Many ways to see your feelings: Successful facial expression recognition occurs with diverse patterns of fixation distributions

Neta Yitzhak, Yoni Pertzov, Nitzan Guy and Hillel Aviezer

Hebrew University of Jerusalem

Date Submitted: 12/11/2019

Author Note

This work was supported by an Israel Science Foundation [ISF#259/18] grant to Hillel Aviezer and by a BSF grant (2013028) to Hillel Aviezer.

Word count for the text: 7831

Corresponding author: Neta Yitzhak

Department of Psychology

The Hebrew University of Jerusalem

Jerusalem, Israel

neta.gabsi@mail.huji.ac.il

Abstract

Facial expression recognition relies on the processing of diagnostic information from different facial regions. For example, successful recognition of anger vs. disgust requires one to process information located in the eye/brow region, or in the mouth/nose region, respectively. Yet, how this information is extracted from the face is less clear. One widespread view, supported by cross-cultural experiments as well as neuropsychological case studies, is that the distribution of gaze fixations on specific diagnostic regions plays a critical role in the extraction of affective information. According to this view, emotion recognition is strongly related to the distribution of fixations to diagnostic regions. Alternatively, facial expression recognition may not rely merely on the exact patterns of fixations, but rather on other factors such as the processing of extra-foveal information. In the present study, we examined this matter by characterizing and utilizing individual differences in fixation distributions during facial expression recognition. We revealed 4 groups of observers that differed in their distribution of fixations towards face regions in a robust and consistent manner. In line with previous studies, we found that different facial emotion categories evoked distinct distribution of fixations according to their diagnostic facial regions. However, individual distinctive patterns of fixations were not correlated with emotion recognition: individuals that strongly focused on the eyes, or on the mouth, achieved comparable emotion recognition accuracy. These findings suggest that extra-foveal processing may play a larger role in emotion recognition from faces than previously assumed. Consequently, successful emotion recognition can rise from diverse patterns of fixations.

Keywords: Emotion recognition; Facial expressions; Eye movements; Individual differences

Many ways to see your feelings: Successful facial expression recognition occurs with diverse patterns of fixation distributions

1. Introduction

Face perception is one of the most developed visual perceptual skills in humans (Haxby, Hoffman & Gobbini, 2000). Facial expressions, as one aspect of facial signaling, convey a wealth of emotional cues that are crucial for social communication. Despite their complexity, a brief glance at a facial expression allows viewers to efficiently (Smith, Cottrell, Gosselin & Schyns, 2005) and rapidly (Tracy & Robins, 2008) perceive and construct impressions of others' emotional states and use this information to engage in social interactions.

While the importance of this skill is undebated, it remains unclear what underlies cultural and individual differences in deciphering facial expressions. One common assumption is that the perception of facial expressions depends, among other factors, on visual attention to particular components of the face. For example, anger and disgust involve diagnostic activity in the upper and lower regions of the face, respectively. Recognizing stereotypical anger relies on decoding the downward drawing of the brows by the corrugator muscle (i.e., AU4[1]). This contrasts with disgust recognition which relies on decoding nose wrinkling (AU9) and upper lip raising (AU10) (Smith et al., 2005).

---

[1] Action units (AUs; Cohn, Ambadar & Ekman, 2007) are independent facial movements that provide an anatomically based coding scheme for measuring facial behavior (The Facial Action Coding System; Ekman & Friesen, 1976; Friesen & Ekman, 1992).

While it is clear that processing the information in diagnostic face regions is important, the specific patterns of visual scanning related to the extraction of this information remains unclear. In line with the assumption of diagnostic facial regions, researchers have examined the possible link between emotion recognition skills and the way people visually scan facial expressions. The precise nature of this link is the focus of the current investigation.

1.1. Gaze Patterns during Facial Identity Recognition

Early studies of eye movements during face recognition have consistently demonstrated that people tend to focus on the internal features of the face, with predominant focus on the eyes, but also on the nose and mouth (Janik, Wellens, Goldberg, & Dell'Osso, 1978; Yarbus, 1965). Recent studies have shown that the strategy employed to explore faces during recognition tasks is not universal, but differs across cultures. Specifically, Western Caucasian observers prominently fixate the eye region, and partially the mouth region, while East Asian observers consistently fixate the nose, or the central region of the face (Blais, Jack, Scheepers, Fiset & Caldara, 2008; Caldara, 2017; Miellet, He, Zhou, Lao & Caldara, 2012). Furthermore, even within culture, large and stable individual differences in gaze patterns of faces were illustrated, showing that individuals demonstrate distinctive performance profiles for visual scanning of faces during identity recognition tasks (Mehoudar, Arizpe, Baker, & Yovel, 2014; Peterson & Eckstein, 2013; Peterson, Lin, Zaun, & Kanwisher, 2016).

Interestingly, these cultural and individual differences in eye movements were not predictive of face identity recognition: observers from different cultures reached comparable performance at recognizing faces, even though they used different scanning patterns (Blais et al., 2008; Miellet et al., 2012). In addition, the wide variation among individuals within the same culture in fixation preferences did not correlate with performance on a face recognition task (Mehoudar et al., 2014; Peterson & Eckstein, 2013; Sekiguchi, 2011). Moreover, when comparing the scanning patterns during face recognition of individuals with Developmental Prosopagnosia (DP) who exhibit striking deficits in face recognition, to the scanning patterns of typical individuals, both groups demonstrate similar fixation locations (Peterson et al., 2019). Hence, at least in the case of DP, atypical fixation strategies cannot explain face processing deficits. However, observers' allowed to use their own idiosyncratic eye movement strategies maximized their task performance (Peterson & Eckstein, 2013). Thus, observers who's personal preference is to fixate the top half of the face perform better in face recognition tasks when forced to fixate on the top half of the face, rather than the bottom half (Peterson & Eckstein, 2013). These personal fixation strategies for face processing were also correlated with unique neural responses, as measured by EEG. The facial regions that were more fixated by specific individuals under free-viewing condition, elicited stronger face-sensitive neural responses in these individuals when fixation on these regions was enforced (Stacchi, Ramon, Lao & Caldara, 2019).

