The Role of Emotional Content and Perceptual Saliency During the Programming of Saccades Toward Faces

Lea Entzmann, Louise Kauffmann, Juliette Lenouvel, Clémence Charles, Carole Peyrin, Roman Vuillaume, Martial Mermillod, Nathalie Guyader

The role of emotional content and perceptual saliency during the programming of saccades toward faces

Léa Entzmann[1, 2], Nathalie Guyader[2], Louise Kauffmann[1], Juliette Lenouvel[1], Clémence Charles[1], Carole Peyrin[1], Roman Vuillaume[3], & Martial Mermillod[1]

[1] Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000, Grenoble, France.

[2] Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France.

[3] Univ. Bourgogne Franche-Comté, ImViA, 21078 Dijon, France.

[×] Corresponding author
Léa Entzmann
LPNC
1251 Avenue Centrale,
Bâtiment Michel Dubois,
38400 Saint-Martin-d'Hères
lea.entzmann@univ.grenoble-alpes.fr

Abstract

Previous studies have shown that the human visual system can detect a face and elicit a saccadic eye movement toward it very efficiently compared to other categories of visual stimuli. In the first experiment, we tested the influence of facial expressions on fast face detection using a saccadic choice task. Face-vehicle pairs were simultaneously presented and participants were asked to saccade toward the target (the face or the vehicle). We observed that saccades toward faces were initiated faster, and more often in the correct direction, than saccades toward vehicles, regardless of the facial expressions (happy, fearful, or neutral). We also observed that saccade endpoints on face images were lower when the face was happy and higher when it was neutral. In the second experiment, we explicitly tested the detection of facial expressions. We used a saccadic choice task with emotional-neutral pairs of faces and participants were asked to saccade toward the emotional (happy or fearful) or the neutral face. Participants were faster when they were asked to saccade toward the emotional face. They also made fewer errors, especially when the emotional face was happy. Using computational modeling, we showed that this happy face advantage can, at least partly, be explained by perceptual factors. Also, saccade endpoints were lower when the target was happy than when it was fearful. Overall, we suggest that there is no automatic prioritization of emotional faces, at least for saccades with short latencies, but that salient local face features can automatically attract attention.

Keywords: Emotional facial expressions; Eye movements; Saccade programming; Neural computation; Time course.

1. Introduction

The human visual system is extremely efficient in the rapid, preferential detection of socially relevant stimuli such as faces. In particular, eye-tracking data have shown that when presented in visual scenes, faces immediately attract the gaze of observers, who then spend most of the exploration time looking at them (Cerf et al., 2009; Coutrot & Guyader, 2014; Foulsham et al., 2010; Marat et al., 2009). Furthermore, saccades toward individual faces can be made continuously, at rates up to 6 faces/second (Martin et al., 2018). Moreover, when presented along with a distractor image (e.g., a vehicle), face stimuli can be detected and elicit a saccade toward them very rapidly, while more time is needed for other objects (Crouzet et al., 2010; Guyader et al., 2017; Kauffmann et al., 2019).

This last result has been highlighted in saccadic choice tasks, where saccades are used as behavioral responses. In such tasks, two images from different categories (e.g., a face and a vehicle) are displayed on the screen and participants are asked to make a saccade as fast as possible toward the image that contains the target category (i.e., the face or the vehicle). Such tasks have revealed that saccades toward faces can be reliably elicited in only 100 ms (Crouzet et al., 2010) and, overall, suggest that faces can attract the gaze very rapidly. This bias for faces during the saccadic choice task has been replicated across many studies and is robust to stimulus manipulations such as grayscaling, thumbnail and phase scrambling or spatial frequency filtering (Boucart et al., 2016, Guyader et al., 2017, Honey et al., 2008, Crouzet et al., 2011, Kauffmann et al., 2021). Furthermore, it persists even if faces are opposed to distractors sharing a similar shape, degree of animacy or structural homogeneity (Boucart et al., 2016, Kauffmann et al., 2021). In everyday life, humans are well accustomed to transmitting and decoding emotional information from faces by means of facial expressions used for social communication. The brain has consequently developed a number of specific and complex mechanisms, which are as yet not fully understood, to process emotionally relevant information from faces (for a review, see Adolphs, 2003). While face stimuli can guide attention very efficiently, it seems appropriate to ask how this can be modulated through facial expressions.

In fact, facial expressions are characterized by both a specific physical facial configuration and an emotion they are assumed to convey (Calvo & Nummenmaa, 2015). Generally, emotional stimuli have been found to be processed more efficiently than neutral ones. For example, an emotional object is more likely to be fixated first than a neutral object (Humphrey et al., 2012; Niu et al., 2012). Also, many studies suggest that humans have evolved

to preferentially orient their attention toward threatening stimuli. At the behavioral level, angry faces seem to be particularly well detected when presented among matrices of other expressions in visual search tasks (Fox & Damjanovic, 2006; Öhman et al., 2001; Schubö et al., 2006; Tipples et al., 2002; for a review, see Frischen et al., 2008). Also, neurophysiological data suggest that this attentional modulation is associated with activations in the limbic system, including the amygdala. In fact, classical models of emotional processes in the brain suppose that threat-related stimuli, in this case fearful faces in particular, can be detected very rapidly through a subcortical pathway involving the superior colliculus, the pulvinar, and the amygdala. This pathway is thought to be magnocellular and to transmit coarse information in parallel to the finer and slower cortical processing (LeDoux, 2000; Morris, 1998; Öhman, 2005; Tamietto & de Gelder, 2010). This idea is supported by intracranial EEG recording in the amygdala showing activations as early as 74ms for coarse fearful faces (Méndez-Bértolo et al., 2016).

However, the existence of such a pathway as well as its involvement in the detection of facial expressions remain a matter of debate, and it is possible that the subcortical route operates for faces irrespective of emotions (Fitzgerald et al., 2006; Garvert et al., 2014; Johnson, 2005; McFadyen et al., 2017). Also, other behavioral studies, especially those using non-schematic faces, have found no prioritization of threatening faces (Becker et al., 2011; Calvo & Marrero, 2009; Calvo & Nummenmaa, 2011; Hunt et al., 2007; Lipp et al., 2009; Juth et al., 2005). For example, Calvo and Nummenmaa (2009) used a saccadic choice task with one emotional and one neutral face presented simultaneously and found that happy, rather than angry or fearful faces, were detected better. These observations are consistent with results from categorization task studies which used manual responses (Calvo & Lundqvist, 2008; Tottenham et al., 2009). In 2011, the same authors assessed the role of perceptual (e.g., luminance, local saliency) and semantic (e.g., affective valence) factors in the discrimination advantage of happy faces, and found only a contribution of mouth saliency. They suggested that this "happy face advantage" relies more on the saliency of perceptual features (like the open mouth) rather than the interpretation of the emotional content (Calvo & Nummenmaa, 2011). Indeed, the smiling mouth is more salient than any other region of happy and non-happy faces (based on a combination of physical image properties such as luminance, contrast and spatial orientation; Itti & Koch, 2000; Calvo & Nummenmaa, 2008), suggesting that it can be used as a shortcut for the quick recognition of happy faces (Calvo & Nummenmaa, 2015). Thus, Horstmann et al., (2012) found that visible teeth in angry or happy faces make them easier to detect in a crowd of neutral faces.

The first goal of this study is to determine whether the presence of an emotional expression can facilitate very fast face detection. Indeed, although many studies have focused on what drives the very fast saccades toward faces, these studies have mostly used neutral faces and it is still unclear whether emotional facial expression influences face detection. In the first experiment, we reproduced a saccadic choice task with face-vehicle pairs, using faces which portrayed either a happy, fearful or neutral expression. We chose fearful rather than angry faces in order to establish a connection with neurophysiological data showing that such faces are processed very efficiently in the brain (LeDoux, 2000; Méndez-Bértolo et al., 2016). Contrary to previous studies, in which images were presented in natural contexts (Crouzet et al., 2010; Kauffmann et al., 2019; Kirchner & Thorpe, 2006), we used more prototypical stimuli. Indeed, classical databases of emotional faces are very prototypical, with all the faces being centered at the same position in the image. Hence, we also used very prototypical vehicle stimuli taken from the stimuli created by Kloth and used by Kloth et al., (2013) in a study in which the authors tested neurophysiological responses to non-face objects that are structurally similar to faces. In line with previous findings, we expected better performances (higher accuracy, shorter latency) for saccades toward face than saccades toward vehicle targets. Regardless of which specific expression can attract the attention the most, we generally expected emotional faces to attract the attention more than neutral faces because of their emotional relevance. Concerning the differences between happy and fearful faces, we formulated two alternative hypotheses. First, assuming that humans evolved to preferentially orient their attention toward threat, we should observe better performances toward fearful than happy face targets. Conversely, assuming that happy faces are easier to detect because of perceptual factors, we should find better performances toward happy than fearful face targets.

We also examined saccade endpoints in order to assess the distribution of attention within the face during saccade programming as a function of the expression. Indeed, saccade programming is thought to occur in a priority map in which both bottom-up (e.g., local saliencies) and top-down (e.g., emotional relevance) information is integrated. This priority map is assumed to follow a retinotopic organization associated with a winner-takes-all mechanism that guides the allocation of attention through orientation of the gaze (Belopolsky, 2015; Bisley & Mirpour, 2019; Fecteau & Munoz, 2006; Klink et al., 2014; Theeuwes, 2019). Saccade endpoints are thus the result of a competition between multiple locations. It is well known that emotional facial expressions have different diagnostic features, mostly in the form of the eyes and the mouth (Eisenbarth & Alpers, 2011; Smith et al., 2005; Wegrzyn et al., 2017). For example, Smith et al. (2005) found that the recognition of happy and surprised faces is

based more on the mouth, whereas the recognition of fearful and angry faces is based more on the eyes. While preparing an eye movement toward a face, we suppose that attention is directed toward it. However, we expected that the diagnostic or salient features of each expression would be able to modulate the allocation of attention within the face by attracting it. For example, if a happy face is presented, attention may be directed more specifically toward the mouth, which is more salient and diagnostic for this expression and, if a fearful face is presented, more attention may be directed toward the eyes. Thus, considering that saccade endpoints can reflect the distribution of attention within the face, we expected that they would be closer to the eyes for fearful than for happy faces.

In the first experiment, the influence of facial expressions was assessed at a very early stage of visual processing (with saccades that are often elicited in less than 200 ms) using a task that does not explicitly require the processing of facial expressions. Some studies suggest that expression decoding only occurs at a later stage of visual processing (after 180 ms; Schyns et al., 2009; Kulke, 2019). It is therefore unclear whether expressions could have been decoded before saccade onset in the first experiment. In the second experiment, we explicitly assessed the detection of facial expressions in order to test the preferential processing of emotional faces displaying task-relevant expressions. In a way similar to Bannermann et al. (2009) and Calvo et al. (2009), neutral and emotional faces (happy and fearful) were presented simultaneously in a saccadic choice task, and participants were asked to saccade toward the emotional or the neutral face. We still expected to find better performances when the target was an emotional than when it was a neutral face. Furthermore, if attention can be preferentially attracted by threat, we should observe better performances for fearful faces. Alternatively, given that happy faces are easier to recognize, better performances might be expected for happy faces. However, even if this were the case, we still expected saccade endpoints to be closer to the eyes for fearful than for happy faces.

Finally, we present a computational model that simulates results from a task similar to the one presented in the second experiment. The goal of this simulation was to show that the happy face advantage that could be expected in the second experiment can be explained through perceptual factors, as suggested by Calvo et al. (2009). Computational models are a useful tool for testing the role of the physical properties of inputs because they are not sensitive to high-level influences, such as the interpretation of the emotional content (Mermillod et al., 2009). If there is a happy face advantage and if it is mainly perceptual in nature, then the neural network should perform better with happy than with fearful faces.

2. Experiment 1: Face vs. Vehicle

2.1. Materials and method

2.1.1. Participants

Sixty-seven participants were recruited at the local university to perform a saccadic choice task. Considering the simple nature of the task, six of them were removed from the statistical analysis due to their low proportion of correct responses (below 0.75 in both sessions), leaving a group of sixty-one participants for inclusion in the statistical analysis (29 females; $M \pm SD$: 21.86 ± 0.47 years; age range: 18-36 years). All of them had normal or corrected-to-normal visual acuity. Undergraduate psychology students received course credits for their participation in the experiment. All participants gave their informed written consent before the experiment, which was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

2.1.2. Stimuli

Stimuli consisted of 160 grayscale photographs, with 120 images containing a face (40 different faces with 3 emotions; fearful, happy, and neutral) and 40 containing a vehicle (Fig. 1). During the experiment, images were systematically presented in face-vehicle pairs. *Face stimuli* were chosen from the Karolinska Directed Emotional Faces database (KDEF; Lundqvist et al., 1998), which is widely used in the field of emotion processing. *Vehicle stimuli* took the form of 40 pictures of cars taken from the stimuli used by Kloth et al. (2013). Each vehicle stimulus was duplicated three times to counterbalance the number of face and vehicle images (just as each of the 40 individual faces was seen three times for the three emotion conditions). The KDEF database is constructed in such a way that the vertical and horizontal positions of the eyes and mouth of each picture are set to the same positions on a digital grid. The original images have a size of 562 pixels x 762 pixels. However, for our experiment, they were cropped by 100 pixels at the top and bottom and were resized to 300 pixels x 300 pixels, corresponding to a coverage of 11 × 11 degrees of visual angle at a 57-cm viewing distance. This preprocessing allowed us to create a database of images with a size matching that used for our previous papers (Kauffmann et al., 2019; Guyader et al., 2017), and to have images with close-up faces. Moreover, the position of the eyes corresponded to the middle of the image as shown in Fig. 3. After resizing, the images were equalized in terms of mean luminance and root mean squared

contrast (mean luminance value of 127 and a mean contrast of 47, for pixel intensities between [0,255], based on mean luminance and contrast values of all the stimuli). Following this equalization step, all the images (faces and vehicles) globally had the same mean luminance and mean RMS contrast. A gamma correction was applied to each image to adapt the stimuli to the screen luminance. Vehicles were set against a gray background and were manually resized so that faces and vehicles had the same average spatial position and size. A training session consisting of six practice trials with 12 additional images not used in the experimental sessions (6 vehicles and 6 faces) was performed at the beginning of the experiment to allow the participants to familiarize themselves with the task.

Figure 1: Example of face and vehicle stimuli used in Experiment 1 (left), and their mean amplitude spectrum (right).

