Fast saccades to faces during the feedforward sweep

Alison Campbell (ID)

Department of Psychology, University of Victoria, Victoria, BC, Canada ✉

James W. Tanaka (ID)

Department of Psychology, University of Victoria, Victoria, BC, Canada ✉

Saccadic choice tasks use eye movements as a response method, typically in a task where observers are asked to saccade as quickly as possible to an image of a prespecified target category. Using this approach, face-selective saccades have been observed within 100 ms poststimulus. When taking into account oculomotor processing, this suggests that faces can be detected in as little as 70 to 80 ms. It has therefore been suggested that face detection must occur during the initial feedforward sweep, since this latency leaves little time for feedback processing. In the current experiment, we tested this hypothesis using backward masking—a technique shown to primarily disrupt feedback processing while leaving feedforward activation relatively intact. Based on minimum saccadic reaction time, we found that face detection benefited from ultra-fast, accurate saccades within 110 to 160 ms and that these eye movements are obtainable even under extreme masking conditions that limit perceptual awareness. However, masking did significantly increase the median SRT for faces. In the manual responses, we found remarkable detection accuracy for faces and houses, even when participants indicated having no visual experience of the test images. These results provide evidence for the view that the saccadic bias to faces is initiated by coarse information used to categorize faces in the feedforward sweep but that, in most cases, additional processing is required to quickly reach the threshold for saccade initiation.

Introduction

Previous research has shown that saccadic eye movements toward faces can be elicited as quickly as 100 ms after image onset (Crouzet, Kirchner, & Thorpe, 2010; Crouzet & Thorpe, 2011; Di Oleggio Castello & Gobbini, 2015; Honey, Kirchner, & VanRullen, 2008). Rapid detection is demonstrated in saccadic choice tasks, where two images of different objects are presented side-by-side and participants are instructed to fixate as quickly as possible on a target from a prespecified category, such as an animal (Guyonneau, Kirchner, & Thorpe, 2006; Kirchner & Thorpe, 2006). When faces are targets, it has been shown that saccadic eye movements are highly accurate (typically around 90%) and faster compared to eye movements directed toward non–face objects, such as animals and vehicles (Crouzet et al., 2010; Crouzet & Thorpe, 2011). In fact, the fastest saccades (elicited within 150 ms of image onset) tend to be directed toward faces, even when they are not the intended target (Experiment 2, Crouzet et al., 2010; Fletcher-Watson, Findlay, Leekam, & Benson, 2008; Little, Jenkins, & Susilo, 2021), suggesting a strong bias toward human faces that generates especially fast saccadic responses. Given that oculomotor responses take around 20 to 35 ms to generate (Heeman, Van der Stigchel, & Theeuwes, 2017; Schiller & Kendall, 2004), the saccadic reaction times suggest that faces can be detected in as little as 70 ms.

Visual processing can be roughly divided into two stages of processing: early, bottom-up processing carried by feedforward activation during the first 150 ms, followed by a later stage of "reentrant" processing carried by feedback activation (Felleman & Van Essen, 1991; Kreiman & Serre, 2020; Lamme & Roelfsema, 2000; Martin, Cox, Scholl, & Riesenhuber, 2019; Ungerleider & Haxby, 1994; VanRullen & Thorpe, 2001). Ultra-fast saccadic latencies place strong constraints on models of face detection, since they imply that face selectivity can be accomplished before the completion of the first feedforward pass through the ventral processing stream. Critically, it leaves little time for feedback connections to exert an effect on visual processing (Lamme & Roelfsema, 2000). It has therefore been claimed that eye movements toward faces are triggered by visual face cues extracted during early feedforward processing (Crouzet et al., 2010; Crouzet & Thorpe, 2011; Honey et al., 2008). The first feedforward sweep is believed to enable the visual system to build a coarse representation that is iteratively refined during later recurrent and feedback processing (Kreiman & Serre, 2020; Thorpe, Fize, & Marlot, 1996; VanRullen, 2007). The initial representation is thought to provide a coarse structure of a stimulus, carried by low spatial

frequencies, before the fine details transmitted by high spatial frequency are processed (Bar, 2007; Goffaux et al., 2011; Hochstein & Ahissar, 2002; Marr, 1982). Computational (Riesenhuber & Poggio, 2002; Serre, Oliva, & Poggio, 2007) and neural (Cauchoix, Crouzet, Fize, & Serre, 2016; DiCarlo, Zoccolan, & Rust, 2012; Hong, Yamins, Majaj, & DiCarlo, 2016; Liu, Harris, & Kanwisher, 2002; VanRullen & Thorpe, 2002) work has indicated that these initial representations are sufficient for rapid categorization, suggesting that fast object categorization is based on coarse representations (Crouzet & Thorpe, 2011; Honey et al., 2008; VanRullen, 2006). Consistent with this, it has been found that observers are faster to orient to coarse, low spatial frequency faces compared to high spatial frequency faces (Guyader, Chauvin, Boucart, & Peyrin, 2017). However, the bias to saccade to faces relative to vehicles has been observed even when images were completely phase-scrambled, indicating that the amplitude spectrum information that still remains after phase-scrambling is an informative cue for rapid categorization (Honey et al., 2008; also Wichmann, Drewes, Rosas, & Gegenfurtner, 2010). In a follow-up study, Crouzet and Thorpe (2011) showed that normalizing amplitude spectrum information across face and vehicle images significantly reduced saccadic accuracy and reaction time to face targets, while having no effect on saccadic movements toward vehicles. However, overall, saccadic accuracy and speed still remained higher for faces over vehicles. This suggests that amplitude information is especially relevant for the early selectivity mechanisms for faces but that both amplitude and phase information contribute to the bias for faces. Most recently, it was shown that the saccadic bias for faces was obtained even when they were inverted or contrast reversed, although saccades were slower overall in these conditions (Little et al., 2021).

While Crouzet et al. (2010) have claimed that the saccadic latency observed for face detection indicates that it must be based on representations computed during the initial feedforward sweep, this hypothesis has not been directly tested. Here, we tested this hypothesis and predicted that selective eye movements to faces would be observed even when feedback processing is interrupted. One way that this may be accomplished is by presenting a second stimulus shortly after an initial target image to create an effect known as backward masking (Breitmeyer & Ogmen, 2000, Breitmeyer & Ogmen, 2006). Perceptual visibility of the initial target image is reduced as the stimulus onset asynchrony (SOA) between the target and masking image is decreased, and at a very short SOA (usually below 50 ms), backward masking can render a stimulus completely invisible (Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe, 2005; Del Cul, Baillet, & Dehaene, 2007; Fahrenfort, Scholte, & Lamme, 2007;

Fahrenfort, Van Leeuwen, Olivers, & Hogendoorn, 2017; Martin et al., 2019). Importantly, this effect has been attributed to a disruption to feedback processing. Electrophysiological data from primates (Cauchoix et al., 2016; Kovács, Vogels, & Orban, 1995; Lamme, Zipser, & Spekreijse, 2002) and humans (Bacon-Macé et al., 2005; Del Cul et al., 2007; Fahrenfort et al., 2007, Fahrenfort et al., 2017; Harris, Schwarzkopf, Song, Bahrami, & Rees, 2011; Martin et al., 2019) support the view that backward masking largely disrupts feedback processing while leaving feedforward processing mostly intact. For example, Fahrenfort et al. (2007) found that masking visual targets had no effect on early occipitotemporal electrophysiological responses observed at approximately 110 ms poststimulus, but it abolished a later occipitotemporal response occurring from 180 to 305 ms. Masking also reduced target detection to chance performance. This is consistent with the current understanding that visual awareness of a stimulus critically depends on recurrent processing in the feedback period (Boehler, Schoenfeld, Heinze, & Hopf, 2008; Camprodon, Zohary, Brodbeck, & Pascual-Leone, 2010; Del Cul et al., 2007; Fahrenfort et al., 2017; Haynes, Driver, & Rees, 2005; Koivisto, Railo, Revonsuo, Vanni, & Salminen-Vaparanta, 2011; Koivisto, Salminen-Vaparanta, Grassini, & Revonsuo, 2016; Lamme, 2010; Lamme, Supèr, Landman, Roelfsema, & Spekreijse, 2000; Lamme & Roelfsema, 2000; Martin et al., 2019; Pascual-Leone & Walsh, 2001; Ro, Breitmeyer, Burton, Singhal, & Lane, 2003).[1]