1.2. Gaze Patterns during Facial Expression Recognition

Both facial identity and expression recognition require visual analysis of faces. However, the recognition of facial expression differs in many aspects from identity recognition. As such, partially distinct cognitive processes and neural pathways are assumed to be involved in these tasks (Bruce & Young, 1986; Calder & Young, 2005). One of the major differences between these two processes is that while identity recognition requires the visual coding of state-invariant properties, expression recognition requires the analysis of changeable properties, such as eye gaze and muscular movements (Haxby et al., 2000). Therefore, different visual cues may be used for detecting each type of facial characteristics – identity or expression.

Furthermore, each expression category conveys unique signals of muscular activity in distinct face regions (Aviezer, Hassin, Perry, Dudarev & Bentin, 2012; Calder, Young, Keane & Dean, 2000; Ekman, 1993; Schyns, Petro, & Smith, 2007; Smith et al., 2005). For example, accurate identification of anger, fear, and sadness entails the detection of activity in the top half of the face, whereas the identification of happiness and disgust relies on detecting activity in the bottom half of the face (Calder et al., 2000). Therefore, according to the diagnostic regions view, the detection of facial activity in these diagnostic regions optimizes emotion recognition (Smith et al., 2005).

Studies have also demonstrated that observers' visual scanning patterns of emotional faces follow the distinct diagnostic facial regions that uniquely characterize different facial expressions. Specifically, observers direct their gaze towards information-rich regions, and therefore distinctive patterns of fixation distribution are induced by different facial expressions. For example, observers tend to look longer at the eye region in facial expressions of sadness, and sometimes also anger and fear, while conversely, they look longer at the mouth region for facial expressions of happiness, and sometimes also disgust (Beaudry, Roy-Charland, Perron, Cormier & Tapp, 2014; Eisenbarth & Alpers, 2011; Schurgin, Nelson, Iida, Ohira, Chiao & Franconeri, 2014; but see Blais, Fiset, Roy, Saumure Régimbald & Gosselin, 2017 for evidence indicating that different expressions evoke similar gaze patterns).

1.3. Fixation Distribution and Emotion Recognition Performance

Based on the aforementioned studies, a plausible assumption is that a gaze pattern that follows the optimal diagnostic sampling would predict optimal emotion recognition performance. By contrast, individuals may display poor recognition of specific emotions if their face scanning profile neglects diagnostic regions. Indeed, previous studies found correlations between fixation distribution and facial expression recognition.

For example, Wong, Cronin-Golomb & Neargarder (2005) have shown that older adults made a greater proportion of fixations to the lower half of the face, regardless of emotion, compared with younger adults, who made proportionally more fixations to the upper half of the face. This tendency to fixate more on the lower half of the face was correlated with poor identification of fear, anger and sadness. Consistent with this result, older adults were more accurate than younger adults at identifying disgusted faces. Since fear, anger and sadness require the processing of the upper half of the face (Calder et al., 2000), the authors suggested that older adults were at a distinct disadvantage when attempting to identify these facial expressions due to their bias towards lower parts of the faces.

Converging evidence for the correlation between fixation distribution and emotion recognition performance comes from neuropsychological case studies. Adolphs et al. (2005) have shown that patient SM who has bilateral amygdala damage, was selectively impaired in recognizing facial expressions of fear. This behavioral impairment was correlated with SM's reduced spontaneous fixations to the eye region. However, when SM was instructed to look at the eyes, her recognition of fearful faces was improved. The authors suggested that SM's impairment in recognizing fear stems from the fact that the recognition of fear relies on viewing the eyes, a region which SM systematically underexplored. By contrast, a similar pattern of facial exploration was demonstrated by patient PS, who had acquired prosopagnosia with bilateral occipitotemporal lesions, and directed most of her fixations towards the mouth region in tasks of facial expression recognition (Fiset et al., 2017). Similarly to SM, PS demonstrated a marked impairment in recognizing facial expressions, specifically with fear, sadness, anger and surprise (Richoz, Jack, Garrod, Schyns & Caldara, 2015). Unlike SM, patient PS remained impaired even when instructed or forced to look at the eye region, suggesting that fixation distribution could not solely explain her emotion recognition deficit (Fiset et al., 2017).

Cross-cultural evidence also hints to the link between fixation distribution and emotion recognition. Jack, Blais, Scheepers, Schyns and Caldara (2009) found differences between East Asian observers and Western Caucasian observers in the visual scanning patterns of facial expressions. East Asian observers systematically biased their fixations toward the eye region of the face, whereas Western Caucasian observers distributed their fixations more evenly across the face. These scanning patterns were correlated with performance: East Asians exhibited a significant deficit in categorizing facial expressions of fear and disgust, compared to Western Caucasian who categorized all facial expressions with comparably high accuracy.