2.1.3. Procedure

Stimuli were displayed on a 24-inch screen with a spatial resolution of 1360×768 pixels and a refresh rate of 60 Hz. This resolution was used with a desktop computer. A keyboard was placed in front of the participants, allowing them to end the breaks by pressing the spacebar. The screen was raised a little at a distance of about 20 cm away from the keyboard and the back of the experimentation room was empty. Eye movements were recorded with an Eyelink 1000 (SR Research) eye-tracker with a 1000-Hz sampling frequency. Viewing was binocular, but only the position of the dominant eye was recorded. Saccades were automatically detected using the Eyelink software. Saccades were detected if they had a minimum velocity of 30 degrees/s, a minimum acceleration of 8000 degrees/s^2 and a minimum motion of 0.15 degrees. Blinks were detected when the pupil was partially or totally occluded, and fixations were detected when there was no blink and no saccade in progress. The experiment was divided into two sessions, with the order being counterbalanced between participants. In each session, each of

the 120 different faces was displayed, once on the left and once on the right side of the screen, leading to 240 trials per session and, therefore, a total of 480 trials at the end of the two sessions. Each face was opposed to a random vehicle. In one session, the target stimulus was the face (vehicle distractor), while in the other session, the target stimulus was the vehicle (face distractor). A calibration phase was performed at the beginning and middle of each session and a drift correction was applied every ten trials (if the drift was larger than 1° then recalibration was performed). During the calibration phase, participants were asked to gaze at 9 white dots appearing sequentially in a 3 × 3 grid covering the entire screen. Matlab (MathWorks, Natick, MA) and the Psychophysics Toolbox (Brainard, 1997) were used to control timing and stimulus display as well as communication with the eye-tracker.

During the experiment, participants were seated on an adjustable chair in a semi-lighted room. The head was stabilized by means of a forehead and a chin-rest at a fixed distance of 57 cm from the screen. A session lasted approximately 20 minutes and the whole experimental procedure took approximately 50 minutes. The target category (face or vehicle) was defined before each session. At the beginning of each trial, participants were asked to fixate a white cross during a pseudo-random time interval ranging between 800 and 1600 ms. After a 200-ms gap, two images (a face and a vehicle) were simultaneously displayed on each side of the screen for 400 ms (Fig. 2), and participants were asked to make a saccade as fast as possible toward the target image. The center of each image was located at a fixed distance of 8° of eccentricity from the center of the screen. Each trial ended with the presentation of a gray background for 1000 ms.

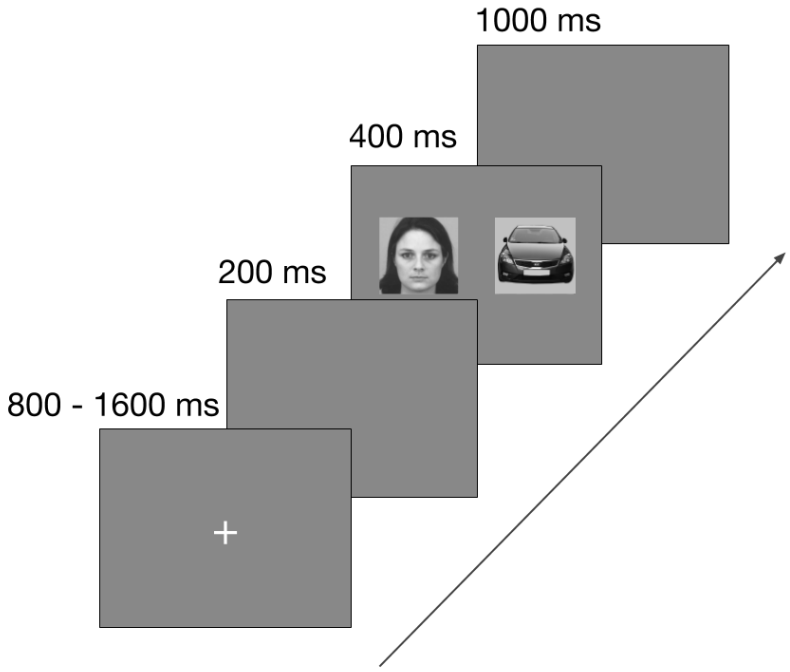

Figure 2: Time course of a trial for Experiment 1.

2.1.4. Data analysis

Before any further analysis, the eye movement data was preprocessed in order to eliminate trials that we considered to be invalid. Valid trials were selected according to the following validity criteria. First, a saccade had to be the first event after stimulus onset, with no blink occurring during its execution. Second, this first saccade had to have a latency greater than 50 ms (to avoid anticipatory saccades), a starting point within a radius of 2° around the center of the screen, and a duration smaller than 100 ms. Moreover, the saccade amplitude had to be greater than 1° and should not go beyond the screen. This preprocessing led to 10.4% of the initial number of trials being rejected. For all valid trials, only the first saccade and fixation were analyzed.

Statistical analyzes were carried out using the open-source software R (R Core Team, 2016) with R Studio 1.1.456 (Racine, 2012). A saccade was considered as "correct" if it was directed toward the side of the display containing the target and as an "error" if it was in the opposite direction (i.e., directed toward the distractor).

In order to quantify the orienting of attention toward local face features such as the mouth or the eyes, we computed saccade endpoints. To do this, we extracted and visualized the coordinates X_e and Y_e of the first saccade endpoint in the image space (i.e., within a square of 11 x 11 degrees, the coordinates X_0 and Y_0 being at the top-left corner). Overall, saccades

tended to land around the eyes, which corresponded to the center of the image. As a measure to compare endpoint positions between conditions, we analyzed the vertical distance between the endpoint of the first correct saccade and the center of the image in each condition. This distance corresponds to the distance between the Y-coordinate of the endpoint Y_e of the saccade and the Y-coordinate of the center of the image Y_c. A visual representation of the distance to the center is presented in Fig. 3, and a negative value corresponds to a saccade landing below the image center. This measure was also computed for the vehicles.

Mean accuracy (in % of correct responses), mean first saccade latency (in ms), and mean vertical distance to the image center (in degrees of visual angle) were computed for each participant in each experimental condition and analyzed as dependent variables. First, a paired samples t-test with the Target (Face, Vehicle) as a within-subject factor was used to assess the main effect of the target. Next, a repeated measures ANOVA with the Emotional Facial target (EFE; Happy, Neutral or Fearful) as a within-subject factor was conducted for saccades when the target was a face. Similarly, a repeated measures ANOVA with the Emotional Facial distractor (EFE; Happy, Neutral or Fearful) as a within-subject factor was also conducted for saccades when the target was a vehicle. If needed (i.e., if a significant effect of the EFE target or EFE distractor was observed), paired samples t-tests were used for pairwise comparisons between Emotional Facial Expressions (EFE; Happy, Neutral or Fearful). Effect sizes were estimated by calculating partial eta-squared (η_p^2) for ANOVAs and Cohen's d for t-tests. An effect was considered significant if its p value was below the threshold $\alpha = .05$.

Before performing the parametric tests, statistical assumptions were tested using a K-S corrected Lilliefors test (Lilliefors, 1967) for normality of distributions (of within-pair differences for t-tests and of the variable for repeated measures ANOVAs; McCrum-Gardner, 2008), and Mauchly's sphericity test was used to test for equality of variances (of the differences between all possible pairs; for repeated measures ANOVAs). When distributions deviated significantly from the normal distribution ($p < .05$), non-parametric tests were used. More precisely, Wilcoxon Signed-ranks tests and Friedman tests were performed on the dependent variables instead of the t-tests and ANOVAs (respectively). Therefore, when the conditions permitted the use of parametric tests, we favored the use of such tests, as they are known to be more powerful than non-parametric tests (Hoskin, 2012).

Finally, we computed the minimum latency (also referred to as the minimum saccadic reaction times in previous studies – Crouzet et al., 2010; Guyader et al., 2017; Kauffmann et al., 2021 – for each Target condition, and for each facial target EFE. To compute the minimum latency, we recorded the latencies of all the first saccades for all participants. We computed

their distribution while taking account of the saccade accuracy (Correct, Error), the type of target (Face, Vehicle) and, in the case of the face targets, the type of EFE (Neutral, Happy, or Fearful). For face and vehicle targets, 13194 (1863 errors) and 13037 (2125 errors) saccades, respectively, were used to compute the distributions. For neutral, happy and fearful faces, the distributions contained 4394 (640 errors), 4416 (601 errors), and 4384 (622 errors) saccades, respectively. The minimum latency corresponds to the time as of which there were significantly more correct than error saccades. More precisely, distributions were divided into 10-ms time bins (e.g., the 170-ms bin contained latencies from 165 to 174 ms), and for each bin we used a $\chi 2$ test (with a criterion of $p < .05$) to test if there were significantly more correct than error saccades. If there were significantly more correct than error saccades in five consecutive bins, the first of these bins was defined as the minimum latency. Note that this procedure was the same as in previous papers using saccadic choice tasks (e.g., Crouzet et al., 2010; Guyader et al., 2017; Kauffmann et al., 2021). Experimental data, analysis code, and simulation code are available at https://osf.io/bjmcy/.

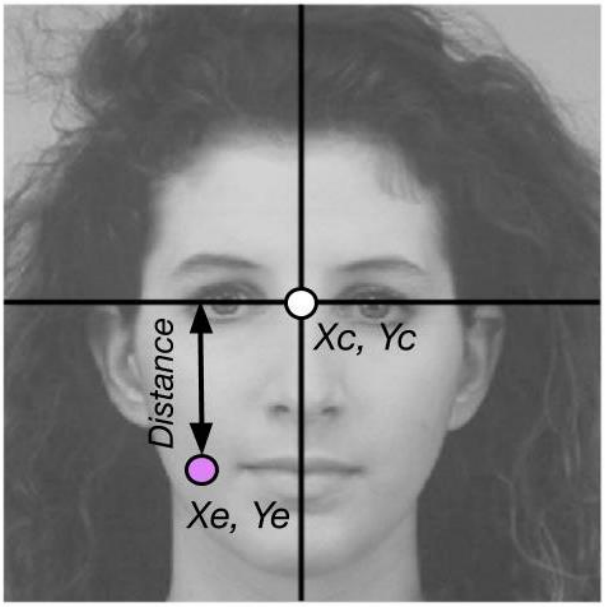

Figure 3: Visual representation of the vertical distance to the center. Xc,Yc denote the central point of the image, and Xe,Ye denote the endpoint of the saccade. The vertical distance to the center corresponds to the difference between Yc and Ye.

2.2. Results

2.2.1. Accuracy

A paired samples t-test performed on mean accuracy (Fig. 4a) indicated a significant effect of the Target ($t(60) = 2.52$, $p = .014$, $d = 0.32$). Participants were more accurate when the target was a face (*M* ± *SD*: .88 ± .076) than when it was a vehicle (*M* ± *SD*: .86 ± .082). Neither an effect of the EFE for facial targets nor an effect of the EFE distractor for vehicles targets were found.

2.2.2. Latency

A paired samples t-test performed on mean saccade latency (Fig. 4b) indicated a significant effect of the Target ($t(60) = -5.43$, $p < .001$, $d = 0.69$). Saccades were elicited faster when the target was a face (*M* ± *SD*: 176 ± 21.5 ms) than when it was a vehicle (*M* ± *SD*: 191 ± 27.2 ms). Neither an effect of the EFE for facial targets nor an effect of the EFE distractor for vehicles targets were found.

2.2.3. Minimum latency

The Minimum latency (Fig. 4d) was found in the 110-ms bin for faces (overall, and also for neutral, happy and fearful faces independently), and in the 130-ms bin for vehicles.

2.2.4. Endpoints

A Wilcoxon Signed-ranks test performed on mean vertical distance to the image center (Fig. 4c) revealed a main effect of the Target ($Z = 5.6$, $p < .001$). Saccades landed higher when the Target was a vehicle (*Mdn* = -0.016°*)* than a face (*Mdn = -0.32* °). A non-parametric Friedman test revealed a main effect of the EFE ($F(2) = 50.6$, $p < .001$) when the target was a face. In this condition, saccades landed higher for neutral faces (*M* ± *SD*: -0.26 ± 0.61°) than for fearful (*M* ± *SD*: -0.29 ± 0.61°; $t(60) = 3.78$, $p < .001$, $d = 0.92$) or happy (*M* ± *SD*: -0.38 ± 0.61 °; $t(60) = -9.04$, $p < .001$, $d = 1.15$) faces. Moreover, saccades toward fearful faces also landed higher than those toward happy faces ($t(60) = -7.21$, $p < .001$, $d = 0.48$). Finally, we did not find any effect of the EFE distractor when the target was a vehicle. Figure 5 presents examples of heat maps computed based on the saccade endpoints of all subjects. For this representation, we chose to display the heat maps on the top of three randomly chosen faces: a neutral, a happy, and a fearful face. Heat maps were obtained by 1) adding all the first correct saccades and 2) convolving a small 2D Gaussian on each endpoint.