More recently, Martin et al. (2019) examined the nature of the neural interference between two successive stimuli with varying intervals between target images. When target images (animal images) were presented approximately 400 ms apart, each stimulus evoked a distinct pattern of EEG activation in posterior channels corresponding to an early feedforward response 150 ms poststimulus and a later feedback response 230 ms poststimulus. However, when the interval between the two targets was reduced, there was an increased overlap between the feedback processing of the first target and the feedforward processing of the second, as well as a greater cost to the behavioral detection of the first target compared to the second. Critically, this interference was reduced when targets were presented in different halves of the visual field to segregate their neural responses in separate hemispheres. These results reveal how feedback processing of the initial image "crashes into" the incoming feedforward signal from the mask and are consistent with an interruption theory of backward masking (Bridgeman, 1980; Di Lollo, Enns, & Rensink, 2000; Fahrenfort et al., 2007, Fahrenfort et al., 2017; Kovács et al., 1995; Lamme et al., 2002). Accordingly, the technique of backward masking has been said to be particularly useful for emphasizing bottom-up processing (Kreiman & Serre, 2020) and to "isolate between feed-forward dominated

versus recurrent processing" (Serre, Kreiman, et al., 2007).

In the current study, we tested the hypothesis that feedforward activation from face images would be sufficient to elicit fast saccadic responses toward faces in a saccadic choice task (Crouzet et al., 2010; Kirchner & Thorpe, 2006). Backward masking was used to interrupt feedback processing and to constrain visual processing of the test images to the initial feedforward pass. Given that category information is primarily carried by phase information (Bar, 2004; Keil, 2008; Oppenheim & Lim, 1981; VanRullen, 2006) and that phase information is the main driver of ultra-rapid face selective responses (Crouzet & Thorpe, 2011), we created masking stimuli by phase-scrambling the test images. Three target-mask SOA conditions were examined: 8 ms, 50 ms, and 400 ms. The 8-ms SOA was chosen because pilot testing showed that visual awareness of the test images was almost entirely suppressed; this SOA therefore allowed us to test the hypothesis that feedforward activation of a face image is sufficient to elicit a saccadic response toward it, even in the absence of conscious perception. However, based on previous findings of the timing of feedforward and feedback processing in posterior occipitotemporal areas, we predicted that feedforward processing of the mask would maximally interfere with the feedback processing of the target at an SOA of 50 ms. At this SOA, the feedforward activation elicited by the mask should occur 150 to 200 after the onset of the target image, thus putting it within the time period when long-range feedback connections are being established for target processing (Fahrenfort et al., 2007; Martin et al., 2019). This SOA allowed us to test the hypothesis that ultra-rapid saccades may be generated even when there is strong neural interference between the target images and the mask but without completely suppressing visual awareness. Finally, we examined saccadic responses with a target-mask SOA of 400 ms. This masking condition was intended to approximate the viewing conditions used in the original Crouzet et al. (2010) study where images were presented for 400 ms. Although a masking stimulus is presented in this condition, the test images are effectively not masked due to the long SOA.

In addition to recording eye movements during the target detection tasks, we also asked participants to manually indicate on which side of the screen the target appeared. This provided an objective measure of target detection and allowed us to examine whether participants can report target location with and without conscious perception. Lastly, participants were asked to provide a subjective visibility rating on each trial to examine the extent to which target-mask SOA affected conscious perception and to confirm whether selective eye movements toward faces (and potentially houses) could be executed independent of subjective perceptual experience.

Under viewing conditions that were comparable to the original study by Crouzet et al. (2010), we replicated the saccadic response profile for faces when they were targets as well as when they were distractors. These saccadic responses were both faster than for house targets and harder to control. Critically, these eye movements were obtainable even under extreme masking conditions that limited perceptual awareness, suggesting that the saccadic bias to faces is initiated by coarse information in the feedforward sweep.

Methods

Participants

Twelve participants recruited from our university's psychology research participation pool successfully completed the experiment. Six additional participants did not display the expected face bias on the 400-ms target-mask SOA trials and were therefore excluded from the final sample (cf. Honey et al., 2008). Of those six, three were excluded due to a median saccadic reaction time over 600 ms and three were excluded based on saccadic response accuracy below 75%. By comparison, with intact images, observers typically show > 90% saccadic accuracy and very few saccades are observed beyond 200 ms (Crouzet et al., 2010). Because these exclusions were based only on the behavior observed on the 400-ms SOA trials, which are effectively unmasked due to the late onset of the mask, we interpret these data to reflect poor participant compliance and not an effect of masking. Although the saccadic face bias is a very strong effect, the proportion of participants who were excluded is likely due to the challenging nature of this version given that perceptual awareness was low on most trials and the brief presentation duration of the images.

The final sample ($N = 12$) ranged from 19 to 31 years of age ($M = 23.8$, $SD = 3.6$, 11 self-report females, 1 self-reported male). All had normal or corrected-to-normal vision. This sample size is comparable to those used in previous studies (Crouzet et al., 2010; Crouzet & Thorpe, 2011; Honey et al., 2008) and was confirmed in our pilot studies to provide stable saccadic response time (SRT) distributions and replicable minimum SRT estimates.

Stimuli

Figure 1 shows sample stimuli of the face, house, and masking images. We used a total of 100 grayscale photographic images taken from an existing database

Figure 1. Faces (A) and houses (B) appeared in the center of the test images and varied in size, viewpoint, lighting, and background. Images were normalized for mean pixel luminance and root-mean-square contrast. Masking stimuli (C) were created for each trial by first phase scrambling each of the test images and then merging the scrambled images into a single image that was used to mask both test images (i.e. the image used to mask the face was the same as the image used to mask the house). The average of the face stimuli (D), the average of the house stimuli (E) and average face minus average house (F) illustrates the low spatial frequency bias that is typical for faces.

of natural scene images (Rossion, Torfs, Jacques, & Liu-Shuang, 2015) with either a face (50 images) or a house (50 images) appearing in the center but that differed in terms of size, viewpoint, lighting, and background. This variability ensures that effects are category specific rather than image specific and that we replicate the same image-invariant effects obtained by Crouzet et al. (2010). Images were normalized for mean pixel luminance and root-mean-square contrast. Masking stimuli were adaptively created for each pair of test images presented during the experiment. For each image in a trial, scrambled versions were created by replacing the phase by random coefficients (Rossion et al., 2015). We then merged the scrambled versions into a single image using alpha blending, so that the amplitude spectrum of the resultant image is a combination of the spatial frequency content of each test image and in equal measure. This single final image was used to mask both the face and house image. All image modifications were done using MATLAB. Presented at a distance of 80 cm, the stimuli subtended approximately 14° of visual angle and were presented so that the center of the image was 4° horizontally away from the center of the screen. Images were presented on a gray background. For each participant, we randomly generated 50 image

pairs with one face image and one house image in each pair.

Apparatus

Participants viewed the stimuli in a dimly lit room with their head in a chinrest to constrain head movements and maintain a viewing distance of 80 cm. Stimuli were displayed on a 25-in. Dell Alienware (AW2521HF) gaming monitor with the screen resolution set to 1,920 × 1,080 pixels and a refresh rate of 240 Hz. The experiment was written in MATLAB, using the Psychophysics Toolbox 3 extension (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997).

Procedure

The experiment was divided into two halves, starting with either a face detection task (faces as targets, houses as distractors) or the house detection task (houses as targets, faces as distractors).

