMed]
- Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*, 759–763. [PubMed]
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302–4311. [PubMed] [Article]
- Kiani, R., Esteky, H., & Tanaka, K. (2005). Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. *Journal of Neurophysiology*, *94*, 1587–1596. [PubMed] [Article]

- Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, *46*, 1762–1776. [PubMed] [Article]
- Langton, S. R., Law, A. S., Burton, A. M., & Schweinberger, S. R. (2008). Attention capture by faces. *Cognition*, *107*, 330–342. [PubMed] [Article]
- Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron*, *62*, 281–290. [PubMed]
- Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, *5*, 910–916. [PubMed]
- Masquelier, T., & Thorpe, S. J. (2007). Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Computational Biology*, *3*, e31. [PubMed] [Article]
- McCarthy, G., Puce, A., Belger, A., & Allison, T. (1999). Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. *Cerebral Cortex*, *9*, 431–444. [PubMed]
- Mormann, F., Kornblith, S., Quiroga, R. Q., Kraskov, A., Cerf, M., Fried, I., et al. (2008). Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. *Journal of Neuroscience*, *28*, 8865–8872. [PubMed] [Article]
- Mouchetant-Rostaing, Y., & Giard, M. H. (2003). Electrophysiological correlates of age and gender perception on human faces. *Journal of Cognitive Neuroscience*, *15*, 900–910. [PubMed]
- Mouchetant-Rostaing, Y., Giard, M. H., Bentin, S., Aguera, P. E., & Pernier, J. (2000). Neurophysiological correlates of face gender processing in humans. *European Journal of Neuroscience*, *12*, 303–310. [PubMed]
- New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences*, *104*, 16598–16603. [PubMed] [Article]
- Ohman, A., Carlsson, K., Lundqvist, D., & Ingvar, M. (2007). On the unconscious subcortical origin of human fear. *Physiology Behavior*, *92*, 180–185. [PubMed] [Article]
- 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. [PubMed]
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, *42*, 107–123. [PubMed] [Article]
- Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, *45*, 2397–2416. [PubMed] [Article]
- Puce, A., Allison, T., Gore, J. C., & McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. *Journal of Neurophysiology*, *74*, 1192–1199. [PubMed]
- Qiu, F. T., Sugihara, T., & von der Heydt, R. (2007). Figure-ground mechanisms provide structure for selective attention. *Nature Neurosciences*, *10*, 1492–1499. [PubMed] [Article]
- Ro, T., Russell, C., & Lavie, N. (2001). Changing faces: A detection advantage in the flicker paradigm. *Psychological Science*, *12*, 94–99. [PubMed]
- Roelfsema, P. R., Tolboom, M., & Khayat, P. S. (2007). Different processing phases for features, figures, and selective attention in the primary visual cortex. *Neuron*, *56*, 785–792. [PubMed]
- Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *Neuroimage*, *39*, 1959–1979. [PubMed] [Article]
- Rousselet, G. A., Mace, M. J., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision*, *3*(6):5, 440–455, <http://journalofvision.org/3/6/5/>, doi:10.1167/3.6.5. [PubMed] [Article]
- Rousselet, G. A., Mace, M. J., & Fabre-Thorpe, M. (2004). Animal and human faces in natural scenes: How specific to human faces is the N170 ERP component? *Journal of Vision*, *4*(1):2, 13–21, <http://journalofvision.org/4/1/2/>, doi:10.1167/4.1.2. [PubMed] [Article]
- Rousselet, G. A., Mace, M. J., Thorpe, S. J., & Fabre-Thorpe, M. (2007). Limits of event-related potential differences in tracking object processing speed. *Journal of Cognitive Neuroscience*, *19*, 1241–1258. [PubMed]
- Schroeder, C. E., Mehta, A. D., & Givre, S. J. (1998). A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. *Cerebral Cortex*, *8*, 575–592. [PubMed]
- Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain: A Journal of Neurology*, *115*, 15–36. [PubMed]

- Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 6424–6429. [PubMed] [Article]
- Sinha, P., Balas, B., & Ostrovsky, Y. (2007). Discovering faces in infancy [Abstract]. *Journal of Vision*, *7*(9):569, 569a, <http://journalofvision.org/7/9/569/>, doi:10.1167/7.9.569.
- Tarr, M. J., & Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, *3*, 764–769. [PubMed]
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, *45*, 643–659. [PubMed] [Article]
- Taubert, J. (2009). Chimpanzee faces are “special” to humans. *Perception*, *38*, 343–356. [PubMed]
- Theeuwes, J., & Van der Stigchel, S. (2006). Faces capture attention: Evidence from inhibition of return. *Visual Cognition*, *13*, 657–665.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522. [PubMed]
- Thorpe, S., & Imbert, M. (1989). Biological constraints on connectionist modelling. In R. Pfeifer, Z. Schreter, F. Fogelman-Soulié, & L. Steels (Eds.), *Connectionism in perspective* (pp. 63–93). Amsterdam, The Netherlands: Elsevier.
- Treue, S. (2001). Neural correlates of attention in primate visual cortex. *Trends in Neurosciences*, *24*, 295–300. [PubMed] [Article]
- Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Review of Neuroscience*, *31*, 411–437. [PubMed] [Article]
- Vanrullen, R. (2006). On second glance: Still no high-level pop-out effect for faces. *Vision Research*, *46*, 3017–3027. [PubMed] [Article]
- VanRullen, R., Gautrais, J., Delorme, A., & Thorpe, S. (1998). Face processing using one spike per neurone. *BioSystems*, *48*, 229–239. [PubMed]
- VanRullen, R., & Thorpe, S. J. (2001a). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects. *Perception*, *30*, 655–668. [PubMed]
- VanRullen, R., & Thorpe, S. J. (2001b). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*, 454–461. [PubMed]
- VanRullen, R., & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Research*, *42*, 2593–2615. [PubMed] [Article]
- Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, *57*, 137–154.
- Vuilleumier, P. (2000). Faces call for attention: Evidence from patients with visual extinction. *Neuropsychologia*, *38*, 693–700. [PubMed] [Article]
- Yarbus, A. F. (1967). *Eye movements and vision*. New York: Plenum Press.