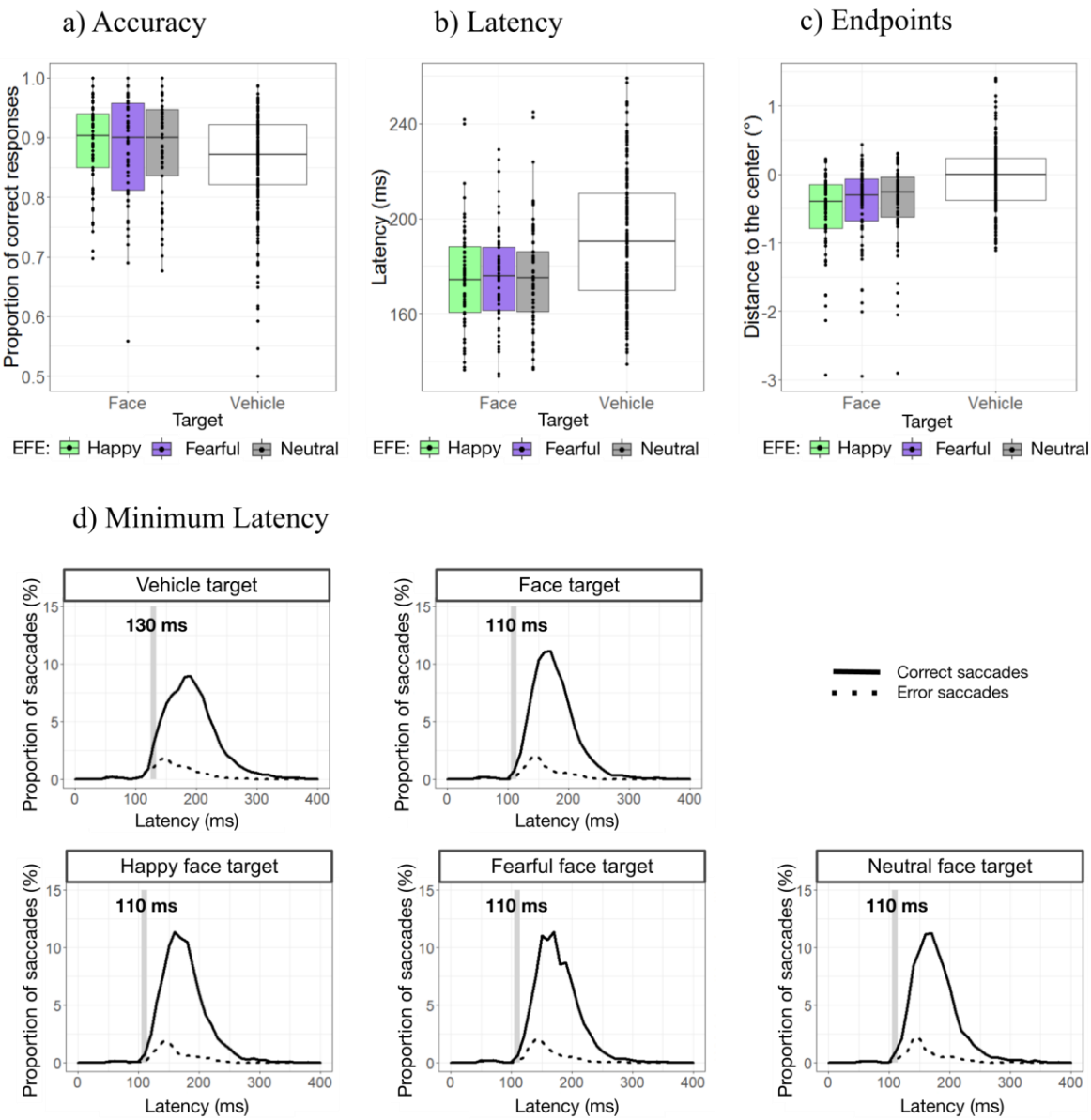

Figure 4: Boxplots for (a) mean proportion of correct responses, (b) mean latency (in ms), and (c) mean distance to the image center (in degrees of visual angle), according to the Target (Face and Vehicle) and the Emotional Facial Expression of face targets (EFE; Happy, Fearful or Neutral). (d) Distribution of saccade latencies for each Target and for each Emotional Facial Expression of face targets. Unbroken lines correspond to correct saccades and dotted lines to error saccades. The gray bar corresponds to the 10-ms bin containing the minimum latency. It should be noted that for the purposes of illustration, and because no significant effect of the EFE distractor was found, all types of face distractors (happy, fearful, and neutral) were recorded for vehicle targets.

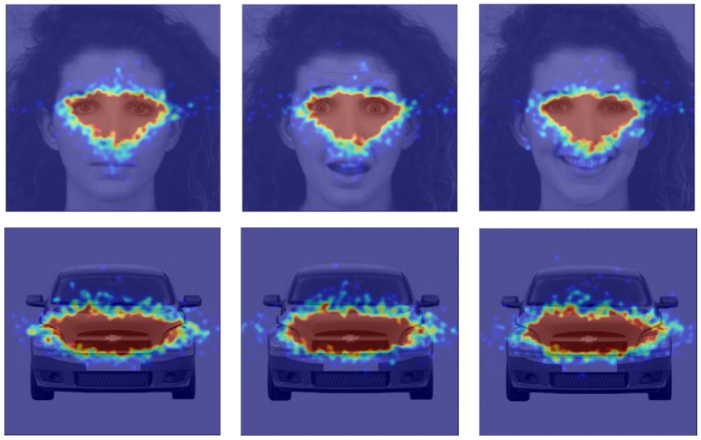

Figure 5: Heat maps computed from the endpoints of all the first correct saccades toward neutral (top left), fearful (top center), or happy (top right) faces, and toward vehicles when the distractor was a neutral (bottom left), fearful (bottom center) or happy (bottom right) face.

2.2.5. Bayes Factor analysis to test the lack of emotional modulation on very fast saccades

Results of Experiment 1 did not show any significant effect of the EFE (Happy, Fearful, and Neutral) on saccade accuracy and saccade latency when the target was a face. Therefore, the null hypothesis (H_0: no effect of EFE on mean saccade accuracy or mean saccade latency) cannot be rejected and no conclusion can be drawn (Hoijtink et al., 2019; Wagenmakers, 2007). To evaluate the probability of the presence or absence of an effect of EFE, we used a method based on Bayesian statistics (Bayes factors; Kass & Raftery, 1995). This method was added to the previous analyses to evaluate the probability of H_0 more precisely compared to the alternative hypothesis (H_a: not H_0). This was done for saccade accuracy and latency independently. We used the bain (Bayesian informative hypotheses evaluation; Gu et al., 2019; Hoijtink et al., 2019) R package. With this package, the variance of the prior distribution for each of the means is computed using a fraction of the information in the data for each group mean (which here renders a prior variance of 0.01 and 706 for the accuracy and latency, respectively).

An ANOVA was computed to estimate the mean accuracy and mean latency when the target was a face in each of the three emotion conditions, happy (*accuracy*, *M* ± *SD*: 0.89 ± 0.072; *latency*, *M* ± *SD*: 175 ± 21.6 ms), fearful (*accuracy*, *M* ± *SD*: 0.88 ± 0.091, *latency*, *M* ± *SD*: 175 ± 20.6 ms), and neutral (*accuracy*, *M* ± *SD*: 0.88 ± 0.084, *latency*, *M* ± *SD*: 176 ± 22.7 ms). Two hypotheses were evaluated,

H_0: $M_{Happy} = M_{Fearful} = M_{Neutral}$

H_a: not H_0

where M_{Happy}, $M_{Fearful}$ and $M_{Neutral}$ denote the mean accuracy or mean latency for happy, fearful or neutral face targets.

For accuracy, the Bayes factor versus H_a was 74, and the posterior probabilities (computed assuming equal prior probabilities) were 0.99 for H_0 and 0.01 for H_a. This Bayes factor suggests that the data are 74 times more likely to occur under H_0 than under H_a, and can be interpreted as providing very strong support for H_0 (for a scale for interpretation of the Bayes factor see Jeffreys, 1998).

For latency, the Bayes factor versus H_a was 88.3, and the posterior probabilities (computed assuming equal prior probabilities) were 0.99 for H_0, and 0.01 for the H_a.. This Bayes factor suggests that the data are 88 times more likely to occur under H_0 than under H_a, and can also be interpreted as providing very strong support for H_0 (Jeffreys, 1998).

2.3. Discussion

Results of Experiment 1 replicate previous findings showing that participants made fewer errors and initiated saccades faster when the target was a face than when it was a vehicle. Accuracy for face targets in our experiment was similar to that reported in previous saccadic choice tasks with face and vehicle targets (88% correct responses on average in this study, compared to 89.6%, 86%, and 87.5 % for previous studies using a saccadic choice task with face-vehicle pairs: respectively, Crouzet et al., 2010; Guyader et al., 2017; Kauffmann et al., 2019). However, accuracy for vehicle targets in this study was higher (86% correct responses on average in this study, compared to 71%, 71%, and 76.6% for previous studies using a saccadic choice task with face-vehicle pairs: respectively, Crouzet et al., 2010; Guyader et al., 2017; Kauffmann et al., 2019). The same pattern (similar results for face detection and better detection of vehicle targets in this experiment compared to previous ones) was found for the mean and minimum latencies. Furthermore, a previous study (Crouzet et al., 2010) had observed a tendency of early saccades (100-140 ms) to go toward the side with the faces, even if the task required a saccade toward the vehicles. However, no such effect was found here.

The better detection of vehicle stimuli in our experiment compared to previous ones may be due to the fact that we used more prototypical images with no background, thus reducing the variability between stimuli. Moreover, vehicle stimuli were duplicated to correspond to the three emotions of the same face (the same vehicle was presented three times), thus contributing to their low variability. It can also be noted that the high degree of within-category homogeneity of faces and cars resulted in both categories having distinct amplitude spectrum properties.

Previous studies have shown that such information is used during the saccadic choice task and could partly explain the bias for faces (Crouzet & Thorpe, 2011; Honey et al, 2008). For example, Honey et al. showed that images of faces for which the phase of the Fourier component (i.e., spatial relations within the image) was disrupted while the amplitude spectrum (AS) was preserved still elicited faster saccades than images of vehicles with similar alterations. A recent study (Kauffmann et al., 2021), however, found that faster saccades toward faces than vehicles could be observed even when the AS of the stimuli was made more similar for faces and cars, suggesting that AS differences between faces and cars cannot entirely explain the bias in favor of faces.

Concerning facial expressions, no effect was observed either on mean latency or on accuracy. The Bayes Factor analysis provides an interesting statistical tool allowing us to draw inferences about the likelihood of a null effect of emotions on saccades toward faces. The conclusion of Experiment 1 is that there is a very strong evidence that emotions do not influence fast saccades toward faces in a saccadic choice task in which participants have to saccade toward a face when a face and a vehicle are simultaneously displayed. Nevertheless, an effect of facial expressions was observed on saccade endpoints. Indeed, overall, saccades tended to land around the eyes, but landed higher when the face was neutral than when it was fearful or happy and also when the face was fearful rather than happy. The fact we found no prioritization of emotional faces contrasts with previous studies suggesting that emotional, and especially threatening, events are automatically (i.e., rapidly and non-intentionally) prioritized (Öhman, 2005). Fast face detection could be the result of "quick and dirty processing" that may be insufficient to decode expressions (Crouzet & Thorpe, 2011). Therefore, one explanation may be that expressions are not yet decoded at such small latencies and before a face is detected (Schyns et al., 2009; Mulckhuyse, 2018; Kulke, 2019).

In the second experiment, we used the same experimental design but with pairs of faces, one emotional (happy or fearful) and the other neutral, in order to directly test the detection of facial expressions. In one session, participants had to saccade toward emotional faces, and in the other, they had to saccade toward neutral faces. We expected to find better performances when the target was an emotional than when it was a neutral face. Furthermore, assuming that humans evolved to preferentially orient their attention toward threat, we should observe better performances toward fearful than happy face targets. However, behavioral studies showing an advantage for threatening faces have used manual responses and may reflect processes occurring at a later stage of visual processing. Thus, even though fearful faces were not prioritized in Experiment 1, we can still suggest that attention may be preferentially attracted

by threat if we consider that this prioritization occurs at a later stage. Alternatively, assuming that happy faces are easier to detect, we should find better performances for happy faces. We still expected saccade endpoints to be higher for fearful than for happy faces.

3. Experiment 2: Emotional face vs. Neutral face

3.1. Materials and methods

3.1.1. Participants

Twenty participants (nine females; mean age ± SD: 23.95 ± 5.26 years; age range: 19-41 years) were recruited from the local university to perform a saccadic choice task. All participants had normal or corrected-to-normal visual acuity. Undergraduate psychology students received course credits for their participation in the experiment. All participants gave their informed written consent before the experiment, which was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

3.1.2. Stimuli

Stimuli were 180 grayscale photographs of emotional and neutral faces chosen from the KDEF database (Fig. 6). More precisely, photographs of sixty different individuals (30 females) displaying three different facial expressions (neutral, happy, and fearful expressions) were included. All individuals presented in Experiment 1 were included in Experiment 2. Also, neutral stimuli were duplicated twice in this experiment to counterbalance the presentation of neutral and emotional stimuli. This led to a total of 240 images. The images again sized 300 × 300 pixels, corresponding to 11 × 11 degrees of visual angle, and were equalized in terms of mean luminance and root mean squared contrast (mean values of 126 and 66, respectively, for pixel intensity values between [0, 255], based on mean luminance and contrast values of all the stimuli) before a gamma correction was applied. A training session with eight practice trials involving 4 additional individuals was performed at the beginning of the experiment to allow participants to familiarize themselves with the task.

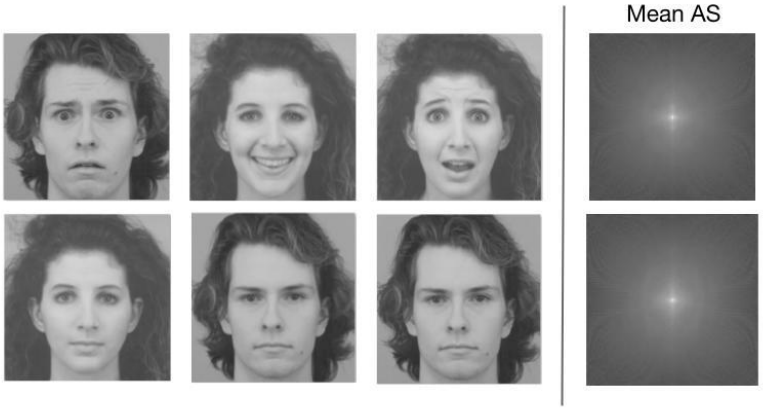

Figure 6: Example of emotional and neutral face stimuli used in Experiment 2 (left) and their mean amplitude spectrum (right). Each emotional face (top) is associated with a neutral face (bottom).

3.1.3. Procedure

The procedure was the same as in the first experiment. In each session, a neutral face was always displayed together with an emotional (happy or fearful) face, and participants were asked to make a saccade as fast as possible toward the target (the emotional or the neutral face), which was defined at the beginning of each session. A trial began with the presentation of a central fixation cross for a pseudo-random time period ranging between 800 and 1600 ms. After a 200 ms gap, an emotional and a neutral face were randomly and simultaneously displayed on either side of the screen for 800 ms (Fig. 7). In each session, each of the 240 different images was displayed, once on the left and once on the right side of the screen, leading to a total of 240 trials per session. The trial ended with the presentation of a gray screen for 1000 ms. One male face and one female face were displayed for each trial.