The detection tasks combined a two-alternative forced choice (2AFC) saccadic choice response, a manual response, and a perceptual awareness rating
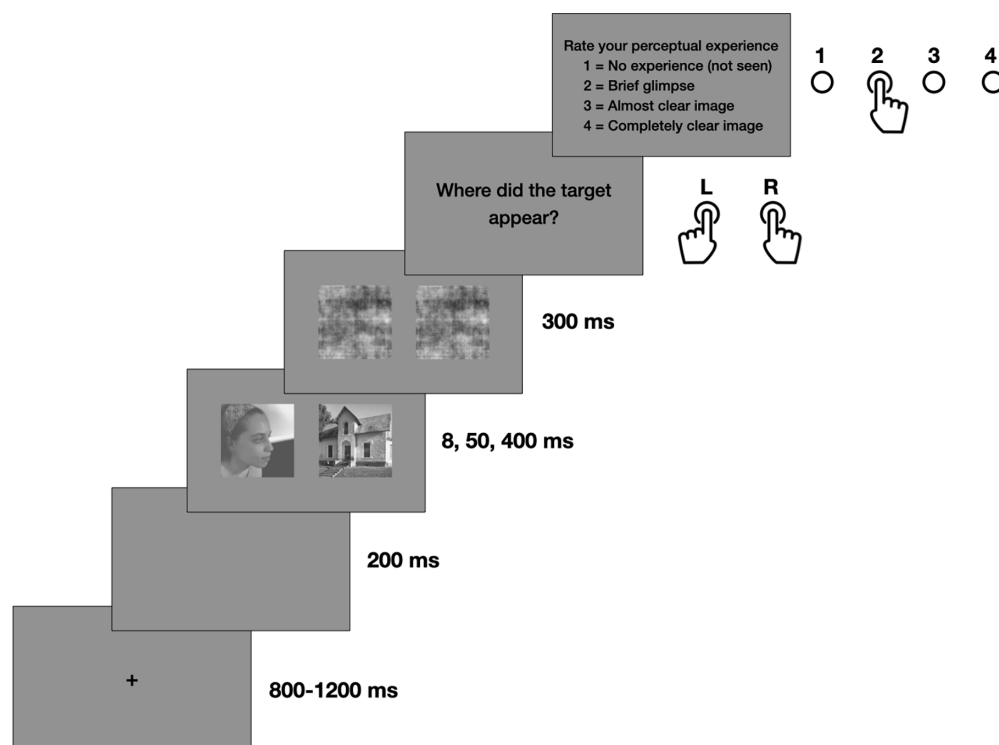
Figure 2. On each trial, a fixation cross appeared for 800–1200 ms, followed by a 200 ms interval, then the target and distractor images appeared for either 8, 50, or 400 ms. A masking image was presented immediately after for 300 ms. Participants were then prompted to indicate by manual response where they saw the target image and to rate their perceptual experience.

([Figure 2](#)). We used the Perceptual Awareness Scale ([Ramsøy & Overgaard, 2004](#)) as a purely introspective measure to examine the quality of participants' conscious perception of the test stimuli on each trial. This 4-point scale includes (1) no experience, (2) brief glimpse, (3) almost clear image, and (4) absolutely clear image and has been shown to have better correspondence to performance compared to other measures of visual awareness, including confidence ratings ([Sandberg, Timmermans, Overgaard, & Cleeremans, 2010](#)). Each trial consisted of the following:

1. A central fixation cross appeared for 800 to 1,200 ms.
2. After a 200-ms gap, an image pair was displayed left and right of the screen center for 8 ms, 50 ms, or 400 ms.
3. Images were replaced by the phase-scrambled composite of each image in the image pair for 300 ms.
4. Instructions appeared to prompt participants to manually indicate which side of the screen the target appeared using the F and J keys on the keyboard (until response).
5. Instructions appeared to prompt participants to manually rate their perceptual experience using

the top number keys 1, 2, 3, or 4 on the keyboard (until response).

Participants were told that their main task was to look as quickly and as accurately as possible to the side containing the face (face detection task) or house (house detection task). To reduce conflicts in motor response planning, participants were told that manual response speed was not important and that they could not respond until after the masking stimuli were removed from the screen and the manual response probe was presented. Each trial was followed by a 1,000-ms black intertrial interval. For each detection task, each participant performed six blocks of 50 trials.

Eye movement recording

Eye movements were recorded with an SR Research EyeLink 1000 system at a sampling rate of 2000 Hz using a 35-mm lens and a 940-nm infrared illuminator. Saccade detection was performed offline using Eyelink's built-in algorithm with standard cognitive thresholds for velocity (30°/s), acceleration (8,000°/s^2), and motion (0.1°). For each trial, the onset of the first saccade after stimulus onset before the manual response probe was considered the SRT. Trials with saccades onsets faster

than 70 ms were considered anticipatory responses and discarded. A 9-point calibration was performed before each detection task.

Data analysis

Minimum SRTs were determined by dividing the SRT distribution for each task and SOA condition into 10-ms time bins (i.e., the 100-ms bin contained latencies from 100 to 109 ms) and performing a chi-square test to determine whether it contained significantly more correct than incorrect responses ($p < 0.05$). Saccades were considered accurate if they were directed toward the intended target. If five consecutive bins were found to be significantly accurate, the first was considered to correspond to the minimum reaction time. Minimum SRTs were obtained from the SRT distributions pooled across all observers (Crouzet et al., 2010; Crouzet & Thorpe, 2011; Honey et al., 2008).

Results

We first confirmed that the effect of the target-mask SOA on subjective visibility using a 2 × 2 repeated-measures analysis of variance (ANOVA) of the perceptual awareness ratings. This showed a main effect of target-mask SOA, $F(2, 22) = 120.03$, $p < 0.001$, $\eta^2 = 0.83$, and post hoc tests showed that ratings significantly differed between all target-mask SOA conditions (all $p < 0.001$, Bonferroni–Holm corrected). As shown in Figure 3, the perceptual ratings indicate that the masking technique was effective in reducing the subjective visibility of the target images: On most trials, participants reported a "brief glimpse" in the 50-ms SOA condition ($M = 2.39$ [2.36, 2.40] 95% bootstrap
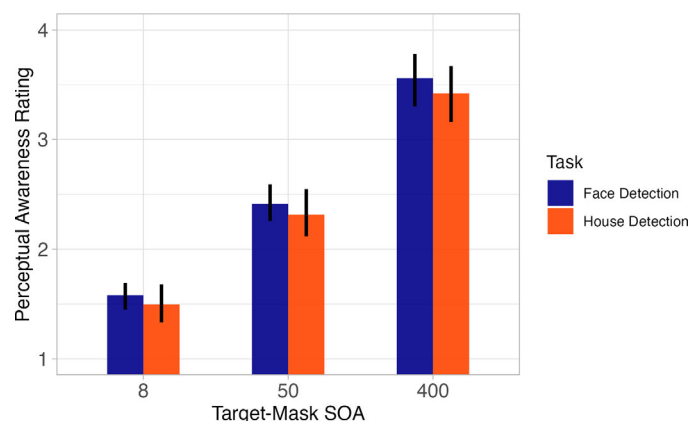

Figure 3. Average ratings on the Perceptual Awareness Scale for each condition and task based on participant responses collected on each trial.
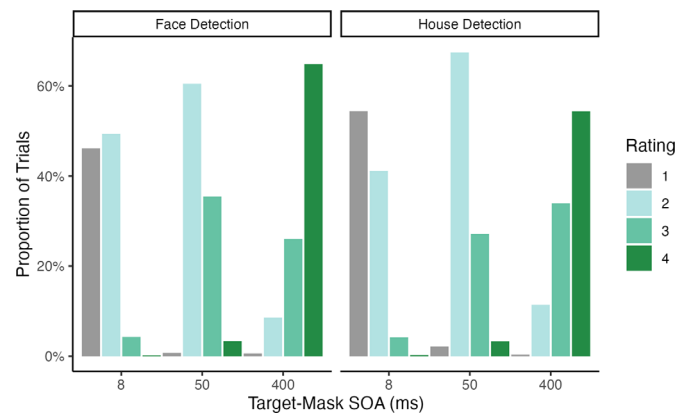

Figure 4. Distribution of responses on the Perceptual Awareness Scale (1 = No experience, 2 = Brief glimpse, 3 = Almost clear image, 4 = Absolutely clear image) for target-mask SOA condition and for each task.

confidence interval) and "no visual experience" in the 8-ms SOA condition ($M = 1.58$ [1.56 1.61]). These were both significantly lower than ratings in the 400-ms SOA condition, which was intended to approximate an unmasked condition and for which participants reported "completely clear" perceptual awareness of the test images on most trials ($M = 3.48$ [3.45, 3.50]). Surprisingly, visibility was not entirely abolished in the 8-ms SOA condition, as participants reported experiencing a brief glimpse of the images (rating 2) on 54% to 41% of trials during the face and house detection tasks, respectively (see Figure 4).