The authors concluded that fixation distribution differences between observers from different cultures lead to performance differences in emotion recognition. Thus, by persistently fixating the eyes, Eastern observers presumably miss emotional diagnostic cues in disgusted and fearful expressions, and therefore their emotion recognition accuracy declines. Although fear recognition is assumed to rely on diagnostic cues in the eye region, the authors suggested that without also extracting visual information from the mouth, fear is highly confused with surprise. Indeed, most of the categorization errors were between these two emotion categories.

Together, these diverse findings reinforce the assumption that some patterns of fixation distribution are more efficient for emotion recognition tasks, while other patterns are inferior. However, this conclusion may be partially challenged by Vaidya et al., (2014) who examined how fixation patterns contribute to emotion recognition in typical participants, and compared the recognition of subtle and intense expressions. They found that the detection of extreme disgust, happiness and surprise (but not fear), and subtle happiness, appeared to be less dependent on fixation patterns. However, for subtle fear, disgust and surprise fixations to the eyes predicted better emotion recognition. On the one hand, these findings indicate that the detection of subtle emotions, which are challenging to detect, is dependent on the observer’s scanning pattern in a way that fixating the eyes may be the optimal scanning strategy for emotion recognition. On the other hand, for intense expressions, fixating on particular positions seems less critical for emotion recognition. These findings suggest a more complex link between visual exploration and emotion recognition.

1.4. Extra-foveal processing and Emotion Recognition Performance

The straightforward assumption that longer fixation time on a diagnostic region predicts enhanced emotion recognition may be weakened when taking into account the complex link between fixation position and visual information extraction and processing. Diagnostic facial features can be efficiently extracted by extra-foveal information processing, while fixation position reveals only the information that undergoes foveal processing. Miellet et al. (2012) showed that Easterners and Westerners reach equal performance in face recognition tasks, despite different fixation distribution: Westerners deployed fixations towards the eyes and the mouth, while Easterners deployed more central fixations. Using a parametric gaze-contingent technique, they demonstrated that Easterners relied more on extra-foveal information sampling during face processing than Westerners, and thus, both groups of observers practically extracted the same facial features for face recognition.

Therefore, observers with different fixation distribution may extract the same information from a facial expression if one is foveating diagnostic information and the other extracts the same features extra-foveally. Despite different gaze patterns, these two observers are expected to achieve similar emotion recognition since the information that is crucial for accurate recognition is processed in both cases.

1.5. The Present Study

As described above, while it is clear that processing diagnostic facial information is important for emotion recognition, the exact manner by which gaze patterns contribute to emotion recognition has not been resolved. In the present study, we took a novel approach to addressing this issue. First, we systematically characterized different profiles of individuals’ fixation distribution patterns, during emotion recognition tasks. Next, we examined if distinct patterns of fixation distributions over facial regions are correlated with emotion recognition. Importantly, because we used a relatively homogenous sample of typical observers, we avoided confounding factors that emerge when observers belong to specific populations (e.g., older adults, neuropsychological patients, or different cultures).

1.5.1. Individual differences in fixation distribution during emotion recognition. While individual differences in fixation distribution patterns have been documented in tasks of facial identity recognition, to the best of our knowledge, this question has not been previously addressed in the field of emotion recognition. Previous studies examining facial identity recognition demonstrated that the distribution of fixations is strongly determined by intrinsic and distinct characteristics of each individual viewer, beyond the effects of stimulus or task characteristics (Mehoudar et al., 2014; Peterson & Eckstein, 2013). These personal strategies of visual scanning were assumed to represent a behavioral trait of the observer. Accordingly, we hypothesized that variability in individuals’ scanning patterns will also emerge in an emotion recognition task.

Previous studies examining individual differences in scanning patterns of faces have focused on measuring the consistency of scanning patterns across individuals, compared with the consistency of scanning pattern within individuals throughout different testing sessions across time (Mehoudar et al., 2014; Peterson & Eckstein, 2013). After assessing individual differences, Peterson & Eckstein also defined two groups of observers –

observers whose gaze was directed to the upper region of the face (eye lookers) and observers whose gaze was directed to the low region of the face (nose lookers).

In the present study we first aimed to group observers into different scanning style profiles using a data driven method, rather than a-priori theoretical assumptions. Then, we aimed to assess the stability of these scanning style profiles across time. Moreover, we examined these individual differences over two different facial expression data sets, in order to generalize these individualized scanning profiles beyond a specific stimulus type. As previous studies demonstrated that subtle and intense facial expressions may induce different scanning strategies (Vaidya et al., 2014), we examined observers' eye movements while viewing both subtle (JeFEE: Yitzhak et al., 2017) and intense (ADFES: Van Der Schalk, Hawk, Fischer & Doosje, 2011) facial expressions. Moreover, we used dynamic, rather than static, facial expressions. It has been shown that dynamic social stimuli may elicit different scanning patterns compared to static social stimuli (Blais et al. 2017; Risko, Laidlaw, Freeth, Foulsham & Kingstone, 2012). Furthermore, as social interactions are dynamic in nature, static stimuli may lack many of their potentially important characteristics. This assumption is particularly relevant for facial expression recognition, which relies on changeable, dynamic cues.