3.1.4. Data analysis

Preprocessing and data analysis methods were the same as in Experiment 1. The preprocessing procedure led to 7.8% of the initial number of trials being rejected. Mean accuracy, mean first saccade latency, as well as mean distance to the center of the image (which again corresponded to the eyes) were computed for each participant in each experimental condition and analyzed as dependent variables. First, a paired samples t-test with the Target (Emotional or Neutral) as a within-subject factor was used to assess the main effect of the Target. Then, a paired samples t-test with the Emotional Facial target (EFE; Happy or Fearful)

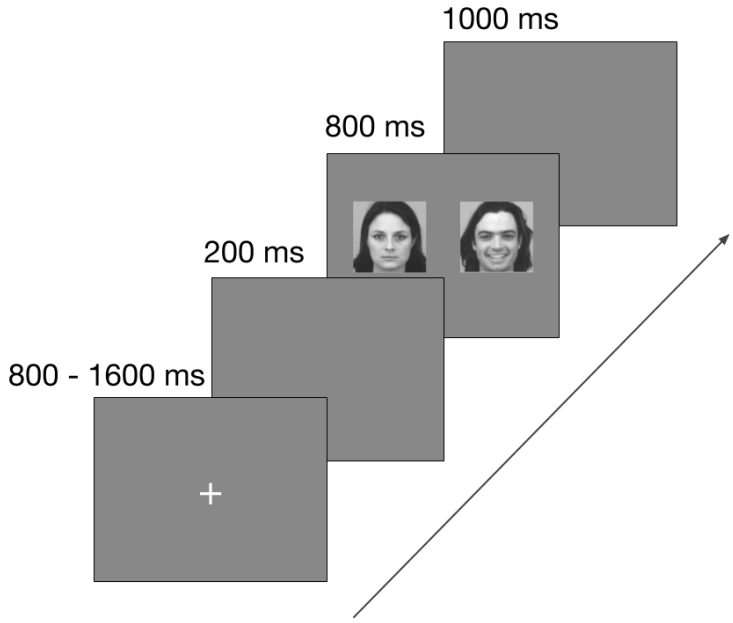

Figure 7: Time course of a trial for Experiment 2.

as a within-subject factor was applied for saccades in cases when the target was an emotional face. A paired samples t-test with the Emotional Facial distractor (EFE; Happy or Fearful) as a within-subject factor was applied for saccades in cases when the target was a neutral face.

Again, statistical assumptions were tested using a K-S corrected Lilliefors test for normal distribution of within-pair differences. When distributions deviated significantly from the normal distribution ($p < .05$), a Wilcoxon Signed-ranks test was applied. Minimum latency was computed in the same way as in the first experiment. For emotional and neutral targets, 4387 (1390 errors) and 4450 (1664 errors) saccades, respectively, were used to compute the distributions. For happy and fearful faces, the distributions contained 2196 (647 errors) and 2191 (743 errors) saccades.

3.2. Results

3.2.1. Accuracy

Paired samples t-tests performed on mean accuracy (Fig. 8a) indicated a significant effect of the Target ($t(19) = 3.91$, $p < .001$, $d = 0.87$) and a significant effect of the EFE target when the Target was an emotional face ($t(19) = -3.55$, $p = .002$, $d = 0.79$). Participants made more correct saccades (i.e., first saccades toward the target) when they were asked to saccade toward the emotional (M ± SD: 0.68 ± 0.11) rather than the neutral face ($M \pm SD$: .62 ± .13), and also

when the emotional Target face was happy (*M* ± *SD*: .70 ± .11) rather than fearful (*M* ± *SD*: .66 ± .11). Finally, we did not find any effect of the EFE distractor when the Target was neutral.

3.2.2. Latency

Paired samples t-tests performed on mean latency (Fig. 8b) showed only a significant effect of the Target ($t(19) = -3.16$, $p = .005$, $d = 0.71$). Saccades were elicited faster when participants were asked to saccade toward an emotional face (*M* ± *SD*: 249 ± 80.1 ms) rather than a neutral face (*M* ± *SD*: 277 ± 96.1 ms). No effect of the EFE was found for emotional targets and there was also no effect of the EFE distractor for neutral targets.

3.2.3. Minimum latency

The minimum latency (Fig. 8d) was located in the 150-ms bin for emotional faces, and in the 290-ms bin for neutral faces. For happy emotional face targets, the minimum latency was 220 ms, and for fearful emotional face targets, it was 260 ms.

3.2.4. Endpoints

Paired samples t-tests performed on the mean vertical distance between the image center and the endpoint (Fig. 8c) revealed only a main effect of the EFE for emotional face targets ($t(19) = -3.68$, $p = .002$, $d = 0.82$). Saccades landed higher when the emotional face was fearful (M ± SD: -1.02 ± 0.7°) than when it was happy (M ± SD: -1.11 ± 0.72°). We did not find any effect of the EFE distractor for neutral targets. Figure 9 presents examples of computed heat maps displayed on top of a randomly chosen face for each emotional condition.

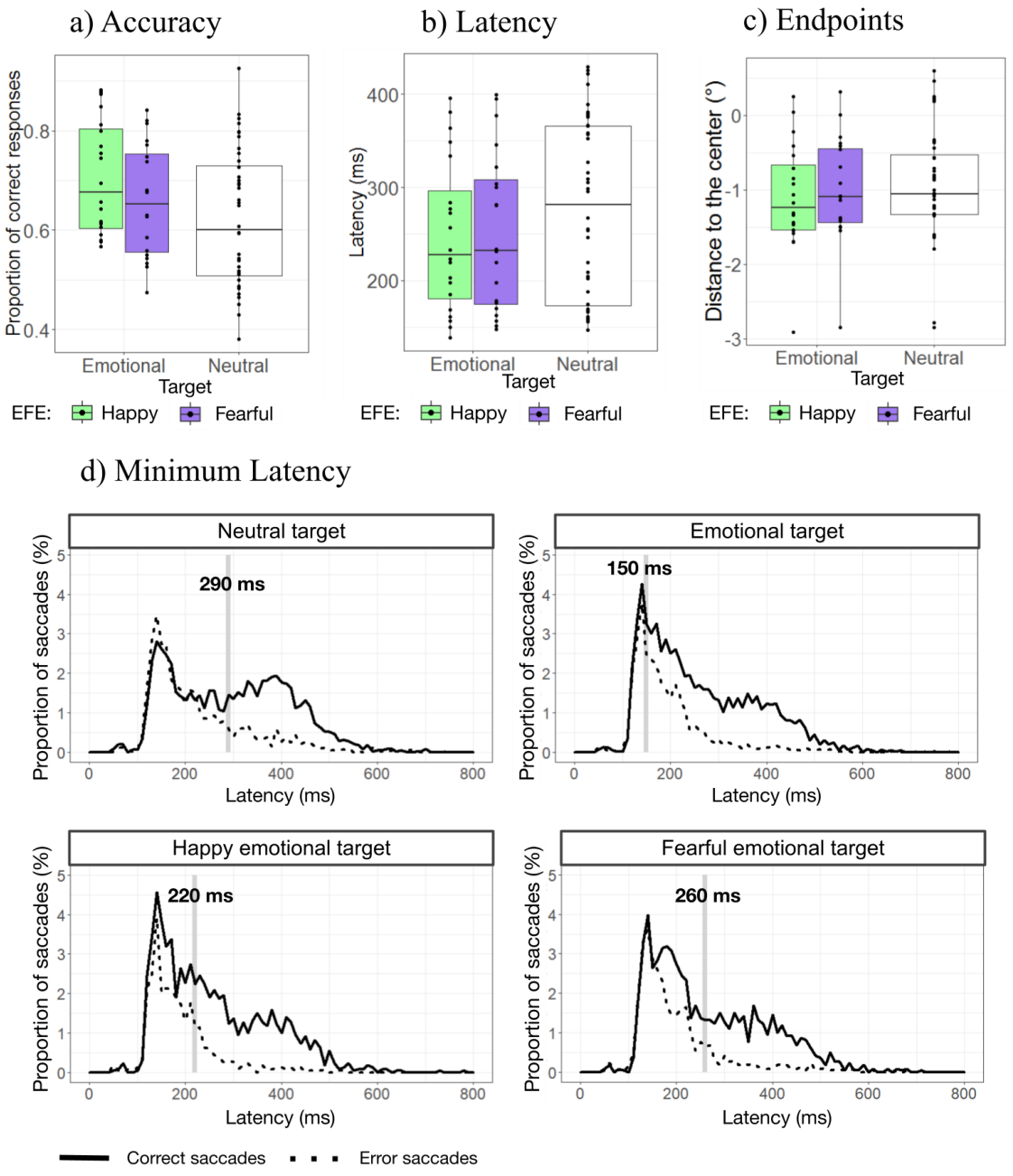

Figure 8: Boxplots for (a) mean proportion of correct responses, (b) mean latency (in ms), and (c) mean distance to the image center (in degrees of visual angle) for correct saccades as a function of the Target (Emotional and Neutral) and the Emotional Facial Expression of emotional targets (EFE; Happy or Fearful). (d) Distribution of saccade latencies for each Target, and for each Emotional Facial Expression of the emotional targets. Unbroken lines correspond to correct saccades and dotted lines to error saccades. The gray bar corresponds to the 10-ms bin containing the minimum latency It should be noted that, for the purposes of

illustration and because no significant effect of the EFE distractor was found, all types of emotional face distractors (happy or fearful) were recorded for neutral targets.

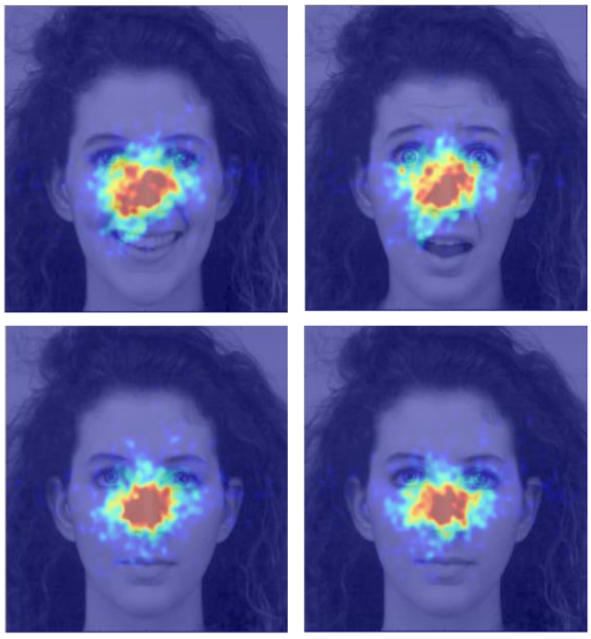

Figure 9: Heat maps computed from all first correct saccade endpoints in the second experiment when the target was a happy face (top left), a fearful face (top right), a neutral face when the distractor was happy (bottom left) or a neutral face when the distractor was fearful (bottom right).

3.3. Discussion

Results of Experiment 2 confirm that emotional faces are easier to detect than neutral faces during a saccadic task in which two faces are simultaneously displayed (Bannerman et al., 2009). First, saccades were more often correct when the target was the emotional face. This was especially the case when the emotional face was happy rather than fearful. Second, saccades were also elicited faster when the target was the emotional face. Surprisingly, however, there was no significant effect of facial expressions on mean saccade latencies, despite the fact that the minimum latency was higher for fearful than happy faces. The distribution of saccade latencies for correct and error saccades in Experiment 2 was quite different from that found in Experiment 1. First, there were more errors in Experiment 2, and we can see that most of the error saccades had short latencies (e.g., below 200 ms) and that they decreased in number as

latencies increased. Reliable saccades occurred at 150 ms for emotional targets, and 290 ms for neutral targets. Moreover, if we consider the distribution of responses in the 140 to 160-ms time windows, saccades tended to be made toward the emotional faces even when the target was the neutral face. It is therefore possible that saccades toward emotional faces are harder to control in this time window. Also, participants might have adopted the strategy of detecting the emotional face first and then deducing the position of the neutral face from this. Since the open mouth is more salient, it might be easier to detect and this would justify such a strategy. With regard to the saccade endpoints, these were still higher when the target was a fearful rather than a happy face. Compared to the first experiment (in which saccades tended to land around the eyes), saccades tended to land around the nose in Experiment 2. This might suggest that even if local features are able to capture attention in an automatic way, the weight allocated to those features can also be modulated by the task.

Overall, it is likely that a parsimonious hypothesis related to the simple perceptual properties of the stimuli is capable of explaining the happy face advantage that we observed in this experiment, without reference to emotional processes (Calvo et al., 2011). Fearful and neutral faces may be statistically more similar than happy and neutral faces. Using computational models, it has already been shown that in categorization tasks, happy faces are easier to recognize (i.e., elicit a higher rate of correct recognition) than fearful faces, and that happy faces are more different from neutral than from angry faces (Dailey et al., 2002; Mermillod et al., 2009). In the next section of the article, we describe an artificial neural network used to quantify the perceptual differences between neutral and fearful or happy faces. Even if we did not directly simulate saccadic responses, the task was nevertheless similar, as it involved finding the location of an emotional compared to a neutral face. More precisely, the neural network was trained and tested on its ability to discriminate between emotional-neutral and neutral-emotional face pairs (with the emotional face being either on the right or on the left side of the pair). Therefore, this task can be considered as a categorization task, as it makes it necessary to classify face pairs on the basis of two categories: the neutral-emotional category and the emotional-neutral category. Hence, its goal was to decide whether the emotional face was on the left or right side of the pair. Next, the network was tested and we computed its performance in the correct categorization of pairs of faces. The results were then further subdivided depending on whether the emotional face on the pair was happy or fearful. The network's performance might have been found to be better when the emotional face was fearful or when it was happy or it might have been the same in the two conditions. Based on results

from Experiment 2, our hypothesis was that the network would be able to discriminate the emotional and neutral faces better when the emotional face was happy.

4. Simulations

The whole simulation procedure was very similar to previous studies that have used a multi-layer perceptron (MLP) for emotion categorization (Dailey et al., 2002; Mermillod et al., 2009, 2010, 2019). It can therefore be subdivided into two steps: a pre-processing step in which Gabor filters are applied to the Fourier transform of the overall image, and a training-testing procedure using the MLP. We opted for this design (e.g., instead of a convolutional neural network) because previous studies have shown that, even though the method is less efficient for artificial intelligence purposes, the use of Gabor filters tuned at different spatial frequencies and orientation channels permits a more biologically plausible simulation of the primary visual cortex (Jones & Palmer, 1987) as well as of the phase invariance properties related to V1 complex cells (Hubel & Wiesel, 1968). Moreover, it produces results that are very similar to humans for the same facial emotion recognition task (Dailey et al., 2002, Li and Cottrell, 2010).