Backward masking significantly reduced the rate of saccadic response. A 2 × 2 repeated-measures ANOVA indicated a main effect of target-mask SOA, $F(2, 18) = 86.2$, $p < 0.001$, $\eta^2 = 0.76$, with significant differences between all conditions (all $p < 0.05$, Bonferroni–Holm corrected). Saccades were observed on 74% of trials ($n = 1,782$) in the 400-ms SOA conditions, but on only 19% ($n = 455$) and 10% ($n = 206$) of trials in the 50-ms and 8-ms SOA conditions, respectively.[2] Pairwise comparisons of the average perceptual awareness rating for trials with and without saccades did not indicate any significant difference in visibility for trials on which a saccade was recorded (Wilcoxen signed-rank tests, all ps > 0.10).

Manual response accuracy across all conditions was very high (Table 1), with accuracy ranging from 98% to 99% for both face and house detection in the 50-ms and 400-ms SOA conditions. Manual response accuracy was also significantly above chance in the 8-ms SOA conditions for both face ($M = 84.1\%$) and house ($M = 83.5\%$) detection as indicated by Wilcoxen signed-rank tests (both $p < 0.001$). Surprisingly, detection remained above chance for trials on which participants provided a rating of 1 (i.e., no visual experience) for both face (69.4% [62.7%, 78.5%], $p = 0.003$) and house

Task	SOA	Manual accuracy (%)	Saccadic accuracy (%)	Minimum SRT (ms)	Median SRT
Face detection	400 ms	99.1 [96.8, 99.7]	95.7 [94.2, 97.0]	120 [110, 120]	177 [175, 180]
	50 ms	98.9 [97.2, 99.6]	86.1 [81.3, 90.4]	130 [120, 160]	278 [271, 286]
	8 ms	84.1 [80.8, 88.5]	83.3 [76.9, 88.6]	140 [120, 150]	255 [234, 266]
House detection	400 ms	99.4 [98.8, 99.7]	82.7 [80.1, 85.2]	190 [180, 200]	235 [230, 242]
	50 ms	99.5 [98.9, 99.8]	78.4 [72.5, 83.8]	250 [220, 280]	303 [292, 314]
	8 ms	83.5 [77.2, 89.3]	63.5 [52.7, 74.3]	290 [210, 300]	246 [211, 267]

Table 1. Target detection accuracy based on saccadic and manual response, as well as median and minimum saccadic reaction time (SRT) for each detection task and target-mask stimulus-onset asynchrony (SOA) condition. 95% bootstrapped confidence intervals for each task and target-mask SOA in parentheses.
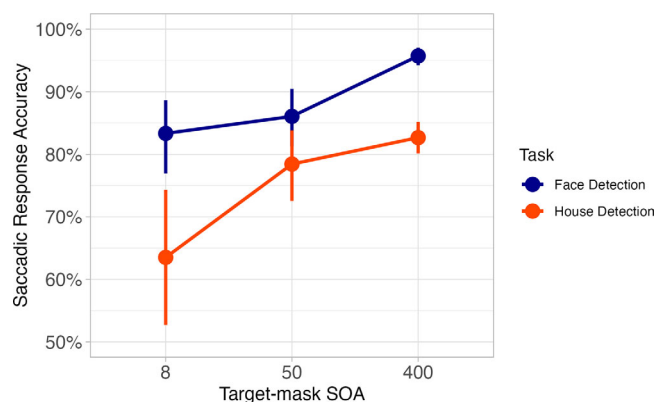

Figure 5. Mean saccadic accuracy for each target-mask SOA and task. Error bars represent 95% bootstrapped confidence intervals.
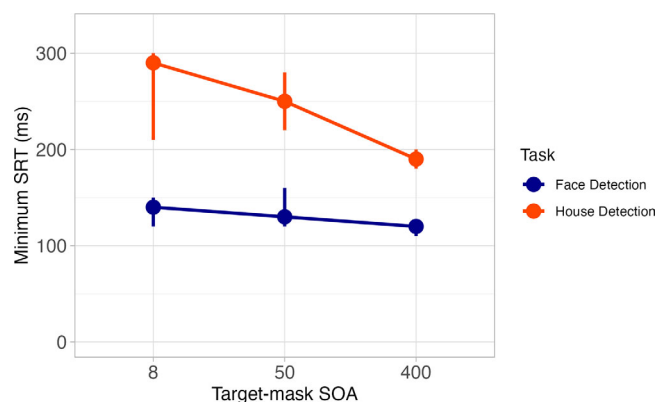

Figure 6. Minimum saccadic reaction time (SRT) for each target-mask SOA and task. Error bars represent 95% bootstrapped confidence intervals.

Figure 7. Median saccadic reaction time (SRT) for each target-mask SOA and task. Error bars represent 95% bootstrapped confidence intervals.

(75.3% [69.1%, 81.6%], $p < 0.001$) detection. A 2×2 repeated-measures ANOVA indicated a main effect of target-mask SOA on accuracy, $F(2, 22) = 42.97$, $p < 0.001$, $\eta^2 = 0.66$, and post hoc tests showed that accuracy in the 8-ms condition was reliably different from accuracy in the two other condition (both $p < 0.001$, Bonferroni–Holm corrected).