1.5.2. Fixation distribution patterns for different emotions. In addition to characterizing fixation distribution profiles of different individuals, our second aim was to characterize fixation distribution patterns for different emotions, across individuals. Thus, we examined whether specific emotions alter observers' fixation distributions according to the assumed diagnostic facial regions of different emotion categories. We hypothesized that beyond individual differences in fixation distribution, the specific facial activity conveyed by different emotion categories will attract observers' fixations towards facial regions that contain the diagnostic muscular action. We followed Schurgin et al., 2014, who demonstrated that observers differentially attend to distinct regions of a face when judging different emotions. In line with their experimental design, we focused on five emotions – sadness, anger, fear, disgust and happiness. According to these previous findings, sadness, anger and fear were expected to attract observers' gaze to the eye region at the top half of the face ("upper" emotions), while disgust and happiness were expected to attract observers' gaze to the nose and mouth regions, at the bottom half of the face ("lower" emotions).

1.5.3. The correlation between fixation distribution and emotion recognition. Following the establishment of individualized fixation distribution profiles and facial expression "attracting regions", our next goal was to examine whether fixation distribution and emotion recognition are related within a relatively homogenous sample. According to the traditional view, gaze patterns that maximize fixations to specific emotionally informative facial features would lead to improved accuracy in emotion recognition. Specifically, observers focusing on the upper face should better recognize emotions with diagnostic regions in the eyes, eyebrows and forehead (i.e., anger, sadness and fear). On the other hand, observers focusing on the lower face should be better at recognizing emotions with diagnostic regions on the nose and mouth (i.e., disgust and happiness).

According to this hypothesis, fixation distribution would also predict emotion recognition at the stimulus level, beyond idiosyncratic scanning profiles. Thus, for trials in which the observed emotion is anger, sadness or fear, longer fixation time on the upper half of the face would be correlated with more accurate emotion recognition. For happiness and disgust, however, longer fixation time on the lower half of the face would be correlated with accurate emotion recognition. Alternatively, longer viewing on a diagnostic region may not predict enhanced emotion recognition. Accordingly, different observers would present similar emotion recognition performance, and longer viewing on the upper or lower half of the face would not predict emotion recognition of specific emotions. This pattern of results would support a weaker link between fixation distribution and emotion recognition then previously assumed, and suggest aspects of gaze behavior, other than direct fixations over time, that affect the processing of visual information.

2. Material and Methods

2.1. Participants

A group of 103 undergraduate students at the Hebrew University received partial course credit or payment in exchange for their participation. Sample size was determined based on previous studies (Jack et al., 2009; Mehoudar et al., 2014; Peterson & Eckstein, 2013) in which a sample size of 40 participants was deemed sufficient to detect medium to large effect sizes for the hypotheses under investigation in the present study. However, since we aimed to characterize different fixation distribution sub-profiles, we increased our sample size beyond that reported in previous studies.

Of these 103 participants that followed through with the experiment, 11 participants were excluded due to low proportion of valid gaze samples (<75%). After exclusion, the remaining participants (N = 92; 73% female; M_{age} = 23.2) had 87.5% valid gaze samples.

A sub group of participants (n = 14, 50% female; M_{age} = 25.3) was invited again to our lab to participate in a second phase of the study, in order to assess the stability of scanning patterns over time. None of these 14 participants met the exclusion criterion (valid gaze samples < 75%), having a group average of 88.5% valid gaze samples. The study was approved by the Hebrew University ethics committee.

2.2. Stimuli

Dynamic facial displays of the six basic emotions (anger, disgust, fear, happiness, sadness and surprise), and neutral, as portrayed by eight actors (4 female), were taken from two published stimulus sets (see Figure 1): The Amsterdam Dynamic Facial Expression Set (ADFES; van der Schalk et al., 2011) and the Jerusalem Facial Expressions of Emotion (JeFEE; Yitzhak et al., 2017). The ADFES stimulus set consists of 6 – 6.5 sec. video clips, in which models display a neutral face for .5 sec., and then express the emotion at apex for 5 – 6 sec. Emotion displays conveyed in a FACS-based prototypical manner, and were considerably intense in terms of facial muscle activity. The JeFEE stimulus set consists of 10 – 11 sec. video clips, in which actors display subtle facial expressions in a non-prototypical manner. Each stimulus begins with a neutral face that gradually changes into an emotional face. Unlike sets in which the actors were strictly instructed and trained to activate specific facial muscles to convey stereotypical configurations, in this set the actors had the liberty of using any facial muscles they wished (for full details of the development and norms of this set, see Yitzhak et al., 2017). Overall, the stimuli consisted of 112 video clips. The height of the characters' face was ~16 cm (∼15 degrees of visual angle, as participants sat at ~60 cm distance from the screen).

In order to test our hypotheses regarding the correlation between fixation distribution and emotion recognition we analyzed eye movements and emotion recognition for the intense ADFES dataset exclusively. This strategy was chosen because the facial expressions in the ADFES dataset are posed based on the FACS prototypical guidelines, and therefore the diagnostic regions of each facial expression are well defined. Emotion displays of the JeFEE dataset, however, are conveyed in a non-prototypical manner. As a result, it is less clear which facial regions are diagnostic for emotion recognition. Nonetheless, we analyzed eye movements for the JeFEE facial expressions in order to generalize the individual differences finding beyond a specific stimulus set. This diversity of stimuli is especially important for eye tracking studies (Risko et al., 2012).

Figure 1. Individual frame examples of (A) prototypical and intense ADFES emotion displays and (B) non-prototypical and subtle JeFEE emotion displays, over time. Both posers present facial expression of disgust.

2.3. Procedure

2.3.1. Design. The study had a 2 (set) X 3 (face region) X 7 (emotion) within-subjects design.

2.3.2. Emotion recognition. Participants were shown displays of anger, disgust, fear, sadness, happiness, surprise and neutral faces from the ADFES and the JeFEE facial expression sets. All displays were presented in random order. Participants viewed the displays on a 15.6" Dell PC and afterwards completed an emotion labelling task using a fixed-choice question, with all of the presented emotions as response options. Participants indicated which emotion label best described the expression. For each participant, we calculated the proportion of accurately recognized displays per emotion category in each of the sets.