4.1. Method

4.1.1. Preprocessing and stimuli

There are many ways to reduce the size of images, but as we wanted to compare the performance of the network with that of our participants, we chose to do this in a biologically plausible way. Indeed, images were described using a bank of Gabor filters applied in the frequency domain. Each filter simulated the functioning of primary visual cells by being sensitive to a particular spatial frequency band and a particular orientation (Hubel & Wiesel, 1968; Jones & Palmer, 1987). Each image was described in terms of its energy at each filter output. More precisely, preprocessing began with the application of a Hanning window on each image to avoid boundary effects. The images were then transferred to the Fourier domain and Gabor filters were applied on the overall image. Therefore, each filter indicated the amount of energy in the image for one frequency channel and one orientation. We used a bank of forty-eight Gabor filters tuned to six spatial frequency channels (central frequencies = {0.82; 1.23; 1.85; 2.78; 4.14; 6.25}, given in cycles/degree) and eight orientations (0, pi/8, 2pi/8, 3pi/8, 4pi/8, 5pi/8, 6pi/8, 7pi/8). The 48 filters were applied to each image and the energy of each filtered image was computed, resulting in an energy vector with 48 values. Each value corresponded to the local energy spectra of the image in the spectral domain multiplied by the kernel of the Gabor filter in a specific spatial frequency band at a specific orientation. Finally,

each image was described by a 48-length vector. These values were normalized between 0 and 1 across all faces and all emotions. Inputs were fed into a MLP, whose task was to associate the descriptor of the face with the output vector of the category.

In order to match the experimental procedure of Experiment 2, we gathered the image vectors together in pairs corresponding to all the possible combinations that could occur in Experiment 2. Consequently, these pairs were necessarily composed of faces of different genders and with different facial expressions: neutral on one side of the vector and emotional (either happy or fearful) on the other side. The association of the vectors was computed by simply concatenating the two vectors, and each pair was represented twice, one with the emotional face on the right and the other with the emotional face on the left. Each pair was also associated with a specific label, which can be considered to represent the probability of each category (the emotional-neutral and the neutral-emotional category): [1, 0] if the emotional face was on the left and [0, 1] if the emotional face was on the right. A visual representation of the preprocessing steps and neural network is presented in Fig. 10.

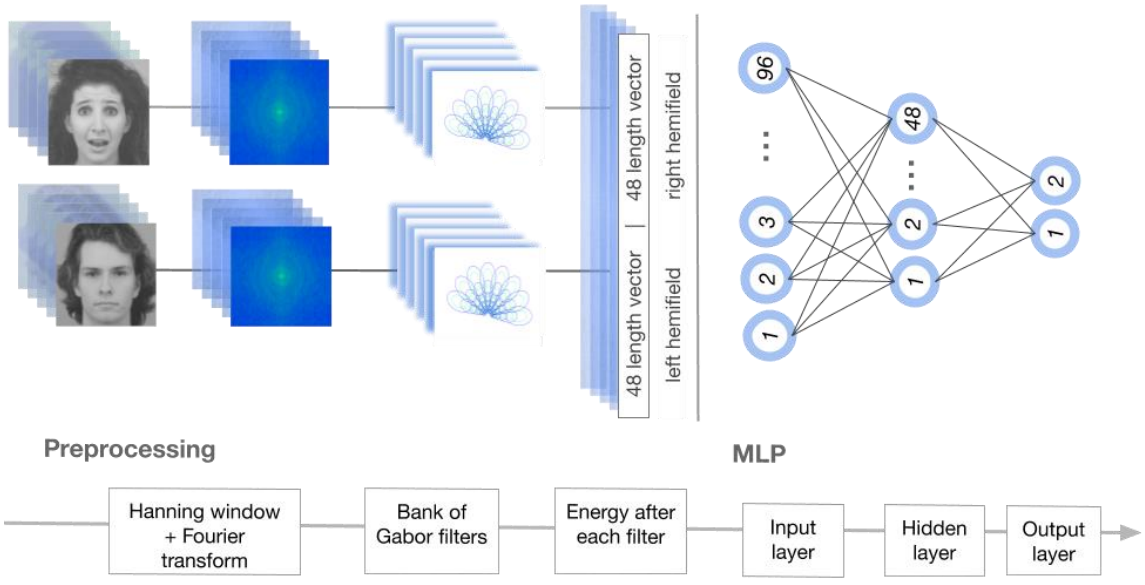

Figure 10: Representation of preprocessing steps and neural network (3-layer MLP with fully interconnected neurons). The module for the Fourier transform is presented for the purposes of illustration.

4.1.2. Network architecture

The purpose of the network was to simulate a discrimination task, similar to the one presented in Experiment 2, involving a simultaneously presented emotional (happy, fearful)

and neutral face. Therefore, the input was not simply a single vector corresponding to a single image, but a combination of two vectors corresponding to a pair of images. The network was trained to discriminate emotional-neutral face pairs from neutral-emotional face pairs and its architecture consisted of 3 different layers with 96 input units, 48 hidden units and 2 output units. On the last layer, a standard sigmoid transfer function was used as an output function, given by:

$$f(x) = \frac{1}{1 + e^{-x}}$$

where x is the weighted input to the layer (i.e., the sum of the layer input multiplied by the weight matrix, which has random initial weights). The standard back-propagation algorithm was used for synaptic weight adjustment during training, with a learning rate set to 0.01, and the Adam algorithm (Kingma & Ba, 2017) for optimization. The error signal used for synaptic weight correction was computed based on the Mean Squared Error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

where n is the number of values in the vector, y the expected output, and $\hat{y}$ the actual output.

4.1.3. Procedure

At the beginning of the procedure, 40 of the 60 different faces from Experiment 2 (20 women) were randomly selected for training and 20 of them (10 women) for testing. Thus, none of the tested faces were used in the training phase and they were therefore unknown to the network. We picked the corresponding vector pairs among all possible vector pairs. As we did not want any individuals to be present in both the training and testing phase, no pairs that contained both a training and a test individual were used. This led to the selection of 800 pairs for testing and 3200 pairs for training for each training-test procedure.

After selecting the training-test sets, the 96-length energy vectors corresponding to the training set were fed into the network using the standard backpropagation algorithm. The network associated the training input vectors with the corresponding output vector over 500 iterations. All pairs were forwarded and back-propagated to the network at the same time and the gradient was thus computed using the whole training dataset (i.e., a batch gradient descent method was used). After learning, the network was tested on 800 new pairs of faces. The output from the model was a vector with 2 values, corresponding to the probabilities of the pair being associated with the emotional-neutral category and the neutral-emotional category. The network classification response was then assigned to the class with the highest probability (i.e., a winner-takes-all procedure was applied), and the accuracy was set to 1 if the expected and

actual responses were the same, and 0 if not. The same training-test procedure was repeated over 50 iterations in order to calculate a stable and reliable average accuracy from 50 different networks.

4.2. Results

On average, the network correctly categorized 88% (*M* ± *SD*: *.88* ± .006) of the new tested face pairs. A paired samples t-test was performed in order to test for significant differences between neutral-happy and neutral-fearful face pairs (Fig. 11). We found that the network was better able to discriminate neutral from emotional faces when the emotional face was happy (*M* ± *SD*: .89 ± .006) than when it was fearful (*M* ± *SD*: .87 ± .005; $t(1,49) = 2.55$, $p = .014$, $d = 0.32$).

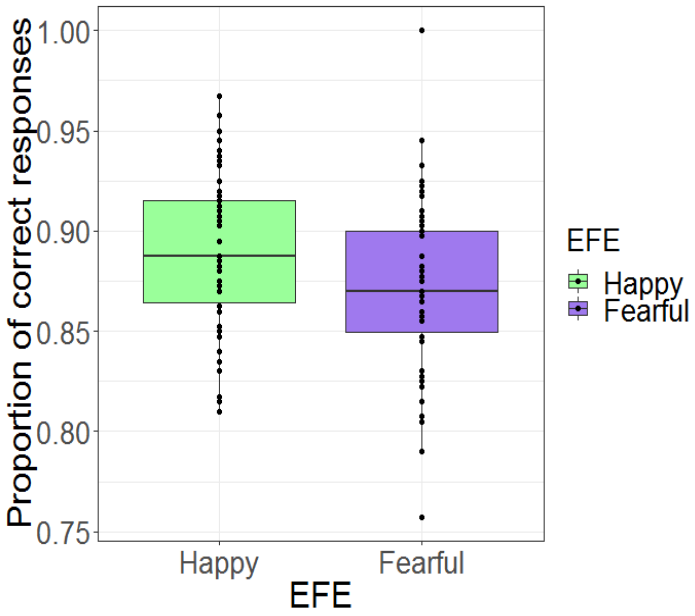

Figure 11: Boxplot for mean accuracy of the model depending on the EFE of the emotional face (happy, fearful).

4.3. Discussion

We used an artificial neural network as a tool to quantify the perceptual differences between neutral and happy compared to neutral and fearful faces. In a similar way to Experiment 2, the neural network reproduced the access to the information from the two hemifields and had to decide which side the emotional face was on based on the perceptual features provided by each hemifield. It is important to clarify that the aim was not to simulate

the processing of facial expression perception in the brain, but rather to compare the results obtained from participants and those from an artificial neural network. Overall, results showed that, with an average accuracy of 88%, the network performed better than the participants (on average 65-70%). This can be explained by the fact that saccadic eye movements can be elicited in a bottom-up fashion and thus do not always follow participants' top-down goals. For example, in one and the same task, participants made more errors with saccadic compared to manual responses (Bannerman et al., 2009). Therefore, the fact that we used a saccadic choice rather than a manual response task in Experiment 2 could have led to a higher rate of erroneous responses.

In line with our hypothesis, we found that the network performances were better when the discrimination involved a happy than a fearful face. This implies that, at a purely perceptual level, fearful and neutral faces may be more similar than happy and neutral faces. Consequently, the fact that participants performed better on happy faces in Experiment 2 might be explicable in terms of perceptual factors. Although we showed that happy faces have a perceptual advantage over fearful faces, our model does not allow us to reject the possibility of a contribution of emotion to the happy face advantage in Experiment 2. For example, we cannot exclude the possibility that happy faces might be prioritized due to their positive valence or because they are encountered more often in everyday life (Bond & Siddle, 1996; Leppänen et al., 2004).

5. General discussion

The purpose of this study was to assess the impact of facial expressions on selection processes. In the first experiment, we wanted to test the impact of facial expressions on very fast face detection. As previously observed, face targets elicited very fast and accurate saccadic responses, whereas participants took longer and made more errors when required to saccade toward vehicle targets. With regard to the effect of facial expressions, we found that they did not influence performances, thus suggesting that emotional faces, whether happy or fearful, are not automatically (i.e., quickly and non-intentionally) prioritized over neutral faces. Nevertheless, saccade endpoints were modulated by facial expressions. Saccades landed lower when the face was happy than when it was fearful or neutral and also when the face was fearful rather than neutral. In the second experiment, we directly tested the detection of neutral and emotional faces. Emotional faces elicited faster and more accurate responses than neutral ones. Also, accuracy was higher when the emotional face was happy, and saccades landed lower for

happy than for fearful face targets. We can note that latencies in Experiment 2 were higher than in Experiment 1 (with mean latencies around 250 ms in Experiment 2 and 170 in Experiment 1), suggesting that faces are detected before expressions are explicitly decoded.

A prioritization of emotional faces?

While emotions modulated performances in the second experiment, emotional faces did not facilitate face detection in the first experiment even though some visual features seemed to differentially attract the gaze depending on facial expression., This observation runs contrary to the idea that attention is reflexively captured by emotional, and especially threatening, events due to evolutionary needs (e.g., Öhman, 2005; Öhman et al., 2001). Nevertheless, similar results (i.e., a fast oculomotor capture by faces irrespective of their expressions) were obtained in an earlier study. Indeed, Devue and Grimshaw (2017) tested the automatic prioritization of non-task-relevant emotional faces in a task in which faces were known to attract the gaze (Devue et al., 2012). A circular array of colored dots was displayed on the screen and participants had to make a saccade toward a color singleton. Pictures of different irrelevant objects (including a neutral or an angry face) were displayed in a concentric circle inside the dot array. The authors found that irrelevant faces attracted the gaze more than other objects but that this occurred irrespective of the expression. In their study, saccades toward faces were elicited quickly, with mean saccade latencies around 200 ms. In fact, this lack of modulation could be explained by a ceiling effect (Mulckhuyse, 2018). Indeed, one might consider that the visual system has evolved so that faces, which are more likely to be socially relevant than other objects, can be detected very rapidly based on low-level features (Baron-Cohen 1995; Haxby, et al., 2000; Leopold et al., 2010), or that visual features have evolved to be easily detected (Emery, 2000; Kobayashi & Kohshima, 1997; Lacruz et al., 2019; Wu et al., 2013). Whether this detection is based on isolated features (i.e., the eyes; Kauffmann et al., 2021; Lewis & Edmonds, 2003) or AS information (Crouzet & Thorpe, 2011; Honey et al, 2008), it may be insufficient to decode expressions. The goal might therefore be to direct the face into central vision and then proceed to further investigations in order to extract more features such as the emotional expression.

Also, in Experiment 1, facial expressions were not task-relevant. However, we suggest that this is not sufficient to explain the lack of emotional modulation. Indeed, there are studies that have used tasks in which expressions were not relevant and which have nevertheless found that they modulated performances. This is the case, for example, of probe categorization tasks in which an emotional face is briefly presented as a prime followed by an emotional probe word

or a visual scene. The participants' task is to ignore the face (which is task-irrelevant) and judge the probe as pleasant or unpleasant. Classically, reaction times are faster when the prime and the probe are affectively congruent, for example for positive words following an expression of happiness (Aguado et al., 2007; Lipp et al., 2009; McLellan et al., 2010; Sassi et al., 2014). Another example can be found in gender categorization tasks. It has been shown that even if expressions are task-irrelevant they can still interfere with gender perception. For example, it has been shown that a happy or a fearful expression biases discrimination toward females (Hess et al., 2009) or, similarly, that gender implicitly interferes with the recognition of emotional expressions (Villepoux et al., 2015). It should be noted that these studies used manual responses, causing reaction times to be relatively high (about 600 ms). If the decoding of facial expressions only occurs after 180-200 ms (Schyns et al., 2009; Kulke, 2019), it is more likely that the lack of emotional prioritization in Experiment 1 was due to the short time window in which saccades were elicited, and that the very fast gaze capture by faces may have preceded emotional facilitation.