The main purpose of our study was to investigate whether the ultra-fast saccades evoked by faces typically observed under normal viewing conditions can escape the disruptive effects of backward masking. We therefore used a standard procedure to estimate the accuracy and reaction times of saccadic responses based on saccadic distributions pooled across all observers (Crouzet et al., 2010; Crouzet & Thorpe, 2011; Honey et al., 2008). Because the 400-ms target-mask SOA is intended to approximate an unmasked condition, we compared the average accuracy for this condition against the 95% bootstrapped confidence interval of the 50-ms and 8-ms target-mask SOA conditions (Figure 5). Face detection accuracy in both the 50-ms (86.1% [81.3%, 90.4%]) and 8-ms (83.3% [76.9%, 88.6%]) conditions was lower than accuracy in the 400-ms (95.7% [94.2%, 97.0%]) condition, although accuracy remained wel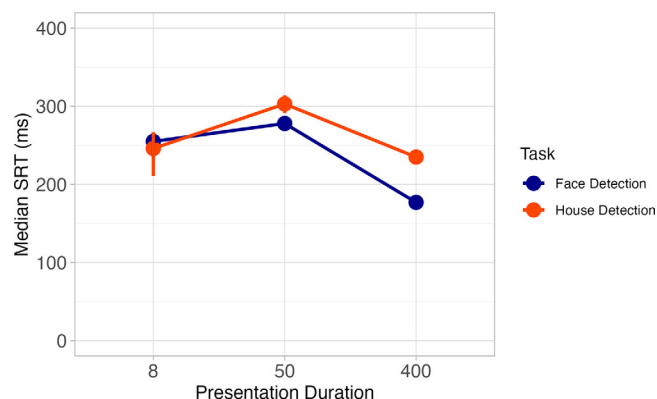l above chance even with strong masking. For house detection, only the 8-ms (63.5% [52.7%, 74.3%]) SOA condition was reliably different in saccadic accuracy from the 400-ms (82.7% [80.1%, 85.2%]) condition, although accuracy also remained above chance. As shown in Figure 5, saccadic response to faces was reliably more accurate than for houses in the 400-ms and the 8-ms target-mask SOA conditions.
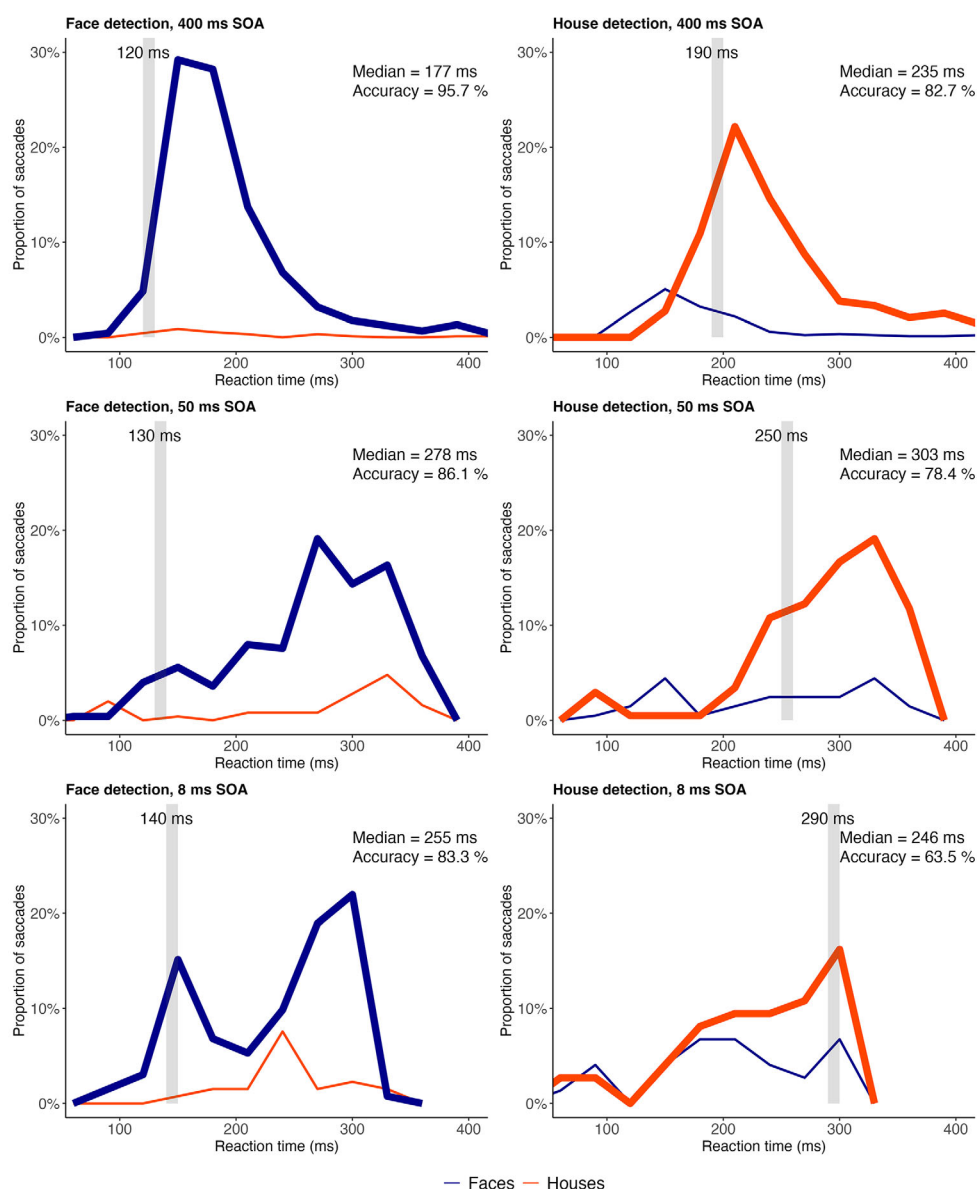
Figure 8. Distributions of saccadic reaction time pooled across observers in the face detection task (left column) and house detection task (right column). These show the relative proportion of saccades to targets (thick lines) and distractors (thin lines) as a function of saccadic latency. The minimum SRT at which saccadic accuracy is reliably greater than chance is indicated by the gray vertical bars.

We then examined the minimum saccadic response times for each condition (Table 1). The minimum SRT represents the first 10-ms time bin in which the cumulative number of correct responses is significantly greater than the number of incorrect responses (chi-square test). We again compared minimum SRT values based on the 95% bootstrapped confidence intervals for data pooled across all observers. As shown in Figure 6, the minimum SRT for face detection remained fast across all target-mask SOA conditions, as the minimum SRT for the 8-ms (140 ms [120, 150]) and 50-ms (130 ms [120, 160]) conditions were not

reliably different from the minimum SRT obtained in the 400-ms (120 ms [110, 120]) condition. By contrast, the minimum SRT for house detection was reliably slower in both the 50-ms (250 ms [220, 280]) and 8-ms (290 ms [210, 300]) SOA conditions compared to the 400-ms condition (190 ms [180, 200]).

Although the minimum time to saccade to faces was comparable across masking conditions, the median saccadic reaction times indicated that backward masking did have an effect on face detection (Figure 7). The median SRT for both the 8-ms (255 ms [234, 266]) and 50-ms (278 ms [271, 286]) SOA conditions was

reliably slower than the median SRT for the 400-ms SOA condition (177 ms [175, 180]). Notably, although the median SRT to detect faces was faster than the median SRT to detect houses in both the 400-ms and 50-ms SOA conditions, the median SRT was not different for faces (255 ms [234, 266]) or houses (246 ms [211, 267]) in the shortest 8-ms SOA condition.

Finally, because accuracy and reaction time estimates for each condition are based on a different number of trials (with the fewest trials in the 8-ms masking condition), we simulated 500 samples with an equal number of trials within each condition and conducted these same analyses. The estimates were remarkably stable and replicated the same pattern of results obtained with the full data set.

The reaction time distributions for correct and incorrect responses in each task are shown in Figure 8. Clear differences can be seen for saccadic responses to faces and houses in the 400-ms SOA condition (Figure 8, top row). For faces, most saccadic responses were initiated within 100 to 250 ms, but for houses, the distribution was shifted and spread over 150 to 300 ms. It is also clear that the fastest saccades tended to be directed toward faces, regardless of the task: In the face detection task, the earliest saccades were directed toward the face targets with almost none toward the house distractors, and in the house detection task, the earliest saccades were directed toward the face distractors with relatively fewer toward the correct house targets. This provides a time window of interest when examining the SRT distributions of the 50-ms and 8-ms SOA conditions for fast, feedforward face detection. There, we see that the earliest face-selective saccades also occurred within this time frame in the stronger masking conditions (130 ms in the 50-ms SOA condition and 140 ms in the 8-ms condition), as indicated by the early difference in the number of correct saccades to face targets and incorrect saccades to house distractors.

For house detection, there again appeared to be an early bias toward face distractors within 120 to 190 ms in the 50-ms masking condition, followed by later selectivity for houses at 250 ms. In the 8-ms SOA condition, saccade direction appeared to be at chance for house detection until 290 ms, when saccades to house targets significantly outnumbered saccades to face targets. Overall, the SRT distributions are consistent with the hypothesis of an early process that is more efficient for detecting faces.