2.3.3. Eye tracking. Gaze position data were collected using an SMI RED-m eye-tracking system (Teltow, Germany), a remote controlled infrared camera with an automatic eye and head tracker. The system collects binocular gaze and pupil data while allowing relatively free head movement within a virtual box of 32 cm X 21 cm at 60 cm distance from screen. A 5 point calibration was performed before each experimental session using RED-M software provided by SMI. Sampling rate was 120 Hz. Parcellation of the raw data to fixations and saccades was performed by the default procedure in BeGaze (Teltow, Germany). Three face regions (eyes, nose and mouth) that were the areas of interest (AOIs) in this study were manually defined for each dynamic stimulus, according to physical features of the faces (see Figure 2). These face regions were redefined and interpolated across time in order to follow the movements of the faces.

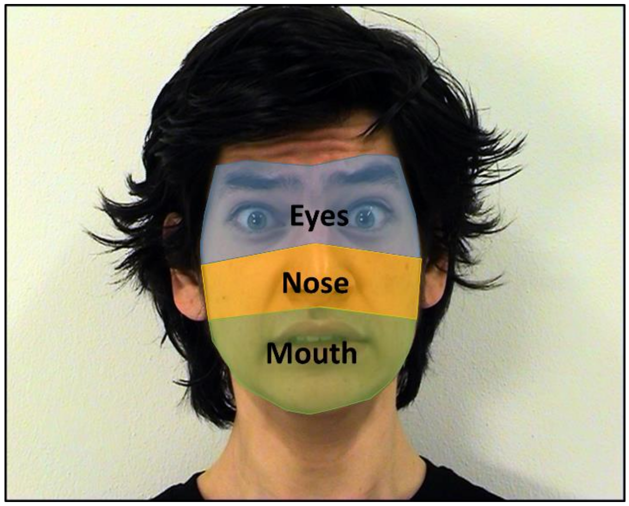

Figure 2. Face region shapes on a snapshot from one dynamic facial expression (ADFES). The face regions of interest included the eye region, nose region and mouth region. Face regions' shape was fixed relative to the face, but dynamic over time and changed according to the facial movements. The face region demarcations were only used for data analyses and were not visible to participants.

2.4. Scanning Patterns – Fixation Distribution

In order to define scanning patterns, we measured for each participant the percentage of time (Fixation time (%)[2]) at which gaze was directed at the three depicted facial regions (eyes, nose and mouth, see Figure 2). After averaging Fixation time (%) for these facial regions over trials from different emotion categories and different stimulus sets, a scanning pattern was defined as a vector of three components: Fixation time (%) within the mouth region, Fixation time (%) within the nose region, and Fixation time (%) within

[2] The eye movement measurement, related here as Fixation time (%) is the Net Dwell Time percentage (NDT%) measurement in the BeGaze software. The NDT% measures the proportion of time that gaze is directed to the region of interest out of the full trial duration.

the eye region. Note that this method of obtaining scanning patterns reflects the overall proportion of time that gaze was directed to the three depicted facial regions. Although "scanning pattern" is a general term that may also incorporate temporal information (i.e., the specific sequence of visually sampling the face regions over time) in this study we only refer to it as fixation distribution over the face.

2.5. Cluster Analysis

Each participant was represented by a vector of averaged Fixation time (%) on the three depicted facial regions. We applied *k*-means clustering[3] to divide the participants into distinct groups with similar patterns of fixation distribution. The *k*-means algorithm ran 20 times from different random starting assignments and the one with the lowest within cluster variation was selected. To ascertain the optimal number of clusters required to map different scanning styles we calculated the within-clusters variance for different *k* values. Based on this measure we identified the range of possible *k* values in which adding another cluster would not significantly improve the modeling of the data. We examined the clustering results for each potential *k* value, evaluated which *k* is most appropriate and established the number of $k = 4$ clusters (for further information on the cluster determination, see Appendix A in Supplemental materials).

3. Results

3.1. Individual Differences in Fixation Distribution during Emotion Recognition

[3] *k*-means clustering was conducted using R software: R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

To identify individualized profiles of fixation distribution we used the Fixation time (%) values over the 3 face regions (eyes, nose, mouth) for viewing facial expressions from the prototypical ADFES dataset (total of 56 exemplars per participant, *N* = 92). Cluster analysis revealed 4 different facial expression scanning styles: Mouth lookers (*n* = 8), Nose lookers (*n* = 5), Moderate eye lookers (*n* = 42) and Extreme eye lookers (*n* = 37). The Fixation time (%) distributions over face regions for different scanning style clusters are shown in Figure 3.