In Experiment 2, emotional faces were detected faster and attracted the gaze more often than neutral faces. This confirms that emotional faces, when decoded, can facilitate the orienting of attention. Nevertheless, there is still some doubt as to whether this process is supported by the interpretation of the emotional content. For example, it is possible that when looking for neutral or emotional faces, participants choose to check for emotional faces first, for example by looking for an open mouth, which is the most salient feature (Calvo et al., 2008; Horstmann et al., 2012; Stuit et al., 2021). Such a strategy would prioritize emotional faces, not because they are meaningful, but because they have configurations that make them easier to see (because happy and most fearful faces do have an open mouth). In support of this idea, one study which compared emotion detection and emotion categorization suggests that visible teeth are particularly useful for emotion detection (Sweeny et al., 2013). In Experiment 2, we also found that happy faces led to higher response accuracy than fearful faces. As suggested by neural computations, this last result could be explained by the fact that happy and neutral faces are more perceptually different than fearful and neutral faces (Mermillod et al., 2009).

Moreover, we found in an additional analysis (see Appendix B) that saccades toward faces in Experiment 1 were elicited faster when the face was presented on the left (i.e. projected in the right hemisphere) compared to the right side of the screen (i.e. projected in the left hemisphere). This result agrees with previous papers using a saccadic choice task with faces and vehicles (Crouzet et al.,2010; Guyader et al., 2017). It is also consistent with theories on the cortical lateralization of face processing which argue in favor of right hemisphere

specialization (Carlei, et al., 2017; Ellis, 1983). Interestingly, saccade endpoints were also lower for targets presented on the left side of the screen. This could also be the result of an enhanced processing of faces in the left hemifield. In Experiment 2, the greater accuracy for emotional than neutral faces was only significant for targets presented on the left side of the screen. This result is consistent with one hypothesis concerning the lateralization of emotional processing which argues in favor of right hemisphere specialization (Demaree et al., 2005).

Even if such results are not particularly suitable for inferring neurophysiological implications, they can nevertheless be considered in the light of recent neurophysiological data. Indeed, there is evidence that face information can be processed via a subcortical pathway connecting the amygdala, the superior colliculus, and the pulvinar (Johnson, 2005). Even though some authors believe that this pathway is modulated by facial expressions (Bayle & Taylor, 2010; Bayle et al., 2009; LeDoux, 2000; Méndez-Bértolo et al., 2016; Vuilleumier et al., 2003), other studies suggest that it could be face-specific and independent of the expressed emotion (Fitzgerald et al., 2006; Garvert et al., 2014; Johnson, 2005; McFadyen et al., 2017). For example, McFadyen et al. used magnetoencephalography (MEG) and dynamic causal modeling (DCM) of participants making gender judgments of neutral and fearful faces to identify the underlying neural networks most likely to carry information to the amygdala. They demonstrated that the most likely subcortical network consisted of a pulvinar-amygdala connection that was not influenced by facial expressions. These results therefore suggest that the emotional content of visual stimuli may not necessarily be the key for entry into the subcortical pathway, whereas a more stereotypical face pattern would be. Overall, even if we cannot affirm that face detection in Experiment 1 was supported by such a pathway (e.g., it could also be supported by the ventral pathway; Crouzet et al., 2010), our results would be consistent with this view.

Finally, we can wonder what would have happened if we had used dynamic stimuli. Indeed, some studies have suggested that the use of dynamic stimuli enhances the processing of fearful faces, and, one functional magnetic resonance imaging (fMRI) study identified enhanced neural activity in response to dynamic fearful but not happy faces (Sato et al., 2004). In addition, an EEG study has shown that dynamic threatening stimuli (e.g., spiders) elicited higher P1 activity than static stimuli or dynamic non-threatening stimuli (Hinojosa et al., 2009). These findings suggest that motion provides additional saliency to threatening stimuli and thus facilitates their detection. However, there are still only a few studies on this topic and the results are sometimes contradictory. For example, another study found no differences between the responses to

dynamic neutral, happy, disgusted and fearful facial expressions (Van der Gaag et al., 2007). It is also likely that the results will be similar for both dynamic and static expressions. Studies using static stimuli similar to those that we used in our study (i.e., basic emotions, with the peak frame of the emotion) have reported a lack of a dynamic advantage for expression recognition (Fiorentini & Viviani, 2011; Gold et al., 2013; see for reviews Kätsyri, 2006; Dobs et al., 2018). Moreover, saccades are likely to be elicited at the very beginning of the video sequence, in particular in our first experiment.

Saccade endpoints as a reflection of the distribution of attention within the face?

In both experiments, we found that saccades landed lower when the target was a happy face than when it was a fearful one. This observation was expected given the visual saliency and diagnosticity of the mouth in happy faces (Calvo & Nummenmaa, 2008; Smith et al., 2005). However, saccades also landed lower when the target was a fearful face than when it was a neutral one in Experiment 1. This can be explained by the fact that in fearful faces, both the eyes and the mouth attract attention. Indeed, in some fearful faces, the mouth is open and this can shift attention toward it more than neutral faces, in which the mouth is always closed. In line with this view, some studies found that both the eyes and the mouth play a critical role in the recognition of fear (Eisenbarth & Alpers, 2011). Given the results of the first experiment, we could have expected that saccades would be lower in the emotional than in the neutral task in Experiment 2. This effect was only marginally significant for happy faces and we suggest that this is due to the small sample size. Nevertheless, perceptual saliency may not be the only factor that can influence the way attention is distributed within the face.

As attested by the heat maps, endpoints were generally located around the eyes in the first experiment, and more around the nose for the second experiment. Such differences between the first and the second experiment can be explained by the different tasks. It is likely that when participants have to process facial expressions, their gaze is naturally more oriented toward the mouth. It is also important to underline that because the eyes were located in the center of the image in this study, it is possible that saccades tended to land around the eyes because of a center of gravity bias (Bindemann, 2010; Tseng et al., 2009; Parkhurst & Niebur, 2003; Tatler, 2007). Furthermore, even if we assume that the positions of the eyes and mouth were the same for all images, some slight differences might nevertheless have occurred, again due to the different, emotion-related shapes of the eyes and the mouth. To explore such differences, we performed an additional analysis (see Appendix A) demonstrating that the positions of the

vertical midline of the eyes and the mouth were not uniform across expressions. Indeed, the mouth position was higher on average in the image of happy than those of neutral or fearful faces, as well as in the images of neutral faces than fearful ones. Also, the mean eye position was higher for fearful than for happy or neutral faces, and for happy than for neutral faces. Therefore, the difference between saccade endpoints on happy and fearful faces (i.e., lower endpoints on happy faces) could be explained by the fact that the eyes are lower in happy faces. However, this explanation on its own would not explain the difference between neutral and fearful faces in Experiment 1 (lower endpoints on fearful faces) since, based on the eye position, we would have expected lower endpoints for neutral faces.

Overall, we assume that attention is shared between multiple locations during saccade programming and that saccade endpoints can reflect the interactions between multiple loci of attention. In this context, when a saccade is executed, the endpoint would reflect the allocation of attention within the face but would not necessarily indicate the exact location that has captured most of the attention. For example, when a saccade has to be executed toward a target (e.g., a dark gray ring) in the presence of a close distractor (e.g., a dark ring), the saccade lands somewhere in-between the target and the distractor. Van der Stigchel and de Vries (2015) have suggested that attention is directed toward the target and distractor location rather than at the intermediate location (where the saccade lands). This is consistent with neurophysiological recordings in such configurations which have shown that peaks of activity in the superior colliculus are focused on cells coding for target and distractor locations (Edelman & Keller, 1998). This phenomenon is often referred to as global effect or saccade averaging, and has, to our knowledge, only been studied using very basic stimuli such as circles with line drawings (Findlay, 1982; Van der Stigchel & Nijboer, 2013). Given that we compared emotional faces, we suggest that attention is mostly shared between face parts which are of diagnostic relevance for expression decoding, which have been shown to be the eyes or the mouth (Eisenbarth & Alpers, 2011; Smith et al., 2005; Wegrzyn et al., 2017). We assume that if the endpoint is lower in one condition, it means that the mouth attracted most attention during saccade programming. Following this view, even though saccade endpoints were located around the nose in the second experiment, this does not necessarily mean that attention was directed there.

6. Conclusions

The present study confirms previous results showing that face stimuli can be detected very efficiently and provides new insights concerning the interaction between facial expression processing and the oculomotor system. Experiment 1 shows that very fast face detection was

not modulated by facial expressions, suggesting that emotional faces, whether fearful or happy, are not automatically prioritized over neutral ones. Experiment 2 showed that emotional faces are detected better than neutral ones in a task in which participants are explicitly asked to process expressions. We suggest that the lack of emotional prioritization in Experiment 1 is due to the short time window in which saccades were elicited (with mean saccade latencies around 170 ms). Indeed, as suggested by some previous studies (Devue et al., 2012; Kulke, 2019; Schyns et al., 2009), fast face detection may occur before expressions are decoded. Experiment 2 also found that it was easier to discriminate between neutral and emotional faces in the case of happy rather than fearful faces. Using computational modeling, we showed that this can at least be partly explained by perceptual factors, as the performances of a neural network were also better with happy faces. Finally, an analysis of saccade endpoints revealed a modulation by facial expressions, even for saccades that were elicited very quickly. We suggest that salient local face features, like the mouth, can automatically shift attention toward themselves and all the more so when they are task-relevant.

7. APPENDICES

APPENDIX A. Analysis of the position of the eyes and the mouth in face images depending on expressions

Here, we present an additional analysis conducted in order to compare the image positions of the eyes or mouth between expressions. More precisely, for each image, we manually drew a rectangular box surrounding the eyes and another surrounding the mouth, as presented in Fig. A.1. We tried to select a rectangle that was as close as possible to the eyes or the mouth but without touching the eyeball or lips, respectively. For each box, we computed the vertical midline (see purple lines in Fig. A.1). We then computed two measurements: (1) the vertical distance between the center of the image (corresponding to a Y coordinate of 150 pixels) and the vertical midline of the eyes, and (2) the vertical distance between the center of the image and the vertical midline of the mouth. T-tests were performed on mean values for each expression to quantify the effect of the Emotional Facial Expression (Fearful, Happy, Neutral) on the position of the eyes and mouth. Measures are in degrees of visual angle to match the endpoint analysis that is presented in the manuscript. In our experimental setup, 1 degree corresponds to 25 pixels.

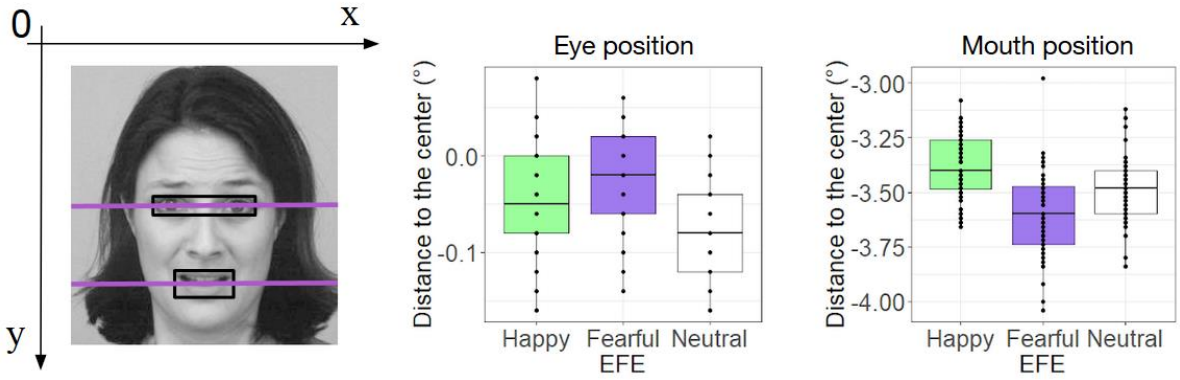

Figure A.1: Example of rectangular boxes surrounding the eyes and the mouth with their vertical midlines (purple lines; left), and boxplots for the vertical distances between the image center and the eyes (middle), and between the image center and the mouth (right). A negative distance means that the feature is below the center.

Results show that (1) Vertical distance between the center of the image and the center of the eyes was smaller for neutral ($M \pm SD$: -0.076 ± 0.043 °) than for happy ($M \pm SD$: -0.05 ± 0.052 °, $t(1,59) = 2.97$, $p = .004$, $d = 0.54$) or fearful ($M \pm SD$: -0.024 ± 0.05 °, $t(1,59) = 6.13$, $p < .001$, $d = 1.12$) faces, and for happy than for fearful faces ($t(1,59) = -2.84$, $p = .005$, $d = 0.52$), and that (2) Vertical distance between the center of the image and the center of the mouth was smaller for fearful ($M \pm SD$: -3.60 ± 0.19 °) than for happy ($M \pm SD$: -3.38 ± 0.14 °, $t(1,59) = 7.31$, $p < .001$, $d = 1.34$) or neutral ($M \pm SD$: -3.49 ± 0.14 °, $t(1,59) = -3.8$, $p < .001$, $d = 0.69$) faces, and for neutral than for happy faces ($t(1,59) = 4.22$, $p < .001$, $d = 0.77$).

In conclusion, based on the regions that we manually selected, there were significant differences concerning the positions of the eyes and the mouth. The differences in the positions of the eyes were small (with the greatest mean difference being 1.25 pixels). The difference was larger when we compared the mean position of the mouth for the three emotions (with the greatest mean difference being 5.5 pixels).

APPENDIX B. The effect of the target location

Here, we present an additional analysis conducted in order to compare participants' performance and saccade endpoints depending on the target location (i.e., left or right visual hemifield). For each experiment, we performed (1) a repeated measures ANOVA on the mean accuracy, mean latency and mean distance to the center, with the Target (Face, Vehicle for Experiment 1; Emotional, Neutral for Experiment 2) and the Location (Left, Right) as between-

subject factors to test for the effect of the Location and for an interaction between Target and Location. Then, for each Target, we performed (2) a repeated measures ANOVA on the mean accuracy, latency and distance to the center of one target condition (face condition for Experiment 1, emotional condition for Experiment 2), with the EFE (Happy, Fearful, Neutral for Experiment 1; Happy, Fearful for Experiment 2) and Location (Left, Right) as between-subject factors to test for an interaction between the EFE of the target and the Location. This procedure was repeated for the other target condition (vehicle condition for Experiment 1, neutral condition for Experiment 2) to test for an interaction between distractor Location and EFE. When required, paired samples t-tests were used for pairwise comparisons. Neither main effects of the Target and the EFE (which are reported in the core paper) nor non-significant effects are reported.