Discussion

Saccadic choice tasks have repeatedly shown that face detection occurs faster and reflexively compared to that for non–face objects and that detection can occur

within 100 ms (Crouzet et al., 2010; Crouzet & Thorpe, 2011; Di Oleggio Castello & Gobbini, 2015; Guyader et al., 2017; Little et al., 2021). Although the timing of the saccadic response to faces strongly implies that detection must occur during feedforward processing, this hypothesis had not yet been directly tested. Here, we examined the saccadic response profile of faces and houses using a backward masking procedure that was intended to disrupt feedback processing and conscious perception. We found that backward masking greatly reduced the number of eye movements elicited by the stimuli, but when saccades were observed, we observed ultra-fast saccades directed toward faces but not houses, similar to those generated under effectively unmasked conditions. We therefore conclude that, although feedback processing typically contributes to normal saccadic responses, ultra-fast saccades to faces can be elicited and are likely initiated during the feedforward sweep.

We used three levels of masking: a strong masking condition with an 8-ms target-mask SOA, a moderate masking condition with a 50-ms target-mask SOA, and a 400-ms target-mask SOA that approximates unconstrained viewing due to the extended delay of the mask onset. Perceptual awareness ratings confirmed that masking reduced subjective visibility, with participants reporting only a brief glimpse on the majority of trials with a 50-ms SOA, yet manual responses showed that categorization was still at ceiling. Thus, even with reduced visibility, sufficient information was available for accurate categorization. Perceptual awareness was not completely abolished in the stronger 8-ms SOA masking condition, but test images were reported to be invisible on roughly half the trials. Despite this substantial loss of visibility, accuracy remained remarkably high for both face and house detection and even when observers reported no visual experience.

Our findings show that just 8 ms of exposure to a face stimulus is capable of eliciting ultra-fast, face-selective saccades originally observed by Crouzet et al. (2010), who presented face stimuli for 400 ms. In the 400-ms SOA condition, the minimum SRT (the earliest saccade latency for above-chance accuracy) for faces was between 110 and 120 ms, which is within the 100- to 135-ms range observed in previous studies (Crouzet et al., 2010; Di Oleggio Castello & Gobbini, 2015; Little et al., 2021). In the 8- and 50-ms SOA conditions, the minimum SRT ranged from 120 to 160 ms, although these estimates were not reliably different from the minimum SRT in the 400-ms SOA condition. These early saccades were also highly accurate, with almost none of the saccades in the early SRT distribution directed toward houses. In other words, saccades made within 160 ms were highly selective for faces and were observed even within strong backward masking.

For house detection in the 400-ms SOA condition, the minimum SRT was between 180 and 200 ms and similar to the 170- to 200-ms range observed for vehicles when faces are distractors (Crouzet et al., 2010; Little et al., 2021). The minimum SRT was slower in the 8-ms and 50-ms masking conditions, indicating that the processes needed for fast object detection were sensitive to backward masking. Notably, we observed a bias for the earliest saccades to move toward the face distractors in both the 50-ms and 400-ms SOA conditions. This provides further evidence that the fastest saccades are both selective for faces and automatic.

Overall, the results show a differential effect of backward masking on the fastest saccades to faces compared to houses. Specifically, backward masking did not affect the minimum SRT for faces, but it significantly slowed the minimum SRT for houses. However, although our findings indicate that face-selective saccades *can* escape backward masking, two different saccade distributions seemed to emerge in the 50-ms and 8-ms SOA conditions (Figure 8). This is consistent with the finding that, for faces, the minimum SRT was not reliably different across the masking conditions, but the median SRT increased as the target-mask SOA decreased. One interpretation is that the saccadic bias to faces is initiated by coarse information in the feedforward sweep but that in most cases, saccadic response also relies on additional recurrent processing. This also aligns with the coarse-to-fine processing account for faces (Goffaux et al., 2011; Petras, Jacobs, et al., 2019; Petras, Ten Oever, et al., 2019; Schuurmans, Bennett, Petras, & Goffaux, 2023), whereby face representations are gradually built up from initial low spatial frequency (LSF) information followed by the integration of high spatial frequency (HSF) information during recurrent processing. For example, some of our face stimuli might have been more easily categorized as a face based on their LSF information, whereas the signal might not have been as strong in others due to lighting or viewpoint. Under normal viewing conditions, the initial LSF information might be expected to efficiently integrate additional information needed to quickly reach the threshold for saccade initiation, but our masking might have slowed that process. This could also explain why there were significantly fewer saccades in our 50-ms and 8-ms SOA conditions. Regardless, it is still clear that only faces benefited from ultra-fast, accurate saccades within 110 to 160 ms and that these particular eye movements are obtainable even under extreme masking conditions.

To the extent that backward masking disproportionately affects feedback and recurrent processing (Bacon-Macé et al., 2005; Cauchoix et al., 2016; Fahrenfort et al., 2007, Fahrenfort et al., 2017), the observed ultra-fast saccades to masked faces

are consistent with early face detection during the feedforward sweep. Given that feedforward processing first establishes a coarse representation carried by LSF content (Bullier, 2001; Hochstein & Ahissar, 2002; Marr, 1982), the interpretation of our results implies that the speed advantage for faces reflects a use of LSF that is specific (or at least more informative) for faces. This account is directly supported by evidence that subjects orient faster to faces filtered for LSF compared to those filtered for HSF (Guyader et al., 2017) and that neural responses reflecting automatic face detection emerge with a minimal amount of spatial frequency content (Quek, Liu-Shuang, Goffaux, & Rossion, 2018). This may be due in part to the physical nature of faces themselves, as natural face images contain more energy in the LSF bands than other objects (Torralba & Oliva, 2003). Furthermore, given that LSF is defined by luminance variations over larger spatial scales, the regularities in LSF across individual faces may provide the basis for the formation of a general face template that is activated when visual stimuli match the spatial structure of a face (Goold & Meng, 2016) and may underlie face pareidolia (the tendency to "see" faces in visual patterns; Caharel et al., 2013) and holistic face perception (Goffaux & Rossion, 2006).

Beyond these representational differences, ultra-fast face detection has also led some to propose a shortcut in the visual system for fast-tracking face detection (Crouzet et al., 2010; Crouzet & Thorpe, 2011; Honey et al., 2008). For example, Campana et al. (2020) found evidence for face-selective responses in V1/V2 within 40 ms of stimulus onset and speculated that such representations could initiate fast motor responses via connections from early visual areas to the superior colliculus (Sherman, 2016). Given that oculomotor responses take around 20 to 35 ms to generate (Heeman et al., 2017; Schiller & Kendall, 2004), face-selective eye movement within 100 to 150 ms poststimulus puts strong constraints on the extent of cortical processing that can occur during detection. The earliest face-selective EEG component, the N170, begins to emerge around 130 ms poststimulus and is thought to originate from the occipital face area and the fusiform face area in the ventral occipital temporal cortex (Jacques et al., 2019; Rossion & Jacques, 2008). Although the face selectivity of this component is dependent on LSF information (Goffaux, Gauthier, & Rossion, 2003; Goffaux, Jemel, Jacques, Rossion, & Schyns, 2003), it is not clear whether it emerges early enough to mediate ultra-fast saccades. Most recently, Schuurmans et al. (2023) examined the processing of intact face images and found that V1 mediates the integration of HSF information after the initial representation of LSF information. Combined with the 40-ms response latency reported by Campana et al. (2020), face representation in V1 is an intriguing candidate for

the initiation of face detection and warrants further investigation.