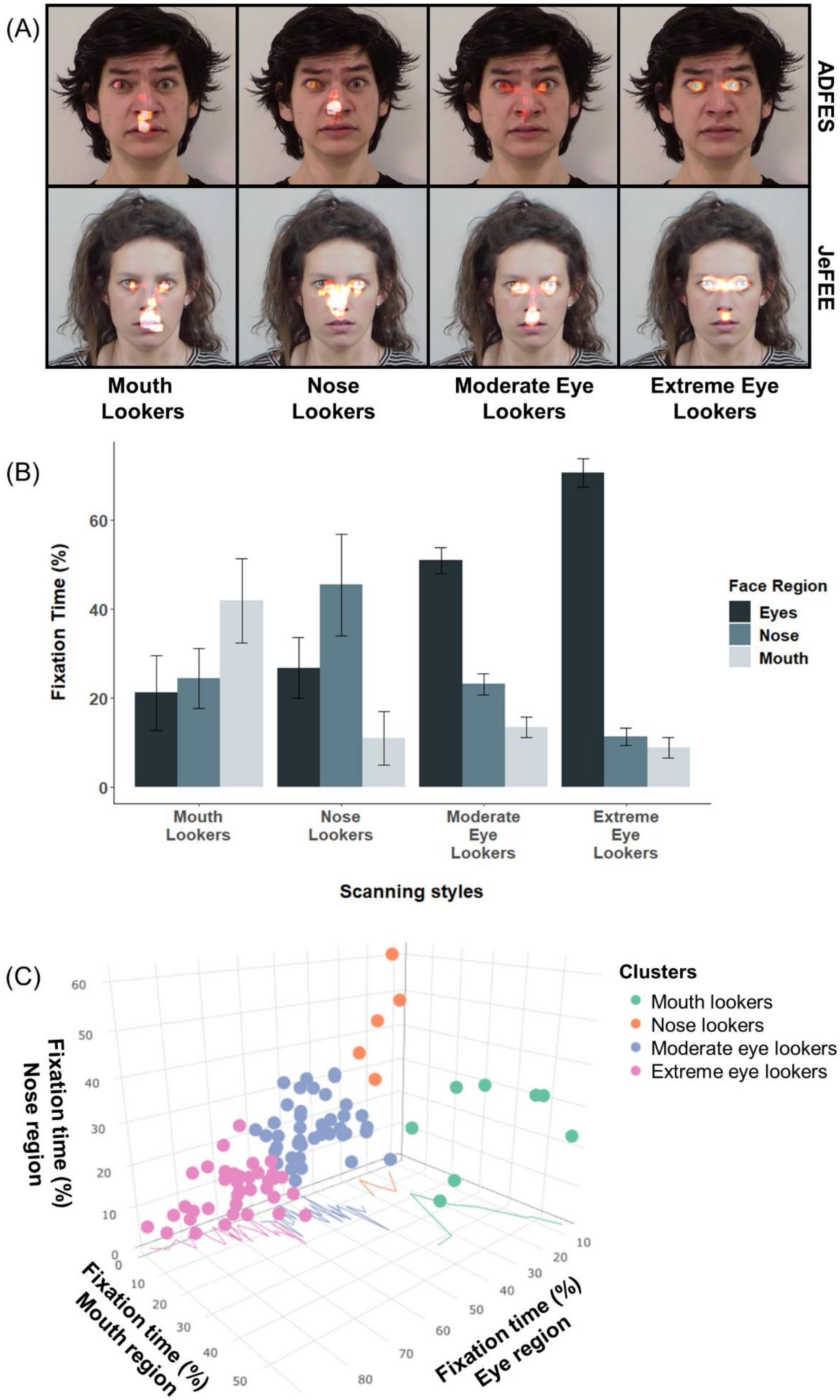
(A)
ADFES
JeFEE
Mouth Lookers
Nose Lookers
Moderate Eye Lookers
Extreme Eye Lookers
(B)
Fixation Time (%)
60
40
20
0
Face Region
Eyes
Nose
Mouth
Mouth Lookers
Nose Lookers
Moderate Eye Lookers
Extreme Eye Lookers
Scanning styles
(C)
Clusters
Mouth lookers
Nose lookers
Moderate eye lookers
Extreme eye lookers
Fixation time (%) Nose region
Fixation time (%) Mouth region
Fixation time (%) Eye region

Figure 3. *k*-means clustering revealed four groups of observers, differing in their facial expression scanning styles: Mouth lookers, Nose lookers, Moderate and Extreme eye lookers. (A) Heat maps of the four scanning style clusters, collapsed across all observers in the same cluster, presented on one exemplar of fear, for the two stimulus sets: ADFES and JeFEE; (B) Fixation time (%) distribution over face regions, averaged across all trials, for different scanning style clusters. Error bars represent 95% confidence intervals; (C) Each dot represents one subject's Fixation time (%) on the three face regions (different axes). The color of a dot represents the subject's cluster. The lines represent the projection of the dots on the x-y plane.

Our next goal was to examine the consistency of these individualized scanning profiles beyond the effects of stimulus characteristics. To this end, we calculated the internal consistency of the Fixation time (%) values of the 3 face regions (eyes, nose, mouth), for viewing facial expressions from ADFES and the JeFEE datasets. Cronbach's α for time spent viewing the eye region, when comparing the ADFES and JeFEE facial expressions, was .98; Cronbach's α for time spent viewing the nose region was .96; and Cronbach's α for time spent viewing the mouth region was .98. These values indicate excellent consistency of observers' fixation distribution over the face across different datasets.

Moreover, we conducted another cluster analysis for the JeFEE stimulus set, and compared it to the clustering that was obtained for the ADFES dataset. Seventy-five out of 92 participants (79%) were classified to the same scanning style category. Together, these results suggest that the scanning profiles that emerged in our sample are not specific to a single stimulus type, but rather they can be generalized across different stimuli. For the full cluster analysis of the JeFEE stimuli see Figure S2, Appendix B in Supplemental materials.