Experiment 1: Face vs Vehicle

Accuracy: A repeated measures ANOVA performed on mean accuracy when the target was a face indicated a marginal effect of Location ($F(1,60) = 3.71$, $p = .059$, $\eta_p^2 = 0.029$). Saccades tended to be more accurate when the face appeared in the left (*M* ± *SD*: .89 ± .087) than the right (*M* ± *SD*: .87 ± .099) hemifield.

Latency: A repeated measures ANOVA performed on mean latency indicated an interaction between the Target and the Location ($F(1,60) = 15.1$, $p < .001$, $\eta_p^2 = 0.2$). When the target was a face, saccades were elicited faster when it was presented in the left (*M* ± *SD*: 178 ± 21.2 ms) than in the right (*M* ± *SD*: 174 ± 23 ms; $p < .001$) visual hemifield.

Endpoints: A repeated measures ANOVA performed on mean distance to the center indicated a significant effect of Location ($F(1,60) = 7.09$, $p < .001$, $\eta_p^2 = 0.11$). Saccades landed lower in the left (*M* ± *SD*: -0.34 ± 0.55) than in the right (*M* ± *SD*: -0.23 ± 0.48) hemifield.

Experiment 2: Emotional vs Neutral face

Accuracy: A repeated measures ANOVA performed on mean accuracy indicated a significant interaction between Target and Location ($F(1,19) = 6$, $p < .001$, $\eta_p^2 = 0.024$). The difference between emotional and neutral faces was only significant for the left hemifield (*accuracy for emotional faces:* *M* ± *SD*: .7 ± .16; *accuracy for neutral faces*: *M* ± *SD*: .59 ± .18; $p < .001$).

Latency and Endpoints: A repeated measures ANOVA performed on mean latency or mean distance to the center indicated no significant effect of Location, nor any interaction with other factors.

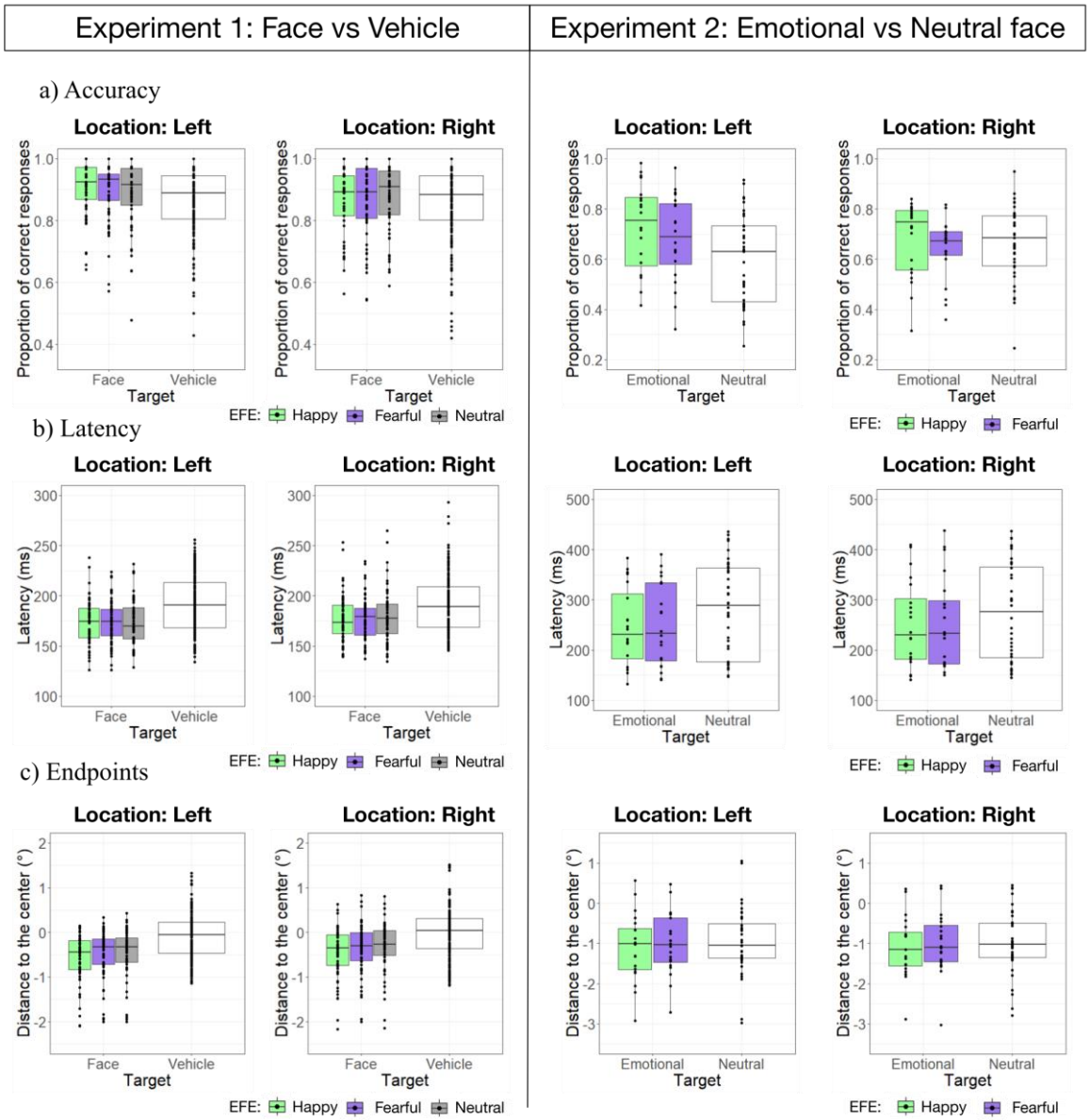

Figure A.2: Boxplots for (a) mean proportion of correct responses, (b) mean latency (in ms), and (c) mean distance to the image center (in degrees of visual angle) for correct saccades according to the Target, the Location of targets, and the Emotional Facial Expression of targets, for Experiment 1 (left) and Experiment 2 (right).

8. Acknowledgments

This work was supported by NeuroCoG IDEX UGA in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02). This work has been partially supported by MIAI @ Grenoble Alpes, (ANR-19-P3IA-0003) to Martial Mermillod.

9. Conflict of interest

The authors declare no conflict of interest.

10. References

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience*, *4*(3), 165–178. https://doi.org/10.1038/nrn1056

Aguado, L., Garcia-Gutierrez, A., Castañeda, E., & Saugar, C. (2007). Effects of Prime Task on Affective Priming By Facial Expressions of Emotion. *The Spanish Journal of Psychology*, *10*(2), 209–217. https://doi.org/10.1017/S1138741600006478

Bannerman, R. L., Milders, M., & Sahraie, A. (2009). Processing emotional stimuli: Comparison of saccadic and manual choice-reaction times. *Cognition & Emotion*, *23*(5), 930–954. https://doi.org/10.1080/02699930802243303

Baron-Cohen, S. (1995) The eye direction detector (EDD) and the shared attention mechanism (SAM): two cases for evolutionary psychology. In Joint Attention: Its Origins and Role in Development (Moore, C. andDunham, P.J., eds), pp. 41–59, Lawrence Erlbaum

Bayle, D. J., & Taylor, M. J. (2010). Attention inhibition of early cortical activation to fearful faces. *Brain Research*, *1313*, 113–123. https://doi.org/10.1016/j.brainres.2009.11.060

Bayle DJ, Henaff M-A, Krolak-Salmon P (2009) Unconsciously Perceived Fear in Peripheral Vision Alerts the Limbic System: A MEG Study. PLoS ONE 4(12): e8207. https://doi.org/10.1371/journal.pone.0008207

Becker, D. V., Anderson, U. S., Mortensen, C. R., Neufeld, S. L., & Neel, R. (2011). The face in the crowd effect unconfounded: Happy faces, not angry faces, are more efficiently detected in single- and multiple-target visual search tasks. *Journal of Experimental Psychology: General*, *140*(4), 637–659. https://doi.org/10.1037/a0024060

Belopolsky, A. V. (2015). Common Priority Map for Selection History, Reward and Emotion in the Oculomotor System. *Perception*, *44*(8–9), 920–933. https://doi.org/10.1177/0301006615596866

Bindemann, M. (2010). Scene and screen center bias early eye movements in scene viewing. *Vision Research*, *50*(23), 2577–2587. https://doi.org/10.1016/j.visres.2010.08.016

Bisley, J. W., & Mirpour, K. (2019). The neural instantiation of a priority map. *Current Opinion in Psychology*, *29*, 108–112. https://doi.org/10.1016/j.copsyc.2019.01.002

Bond, N. W., & Siddle, D. A. T. (1996). The preparedness account of social phobia: Some data and alternative explanations. In R.M. Rapee (Ed.), Current controversies in the anxiety disorders (pp. 291–316). London: Guilford Press.

Boucart, M., Lenoble, Q., Quettelart, J., Szaffarczyk, S., Despretz, P., & Thorpe, S. J. (2016). Finding faces, animals, and vehicles in far peripheral vision. Journal of vision, 16(2), 10-10.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior Research Methods*, *40*(1), 109–115. https://doi.org/10.3758/BRM.40.1.109

Calvo, M. G., & Marrero, H. (2009). Visual search of emotional faces: The role of affective content and featural distinctiveness. *Cognition & Emotion*, *23*(4), 782–806. https://doi.org/10.1080/02699930802151654

Calvo, M. G., & Nummenmaa, L. (2008). Detection of emotional faces: Salient physical features guide effective visual search. *Journal of Experimental Psychology: General*, *137*(3), 471–494. https://doi.org/10.1037/a0012771

Calvo, M. G., & Nummenmaa, L. (2009). Eye-movement assessment of the time course in facial expression recognition: Neurophysiological implications. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(4), 398–411. https://doi.org/10.3758/CABN.9.4.398

Calvo, M. G., & Nummenmaa, L. (2011). Time course of discrimination between emotional facial expressions: The role of visual saliency. *Vision Research*, *51*(15), 1751–1759. https://doi.org/10.1016/j.visres.2011.06.001

Calvo, M. G., & Nummenmaa, L. (2015). Perceptual and affective mechanisms in facial expression recognition: An integrative review. *Cognition and Emotion*, *30*(6), 1081–1106. https://doi.org/10.1080/02699931.2015.1049124

Carlei, C., Framorando, D., Burra, N., & Kerzel, D. (2017). Face processing is enhanced in the left and upper visual hemi-fields. *Visual Cognition*, *25*(7-8), 749-761.

Cerf, M., Frady, E. P., & Koch, C. (2009). Faces and text attract gaze independent of the task: Experimental data and computer model. *Journal of Vision*, *9*(12), 10–10. https://doi.org/10.1167/9.12.10

Coutrot, A., & Guyader, N. (2014). How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of Vision*, *14*(8), 5. https://doi.org/10.1167/14.8.5

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

Crouzet, S. M., & Thorpe, S. J. (2011). Low-level cues and ultra-fast face detection. *Frontiers in psychology*, *2*, 342.

Dailey, M. N., Cottrell, G. W., Padgett, C., & Adolphs, R. (2002). EMPATH: A Neural Network that Categorizes Facial Expressions. *Journal of Cognitive Neuroscience*, *14*(8), 1158–1173. https://doi.org/10.1162/089892902760807177

Demaree, H. A., Everhart, D. E., Youngstrom, E. A., & Harrison, D. W. (2005). Brain lateralization of emotional processing: historical roots and a future incorporating "dominance". *Behavioral and cognitive neuroscience reviews*, *4*(1), 3-20.

Devue, C., Belopolsky, A. V., & Theeuwes, J. (2012). Oculomotor Guidance and Capture by Irrelevant Faces. *PLoS ONE*, *7*(4), e34598. https://doi.org/10.1371/journal.pone.0034598

Devue, C., & Grimshaw, G. M. (2017). Faces are special, but facial expressions aren't: Insights from an oculomotor capture paradigm. *Attention, Perception, & Psychophysics*, *79*(5), 1438–1452. https://doi.org/10.3758/s13414-017-1313-x

Dobs, K., Bülthoff, I., & Schultz, J. (2018). Use and usefulness of dynamic face stimuli for face perception studies—A review of behavioral findings and methodology. *Frontiers in psychology*, *9*, 1355.

Edelman, J. A., & Keller, E. L. (1998). Dependence on Target Configuration of Express Saccade-Related Activity in the Primate Superior Colliculus. *Journal of Neurophysiology*, *80*(3), 1407–1426. https://doi.org/10.1152/jn.1998.80.3.1407

Eisenbarth, H., & Alpers, G. W. (2011). Happy mouth and sad eyes: Scanning emotional facial expressions. *Emotion*, *11*(4), 860–865. https://doi.org/10.1037/a0022758

Ellis, H. D., & Young, A. (1983). The role of the right hemisphere in face perception. *Functions of the right cerebral hemisphere*, 33-64.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & biobehavioral reviews*, *24*(6), 581-604.

Farah, M. J., Wilson, K. D., & Drain, M. (1998). *What Is 'Special' About Face Perception?* 17.

Fecteau, J., & Munoz, D. (2006). Salience, relevance, and firing: A priority map for target selection. *Trends in Cognitive Sciences*, *10*(8), 382–390. https://doi.org/10.1016/j.tics.2006.06.011

Findlay, J. M. (1982). Global visual processing for saccadic eye movements. *Vision Research*, *22*(8), 1033–1045. https://doi.org/10.1016/0042-6989(82)90040-2

Fiorentini, C., & Viviani, P. (2011). Is there a dynamic advantage for facial expressions?. *Journal of Vision*, *11*(3), 17-17.

Fitzgerald, D. A., Angstadt, M., Jelsone, L. M., Nathan, P. J., & Phan, K. L. (2006). Beyond threat: Amygdala reactivity across multiple expressions of facial affect. *NeuroImage*, *30*(4), 1441–1448. https://doi.org/10.1016/j.neuroimage.2005.11.003

Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition*, *117*(3), 319–331. https://doi.org/10.1016/j.cognition.2010.09.003

Fox, E., & Damjanovic, L. (2006). The eyes are sufficient to produce a threat superiority effect. *Emotion*, *6*(3), 534–539. https://doi.org/10.1037/1528-3542.6.3.534

Fox, J., Friendly, G. G., Graves, S., Heiberger, R., Monette, G., Nilsson, H., ... & Suggests, M. A. S. S. (2007). The car package. *R Foundation for Statistical Computing*.