The second major finding was the novel evidence for accurate face detection in the absence of conscious report. In the saccadic response measure, we observed a similar saccadic response profile for clearly visible faces as we did for faces with little to no visibility. In the manual response measure, we found remarkably high accuracy even when observers reported no conscious perception for both faces and houses (although it is possible that the presence of the face was used to guide responses in the house detection task). This suggests that fast feedforward representations may be sufficient to activate motor regions and form a decision variable in the frontal cortex before the onset of conscious perception (Freedman, Riesenhuber, Poggio, & Miller, 2003; Thorpe et al., 1996; VanRullen & Thorpe, 2001). Future work is needed to clarify the exact mechanism, but at least one study found that undetected (masked) stimuli elicited EEG responses associated with motor response preparation and influenced subsequent behavioral response (Dehaene et al., 1998). In primates, single-unit recording indicated that neurons in the frontal eye field (FEF) involved in transforming visual signals into motor commands were activated by both detected and undetected (masked) shape targets (Thompson & Schall, 1999). Although the FEF response was stronger for detected targets, the activation of the FEF in the absence of detection is consistent with our saccadic data for strongly masked face images.

There are two major caveats to the current findings. First, we used only houses as a distractor category for faces, so we cannot conclude that ultra-fast, feedforward saccades are specific to faces. In fact, like the tendency to saccade to a face when a vehicle is the target (Crouzet et al., 2010; Little et al., 2021), observers show the same tendency to saccade to an animal when a vehicle is a target (Crouzet, Joubert, Thorpe, & Fabre-Thorpe, 2012). Moreover, when contrasted with vehicles, the minimum SRT for animals was 120 ms, and this was faster than scene categorization (minimum SRT = 160 ms). This would predict that the saccadic bias for faces might be weaker if animals were the distractor. Feedforward representations might therefore enhance detection of categories that can be reliably detected based on LSF content (i.e., those that have the regularities in spatial structure needed to categorize them) and/or benefit from the quickest behavioral response. Moreover, face- and animal-selective cortices along the ventral visual pathway are found in more lateral aspects of the cortex, whereas scene- and place-selective areas are found along the medial aspect (Grill-Spector & Weiner, 2014). Additional work might examine whether this functional organization is relevant to the speed of categorization, automatic detection, and LSF sensitivity.

Second, it could be argued that the rapid stimulus offset and masking image onset disrupted the initiation of saccadic eye movements. To our knowledge, only one other study has incorporated backward masking into a saccadic detection task in which observers responded to simple "X" and "O" shapes with a 7-ms target-mask SOA, but saccadic responses were observed on over 90% of trials (Crouzet, Overgaard, & Busch, 2014). This indicates that it is theoretically possible to initiate a saccade despite rapid stimulus presentation, at least with very simple stimuli. However, an important difference in their study was that the target appeared at varying locations and among a number of distractors, and so the task involved a visual search component that may not be possible to perform without eye movements. By contrast, the target and distractor location is predictable in the standard 2AFC saccadic choice task used here, and maintaining a point of fixation may be advantageous when stimulus exposure is so limited. However, because manual response accuracy (especially for faces) indicates that the stimulus category was still being accurately encoded, we believe that the saccadic responses that were captured most likely reflect those underlying category representations.

To conclude, the current work replicates and extends the landmark finding of ultra-fast saccades to faces by testing the prediction that face-selective saccadic responses would be evoked even under strong backward masking. Whereas other studies have sought to characterize the image properties that contribute to fast face selection, such as phase information (Honey et al., 2008), spatial frequency (Guyader et al., 2017), amplitude spectrum (Crouzet & Thorpe, 2011), and orientation and contrast (Little et al., 2021), our study examined when that selectivity occurs with respect to feedforward and feedback processing and its relationship to conscious processing. Our findings not only support the claim that face-selective saccades are mediated by feedforward representations but also demonstrate a capacity for accurate response selection in the absence of perceptual awareness. By contrast, masking had a strong effect on both saccadic latency and accuracy for house detection. These divergent response profiles suggest that face detection is more sensitive to the coarse structure of the input carried by the feedforward signal and less dependent on recurrent processing than other object categories. Overall, these findings clarify the mechanism underlying fast saccades toward faces and reveal more about the role of unconscious and early visual processing in attention and eye movements for faces.

Keywords: saccadic choice, face detection, backward masking, rapid visual categorization, unconscious processing

Acknowledgments

Footnotes

[1]While recurrent processing may be necessary for conscious perception, recurrent processing has been observed in the absence of conscious perception, indicating that it is not sufficient (Fahrenfort et al., 2017).
[2]Although the saccadic response rate is typically much higher than reported in our 400-ms SOA conditions, the current design asks subjects to make both a saccadic response ("fixate your eyes on the target") and a manual response (using the keyboard), whereas subjects in previous saccadic choice tasks could respond by eye movement only. It is possible that having the additional manual response caused participants to be less compliant about making eye movements or to favor the manual response. However, the response profile of saccades observed in the 400-ms SOA condition replicates that observed in previous studies, suggesting that the data from the saccades that were initiated are generalizable.

References

Bacon-Macé, N., Macé, M. J. M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research, 45*(11), 1459–1469.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*(8), 617–629.

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences, 11*(7), 280–289.

Boehler, C. N., Schoenfeld, M. A., Heinze, H.-J., & Hopf, J.-M. (2008). Rapid recurrent processing gates awareness in primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 105*(25), 8742–8747.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & Psychophysics, 62*(8), 1572–1595.

Breitmeyer, B. G., & Ogmen, H. (2006). *Visual masking*. Oxford, UK: Oxford University Press, https://doi.org/10.1093/acprof:oso/9780198530671.001.0001.

Bridgeman, B. (1980). Temporal response characteristics of cells in monkey striate cortex measured with metacontrast masking and brightness discrimination. *Brain Research, 196*(2), 347–364.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews, 36*(2–3), 96–107.

Caharel, S., Leleu, A., Bernard, C., Viggiano, M.-P., Lalonde, R., & Rebaï, M. (2013). Early holistic face-like processing of Arcimboldo paintings in the right occipito-temporal cortex: Evidence from the N170 ERP component. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology, 90*(2), 157–164.

Campana, F., Martin, J. G., Bokeria, L., Thorpe, S., Jiang, X., & Riesenhuber, M. (2020). Evidence for face selectivity in early vision. bioRxiv, https://doi.org/10.1101/2020.03.14.987735.

Camprodon, J. A., Zohary, E., Brodbeck, V., & Pascual-Leone, A. (2010). Two phases of V1 activity for visual recognition of natural images. *Journal of Cognitive Neuroscience, 22*(6), 1262–1269.

Cauchoix, M., Crouzet, S. M., Fize, D., & Serre, T. (2016). Fast ventral stream neural activity enables rapid visual categorization. *NeuroImage, 125*, 280–290.

Crouzet, S. M., Joubert, O. R., Thorpe, S. J., & Fabre-Thorpe, M. (2012). Animal detection precedes access to scene category. *PLoS One, 7*(12), 1–9.

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, https://doi.org/10.1167/10.4.16.

Crouzet, S. M., Overgaard, M., & Busch, N. A. (2014). The fastest saccadic responses escape visual masking. *PLoS One, 9*(2), e87418.

Crouzet, S. M., & Thorpe, S. J. (2011). Low-level cues and ultra-fast face detection. *Frontiers in Psychology, 2*, 1–9, https://doi.org/10.3389/fpsyg.2011.00342.

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., … Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature, 395*(6702), 597–600.

Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biology, 5*(10), 2408–2423.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron, 73*(3), 415–434.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General, 129*(4), 481–507.

Di Oleggio Castello, M. V., & Gobbini, M. I. (2015). Familiar face detection in 180 ms. *PLoS One,*

10(8), e0136548, https://doi.org/10.1371/journal.pone.0136548.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2007). Masking disrupts reentrant processing in human visual cortex. *Journal of Cognitive Neuroscience, 19*(9), 1488–1497.

Fahrenfort, J. J., Van Leeuwen, J., Olivers, C. N. L., & Hogendoorn, H. (2017). Perceptual integration without conscious access. *Proceedings of the National Academy of Sciences of the United States of America, 114*(14), 3744–3749.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*(1), 1–47.

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception, 37*(4), 571–583.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 23*(12), 5235–5246.