After demonstrating individualized scanning profiles among observers that are consistent across datasets, our next goal was to examine the consistency of these patterns over time. If these 4 scanning-style groups indeed capture observers' personal scanning traits, they are expected to be stable over time to a considerable degree (Asendorpf, 1992). To examine this, we invited participants (n = 14) to participate in a second examination session of the study. Participants were invited based on their scanning style profile, in an effort to represent all the 4 different profiles that were obtained in the first session of the study (out of the 14 participants that participated in this second session: $n_{\text{mouth lookers}} = 3$, $n_{\text{nose lookers}} = 3$, $n_{\text{moderate eye lookers}} = 2$, $n_{\text{extreme eye lookers}} = 6$). In the second session participants went through the same procedure as in the first session, and their scanning patterns were assessed in a similar manner. The time interval between the two experimental sessions ranged between 2 weeks and 12 months (M_{Interval} = 5.2 months, SD_{interval} = 4.66). For more details, see Appendix C in supplemental materials.

A similarity matrix for fixation distribution obtained in the two experimental sessions is presented in Figure 4. The values in the similarity matrix represent the overall Pearson correlations between scanning pattern vectors from two sessions for each participant (the diagonal values of the matrix), and between participants (non-diagonal values of the matrix). Therefore, the maximal similarity that can be obtained for two sessions with identical vectors is $r_p = 1$, representing perfect stability of scanning patterns over time. A minimal similarity of $r_p = -1$ can be obtained for two sessions with perfectly opposite vectors. Similarity of $r_p = 0$ indicates two sessions with uncorrelated scanning style vectors.

The correlation coefficient between scanning patterns within individuals for the two sessions over time was $r_p = .92$, $p < .001$, indicating strong consistency of scanning styles. Strong correlations are also shown within groups of scanning styles (mean correlation within mouth lookers: $r_p = .97$; mean correlation within nose lookers: $r_p = .92$; mean correlation within extreme eye lookers: $r_p = .71$; mean correlation within extreme eye lookers: $r_p = .94$). Between these groups, however, correlation is weak or even negative. See Figure 4 for more details.

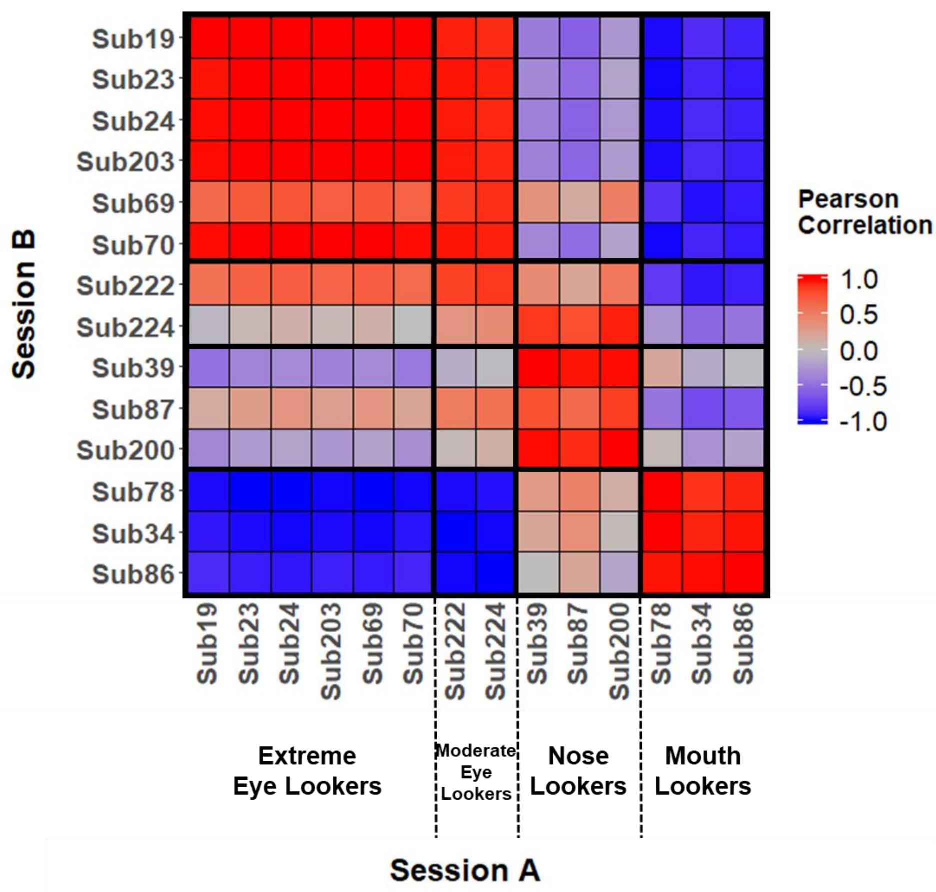

Figure 4. Matrix of correlations between fixation distribution vectors in two separate sessions, for 14 individuals with various scanning styles. Session 2 was a replication of session 1. The correlations between two sessions of the same participants are on the diagonal ($r_p = .92$), indicating strong consistency of scanning patterns over time. Strong correlations are also shown within groups of scanning styles, while between these groups correlation is weak, or even negative.

Next, we wanted to confirm that the fixation distribution profiles that were demonstrated were consistent along each trial. This was particularly important since observers' eye movements were recorded during relatively long viewing durations (6 seconds for the ADFES video clips), and reporting the perceived emotion was allowed only after each full video clip was presented. Therefore, a possible concern was that observer's internal scanning strategy for emotion recognition performance emerged at the beginning of each trial, and then it was masked by later eye movement data that were collected after emotion identification had been accomplished.