Friedman, M. (1937). The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association*, *32*(200), 675–701. https://doi.org/10.1080/01621459.1937.10503522

Frischen, A., Eastwood, J. D., & Smilek, D. (2008). Visual search for faces with emotional expressions. *Psychological Bulletin*, *134*(5), 662–676. https://doi.org/10.1037/0033-2909.134.5.662

Garvert, M. M., Friston, K. J., Dolan, R. J., & Garrido, M. I. (2014). Subcortical amygdala pathways enable rapid face processing. *NeuroImage*, *102*, 309–316. https://doi.org/10.1016/j.neuroimage.2014.07.047

Gold, J. M., Barker, J. D., Barr, S., Bittner, J. L., Bromfield, W. D., Chu, N., ... & Srinath, A. (2013). The efficiency of dynamic and static facial expression recognition. *Journal of vision*, *13*(5), 23-23.

Gu, X., Hoijtink, H., Mulder, J., & Rosseel, Y. (2019). Bain: A program for Bayesian testing of order constrained hypotheses in structural equation models. *Journal of Statistical Computation and Simulation*, *89*(8), 1526-1553.

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. https://doi.org/10.1016/j.visres.2016.12.019

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in cognitive sciences*, *4*(6), 223-233.

Hess, U., Adams, R. B., Grammer, K., & Kleck, R. E. (2009). Face gender and emotion expression: Are angry women more like men? *Journal of Vision*, *9*(12), 19–19. https://doi.org/10.1167/9.12.19

Hinojosa, J. A., Carretié, L., Valcárcel, M. A., Méndez-Bértolo, C., & Pozo, M. A. (2009). Electrophysiological differences in the processing of affective information in words and pictures. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(2), 173-189.

Hoijtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods*, *24*(5), 539

Honey, C., Kirchner, H., & VanRullen, R. (2008). Faces in the cloud: Fourier power spectrum biases ultrarapid face detection. *Journal of vision*, *8*(12), 9-9.

Horstmann, G., Lipp, O., Becker, S. (2012). Of toothy grins and angry snarls—Open mouth displays contribute to efficiency gains in search for emotional faces. *Journal of Vision*, 12(5):7. doi: https://doi.org/10.1167/12.5.7.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1), 215–243. https://doi.org/10.1113/jphysiol.1968.sp008455

Humphrey, K., Underwood, G., & Lambert, T. (2012). Salience of the lambs: A test of the saliency map hypothesis with pictures of emotive objects. *Journal of Vision*, *12*(1), 22–22. https://doi.org/10.1167/12.1.22

Hunt, A. R., Cooper, R. M., Hungr, C., & Kingstone, A. (2007). The effect of emotional faces on eye movements and attention. *Visual Cognition*, *15*(5), 513–531. https://doi.org/10.1080/13506280600843346

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research, 40, 1489–1506.

Jeffreys, H. (1998). The theory of probability. OUP Oxford.

Johnson, M. H. (2005). Subcortical face processing. *Nature Reviews Neuroscience*, *6*(10), 766–774. https://doi.org/10.1038/nrn1766

Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, *58*(6), 1233–1258. https://doi.org/10.1152/jn.1987.58.6.1233

Juth, P., Lundqvist, D., Karlsson, A., & Ohman, A. (2005). Looking for foes and friends: Perceptual and emotional factors when finding a face in the crowd. Emotion, 5(4), 379–395.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, *90* (430), 773-795.

Kätsyri, J. (2006). Human recognition of basic emotions from posed and animated dynamic facial expressions.

Kauffmann, L., Peyrin, C., Chauvin, A., Entzmann, L., Breuil, C., & Guyader, N. (2019). Face perception influences the programming of eye movements. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-018-36510-0

Kauffmann, L., Khazaz, S., Peyrin, C., & Guyader, N. (2021). Isolated face features are sufficient to elicit ultra-rapid and involuntary orienting responses toward faces. *Journal of Vision*,*21*(2), 4-4.

Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization. *ArXiv:1412.6980 [Cs]*. http://arxiv.org/abs/1412.6980

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, *46*(11), 1762–1776. https://doi.org/10.1016/j.visres.2005.10.002

Klink, P. C., Jentgens, P., & Lorteije, J. A. M. (2014). Priority Maps Explain the Roles of Value, Attention, and Salience in Goal-Oriented Behavior. *Journal of Neuroscience*, *34*(42), 13867–13869. https://doi.org/10.1523/JNEUROSCI.3249-14.2014

Kloth, N., Itier, R. J., & Schweinberger, S. R. (2013). Combined effects of inversion and feature removal on N170 responses elicited by faces and car fronts. *Brain and Cognition*, *81*(3), 321–328. https://doi.org/10.1016/j.bandc.2013.01.002

Kobayashi, H., & Kohshima, S. (1997). Unique morphology of the human eye. *Nature*, *387*(6635), 767-768.

Kulke, L. (2019). Neural mechanisms of overt attention shifts to emotional faces. *Neuroscience*, *418*, 59-68.

Lacruz, R. S., Stringer, C. B., Kimbel, W. H., Wood, B., Harvati, K., O'Higgins, P., ... & Arsuaga, J. L. (2019). The evolutionary history of the human face. *Nature ecology & evolution*, *3*(5), 726-736.

LeDoux, J. E. (2000). *Emotion Circuits in the Brain*. 31.

Leppänen, J. M., & Hietanen, J. K. (2004). Emotionally positive facial expressions are processed faster than negative facial expressions, but why? Psychological Research, 69, 2229.

Leopold, D. A., & Rhodes, G. (2010). A comparative view of face perception. *Journal of Comparative Psychology*, *124*(3), 233.

Lewis, M. B., & Edmonds, A. J. (2003). Face detection: Mapping human performance. *Perception*, *32*(8), 903-920.

Li, R., & Cottrell, G. (2012). A new angle on the EMPATH model: Spatial frequency orientation in recognition of facial expressions. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 34, No. 34).

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of the American Statistical Association*, *62*(318), 399–402. https://doi.org/10.1080/01621459.1967.10482916

Lipp, O. V., Price, S. M., & Tellegen, C. L. (2009). No effect of inversion on attentional and affective processing of facial expressions. *Emotion*, *9*(2), 248–259. https://doi.org/10.1037/a0014715

Lundqvist, D., Flykt, A., & Ohman, A. (1998). *The Karolinska directed emotional faces (KDEF)*.

Marat, S., Ho Phuoc, T., Granjon, L., Guyader, N., Pellerin, D., & Guérin-Dugué, A. (2009). Modelling Spatio-Temporal Saliency to Predict Gaze Direction for Short Videos. *International Journal of Computer Vision*, *82*(3), 231–243. https://doi.org/10.1007/s11263-009-0215-3

Martin, J. G., Davis, C. E., Riesenhuber, M., & Thorpe, S. J. (2018). Zapping 500 faces in less than 100 seconds: Evidence for extremely fast and sustained continuous visual search. *Scientific reports*, *8*(1), 1-12.

McCrum-Gardner, E. (2008). Which is the correct statistical test to use? *British Journal of Oral and Maxillofacial Surgery*, *46*(1), 38–41. https://doi.org/10.1016/j.bjoms.2007.09.002

McFadyen, J., Mermillod, M., Mattingley, J. B., Halász, V., & Garrido, M. I. (2017). A Rapid Subcortical Amygdala Route for Faces Irrespective of Spatial Frequency and Emotion. *The Journal of Neuroscience*, *37*(14), 3864–3874. https://doi.org/10.1523/JNEUROSCI.3525-16.2017

McLellan, T., Johnston, L., Dalrymple-Alford, J., & Porter, R. (2010). Sensitivity to genuine versus posed emotion specified in facial displays. *Cognition & Emotion*, *24*(8), 1277–1292. https://doi.org/10.1080/02699930903306181

Méndez-Bértolo, C., Moratti, S., Toledano, R., Lopez-Sosa, F., Martínez-Alvarez, R., Mah, Y. H., Vuilleumier, P., Gil-Nagel, A., & Strange, B. A. (2016). A fast pathway for fear in human amygdala. *Nature Neuroscience*, *19*(8), 1041–1049. https://doi.org/10.1038/nn.4324

Mermillod, M., Bonin, P., Mondillon, L., Alleysson, D., & Vermeulen, N. (2010). Coarse scales are sufficient for efficient categorization of emotional facial expressions: Evidence from neural computation. *Neurocomputing*, *73*(13–15), 2522–2531. https://doi.org/10.1016/j.neucom.2010.06.002

Mermillod, M., Bourrier, Y., David, E., Kauffmann, L., Chauvin, A., Guyader, N., Dutheil, F., & Peyrin, C. (2019). The importance of recurrent top-down synaptic connections for the anticipation of dynamic emotions. *Neural Networks*, *109*, 19–30. https://doi.org/10.1016/j.neunet.2018.09.007

Mermillod, M., Vermeulen, N., Lundqvist, D., & Niedenthal, P. M. (2009). Neural computation as a tool to differentiate perceptual from emotional processes: The case of anger superiority effect. *Cognition*, *110*(3), 346–357. https://doi.org/10.1016/j.cognition.2008.11.009

Morris, J. (1998). A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain*, *121*(1), 47–57. https://doi.org/10.1093/brain/121.1.47

Mulckhuyse, M. (2018). The influence of emotional stimuli on the oculomotor system: A review of the literature. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(3), 411–425. https://doi.org/10.3758/s13415-018-0590-8

Niu, Y., Todd, R. M., Anderson, A. (2012). Affective salience can reverse the effects of stimulus-driven salience on eye movements in complex scenes. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2012.00336

Öhman, A. (2005). The role of the amygdala in human fear: Automatic detection of threat. *Psychoneuroendocrinology*, *30*(10), 953–958. https://doi.org/10.1016/j.psyneuen.2005.03.019

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology*, *80*(3), 381–396. https://doi.org/10.1037/0022-3514.80.3.381

Parkhurst, D., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision*, *16*(2), 125–154. https://doi.org/10.1163/15685680360511645

Racine, J. S. (2012). RStudio: A platform-independent IDE for R and Sweave. *Journal of Applied Econometrics*.

Sassi, F., Campoy, G., Castillo, A., Inuggi, A., & Fuentes, L. J. (2014). Task Difficulty and Response Complexity Modulate Affective Priming by Emotional Facial Expressions. *Quarterly Journal of Experimental Psychology*, *67*(5), 861–871. https://doi.org/10.1080/17470218.2013.836233

Sato, W., Kochiyama, T., Yoshikawa, S., Naito, E., & Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. *Cognitive Brain Research*, *20*(1), 81-91.

Schubö, A., Gendolla, G. H. E., Meinecke, C., & Abele, A. E. (2006). Detecting emotional faces and features in a visual search paradigm: Are faces special? *Emotion*, *6*(2), 246–256. https://doi.org/10.1037/1528-3542.6.2.246

Schyns, P. G., Petro, L. S., & Smith, M. L. (2009). Transmission of Facial Expressions of Emotion Co-Evolved with Their Efficient Decoding in the Brain: Behavioral and Brain Evidence. *PLoS ONE*, *4*(5), e5625. https://doi.org/10.1371/journal.pone.0005625

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and Decoding Facial Expressions. *Psychological Science*, *16*(3), 184–189. https://doi.org/10.1111/j.0956-7976.2005.00801

Stuit, S. M., Kootstra, T. M., Terburg, D., van den Boomen, C., van der Smagt, M. J., Kenemans, J. L., & Van der Stigchel, S. (2021). The image features of emotional faces that predict the initial eye movement to a face. *Scientific Reports*, *11*(1), 1-14.

Sweeny, T. D., Suzuki, S., Grabowecky, M., & Paller, K. A. (2013). Detecting and categorizing fleeting emotions in faces. Emotion, 13(1), 76–91. https://doi.org/10.1037/a0029193

Tamietto, M., & de Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. *Nature Reviews Neuroscience*, *11*(10), 697–709. https://doi.org/10.1038/nrn2889

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, *7*(14), 4. https://doi.org/10.1167/7.14.4

Theeuwes, J. (2019). Goal-driven, stimulus-driven, and history-driven selection. *Current Opinion in Psychology*, *29*, 97–101. https://doi.org/10.1016/j.copsyc.2018.12.024

Tipples, J., Atkinson, A. P., & Young, A. W. (2002). The eyebrow frown: A salient social signal. *Emotion*, *2*(3), 288–296. https://doi.org/10.1037/1528-3542.2.3.288

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from

untrained research participants. *Psychiatry Research*, *168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Tseng, P. H., Carmi, R., Cameron, I. G. M., Munoz, D. P., & Itti, L. (2009). Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of Vision*, *9*(7), 4–4. https://doi.org/10.1167/9.7.4

Van der Gaag, C., Minderaa, R. B., & Keysers, C. (2007). The BOLD signal in the amygdala does not differentiate between dynamic facial expressions. *Social cognitive and affective neuroscience*, *2*(2), 93-103.

Van der Stigchel, S., & de Vries, J. P. (2015). There is no attentional global effect: Attentional shifts are independent of the saccade endpoint. *Journal of Vision*, *15*(15), 17. https://doi.org/10.1167/15.15.17

Van der Stigchel, S., & Nijboer, T. C. W. (2013). How global is the global effect? The spatial characteristics of saccade averaging. *Vision Research*, *84*, 6–15. https://doi.org/10.1016/j.visres.2013.03.006

Villepoux, A., Vermeulen, N., Niedenthal, P., & Mermillod, M. (2015). Evidence of fast and automatic gender bias in affective priming. *Journal of Cognitive Psychology*, *27*(3), 301-309

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature Neuroscience*, *6*(6), 624–631. https://doi.org/10.1038/nn1057

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review*, *14*(5), 779-804.

Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., & Kissler, J. (2017). Mapping the emotional face. How individual face parts contribute to successful emotion recognition. *PLOS ONE*, *12*(5), e0177239. https://doi.org/10.1371/journal.pone.0177239

Wilcoxon, F. (1992). Individual Comparisons by Ranking Methods. *Biometrics Bulletin, Vol. 1, No. 6. (Dec., 1945), Pp. 80-83.*

Wu, D. W. L., Bischof, W. F., & Kingstone, A. (2014). Natural gaze signaling in a social context. *Evolution and Human Behavior*, *35*(3), 211-218.