Goffaux, V., Gauthier, I., & Rossion, B. (2003). Spatial scale contribution to early visual differences between face and object processing. *Brain Research: Cognitive Brain Research, 16*(3), 416–424.

Goffaux, V., Jemel, B., Jacques, C., Rossion, B., & Schyns, P. G. (2003). ERP evidence for task modulations on face perceptual processing at different spatial scales. *Cognitive Science, 27*(2), 313–325.

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2011). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral Cortex, 21*(2), 467–476.

Goffaux, V., & Rossion, B. (2006). Faces are "spatial"—Holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance, 32*(4), 1023–1039.

Goold, J. E., & Meng, M. (2016). Visual search of Mooney faces. *Frontiers in Psychology, 7*, 155.

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience, 15*(8), 536–548.

Guyader, N., Chauvin, A., Boucart, M., & Peyrin, C. (2017). Do low spatial frequencies explain the extremely fast saccades towards human faces? *Vision Research, 133*, 100–111.

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), 1008–1017, https://doi.org/10.1167/6.10.1.

Harris, J. J., Schwarzkopf, D. S., Song, C., Bahrami, B., & Rees, G. (2011). Contextual illusions reveal the limit of unconscious visual processing. *Psychological Science, 22*(3), 399–405.

Haynes, J.-D., Driver, J., & Rees, G. (2005). Visibility reflects dynamic changes of effective connectivity between V1 and fusiform cortex. *Neuron, 46*(5), 811–821.

Heeman, J., Van der Stigchel, S., & Theeuwes, J. (2017). The influence of distractors on express saccades. *Journal of Vision, 17*(1):35, 1–17, https://doi.org/10.1167/17.1.35.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*(5), 791–804.

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, https://doi.org/10.1167/8.12.9.

Hong, H., Yamins, D. L. K., Majaj, N. J., & DiCarlo, J. J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature Neuroscience, 19*(4), 613–622.

Jacques, C., Jonas, J., Maillard, L., Colnat-Coulbois, S., Koessler, L., & Rossion, B. (2019). The inferior occipital gyrus is a major cortical source of the face-evoked N170: Evidence from simultaneous scalp and intracerebral human recordings. *Human Brain Mapping, 40*(5), 1403–1418.

Keil, M. S. (2008). Does face image statistics predict a preferred spatial frequency for human face processing? *Proceedings. Biological Sciences/The Royal Society, 275*(1647), 2095–2100.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46*(11), 1762–1776.

Kleiner, M., Brainard, D., & Pelli, D. (2007). *What's new in Psychtoolbox-3?* https://pure.mpg.de/rest/items/item_1790332/component/file_3136265/content.

Koivisto, M., Railo, H., Revonsuo, A., Vanni, S., & Salminen-Vaparanta, N. (2011). Recurrent processing in V1/V2 contributes to categorization of natural scenes. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 31*(7), 2488–2492.

Koivisto, M., Salminen-Vaparanta, N., Grassini, S., & Revonsuo, A. (2016). Subjective visual awareness emerges prior to P3. *The European Journal of Neuroscience, 43*(12), 1601–1611.

Kovács, G., Vogels, R., & Orban, G. A. (1995). Cortical correlate of pattern backward masking. *Proceedings of the National Academy of Sciences of the United States of America, 92*(12), 5587–5591.

Kreiman, G., & Serre, T. (2020). Beyond the feedforward sweep: Feedback computations in the visual cortex. *Annals of the New York Academy of Sciences, 1464*(1), 222–241.

Lamme, V. A. F. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience, 1*(3), 204–220.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23*(11), 571–579.

Lamme, V. A. F., Supèr, H., Landman, R., Roelfsema, P. R., & Spekreijse, H. (2000). The role of primary visual cortex (V1) in visual awareness. *Vision Research, 40*(10–12), 1507–1521, https://doi.org/10.1016/s0042-6989(99)00243-6.

Lamme, V. A. F., Zipser, K., & Spekreijse, H. (2002). Masking interrupts figure-ground signals in V1. *Journal of Cognitive Neuroscience, 14*(7), 1044–1053, https://doi.org/10.1162/089892902320474490.

Little, Z., Jenkins, D., & Susilo, T. (2021). Fast saccades towards faces are robust to orientation inversion and contrast negation. *Vision Research, 185*, 9–16.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience, 5*(9), 910–916.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman.

Martin, J. G., Cox, P. H., Scholl, C. A., & Riesenhuber, M. (2019). A crash in visual processing: Interference between feedforward and feedback of successive targets limits detection and categorization. *Journal of Vision, 19*(12):6, 1–21, https://doi.org/10.1167/19.12.20.

Oppenheim, A. V., & Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE, 69*(5), 529–541.

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science, 292*(5516), 510–512.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Petras, K., Jacobs, C., ten Oever, S., & Goffaux, V. (2019). Coarse image information guides integration of fine details (if you let it). *Perception, 48*, 180–181.

Petras, K., Ten Oever, S., Jacobs, C., & Goffaux, V. (2019). Coarse-to-fine information integration in human vision. *NeuroImage, 186*, 103–112.

Quek, G. L., Liu-Shuang, J., Goffaux, V., & Rossion, B. (2018). Ultra-coarse, single-glance human face detection in a dynamic visual stream. *NeuroImage, 176*, 465–476.

Ramsøy, T. Z., & Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences, 3*(1), 1–23.

Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology, 12*(2), 162–168.

Ro, T., Breitmeyer, B., Burton, P., Singhal, N. S., & Lane, D. (2003). Feedback contributions to visual awareness in human occipital cortex. *Current Biology, 13*(12), 1038–1041.

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*(4), 1959–1979.

Rossion, B., Torfs, K., Jacques, C., & Liu-Shuang, J. (2015). Fast periodic presentation of natural images reveals a robust face-selective electrophysiological response in the human brain. *Journal of Vision, 15*(1):18, 1–18, https://doi.org/10/1167.15.1.18.

Sandberg, K., Timmermans, B., Overgaard, M., & Cleeremans, A. (2010). Measuring consciousness: Is one measure better than the other? *Consciousness and Cognition, 19*(4), 1069–1078.

Schiller, P. H., & Kendall, J. (2004). Temporal factors in target selection with saccadic eye movements. *Experimental Brain Research, 154*(2), 154–159.

Schuurmans, J. P., Bennett, M. A., Petras, K., & Goffaux, V. (2023). Backward masking reveals coarse-to-fine dynamics in human V1. *NeuroImage, 274*, 120139.

Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., & Poggio, T. (2007). A quantitative theory of immediate visual recognition. *Progress in Brain Research, 165*, 33–56.

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*(15), 6424–6429.

Sherman, S. M. (2016). Thalamus plays a central role in ongoing cortical functioning. *Nature Neuroscience, 19*(4), 533–541.

Thompson, K. G., & Schall, J. D. (1999). The detection of visual signals by macaque frontal eye field during

masking. *Nature Neuroscience, 2*(3), 283–288, https://doi.org/10.1038/6398.

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

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems, 14*(3), 391–412.

Ungerleider, L. G., & Haxby, J. V. (1994). "What" and "where" in the human brain. *Current Opinion in Neurobiology, 4*(2), 157–165.

VanRullen, R. (2006). On second glance: Still no high-level pop-out effect for faces [Review of *On second glance: Still no high-level pop-out effect for faces*]. *Vision Research, 46*(18), 3017–3027.

VanRullen, R. (2007). The power of the feed-forward sweep. *Advances in Cognitive Psychology, 3*(1–2), 167–176.

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience, 13*(4), 454–461.

VanRullen, R., & Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Research, 42*(23), 2593–2615.

Wichmann, F. A., Drewes, J., Rosas, P., & Gegenfurtner, K. R. (2010). Animal detection in natural scenes: Critical features revisited. *Journal of Vision, 10*(4):6, 1–27, https://doi.org/10.1167/10.4.6.