To rule out this alternative explanation we analyzed the eye movement data along trials to examine the consistency of fixation distribution from the beginning to the end of each trial. As can be seen in Figure 5, the findings suggest that fixation distribution over different face regions is considerably stable over trials, and the amount of time that participants from different clusters spend in different AOIs does not dramatically change over time within trials. The exception to this consistency is the first 1-2 fixations of each trial: at the first 200 ms of stimulus presentation, most of the fixations are directed to the eye region, regardless of scanning style. This deviation is expected, since the stimuli follow a fixation cross that appears at the center of the screen. However, after 500 ms (the time in which the neutral face changes into an emotional expression), observers with different scanning style demonstrate distinct patterns of facial exploration. From this time point and until the end of the stimulus, these systematic differences across observers are consistent along the trial. In Figure 6 we sampled 3 time intervals (first fixations: 0 – 200ms; first second of facial expressions: 500 – 1500 ms; relatively late time interval:

3500 – 4500 ms) to demonstrate this course. Therefore, we can conclude that the use of dynamic stimuli with long duration did not bias observers' scanning patterns, which remained considerably stable over the trials.

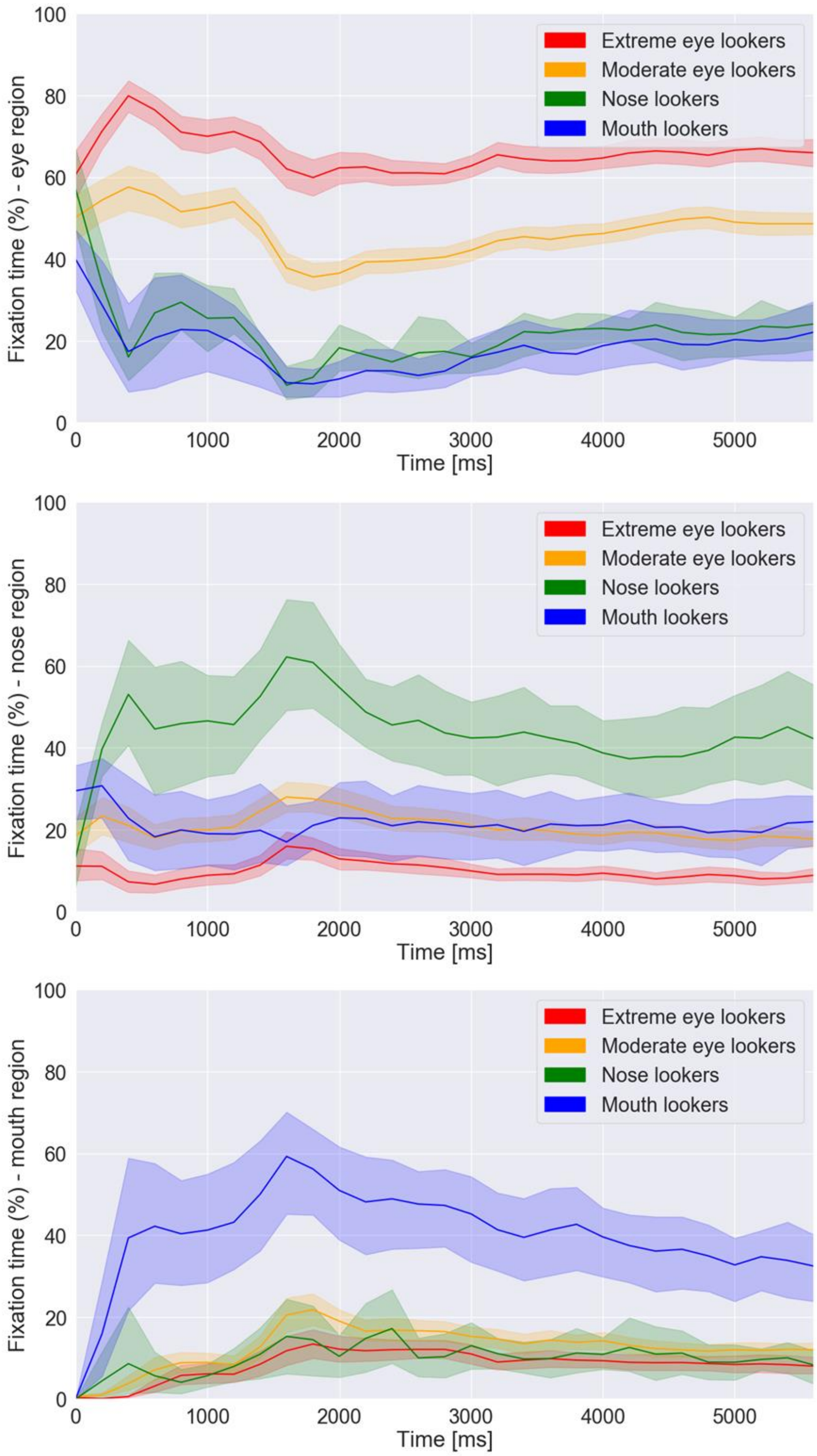

Figure 5. Time course of fixation time (%) on different face regions, averaged across all participants from each of the scanning style groups. X axis represents the trial duration, and each bin is 200 ms out of the full trial period (6 seconds). Y axis is the proportion of

fixation time, and different colors represent different scanning profiles. As can be seen in the charts, within each group of scanning style, the scanning patterns are relatively consistent over the trial, and no qualitative change is apparent following the first time bin. The distribution of the mean value at each time point is estimated based on bootstrapping. This distribution represents 95% confidence intervals.

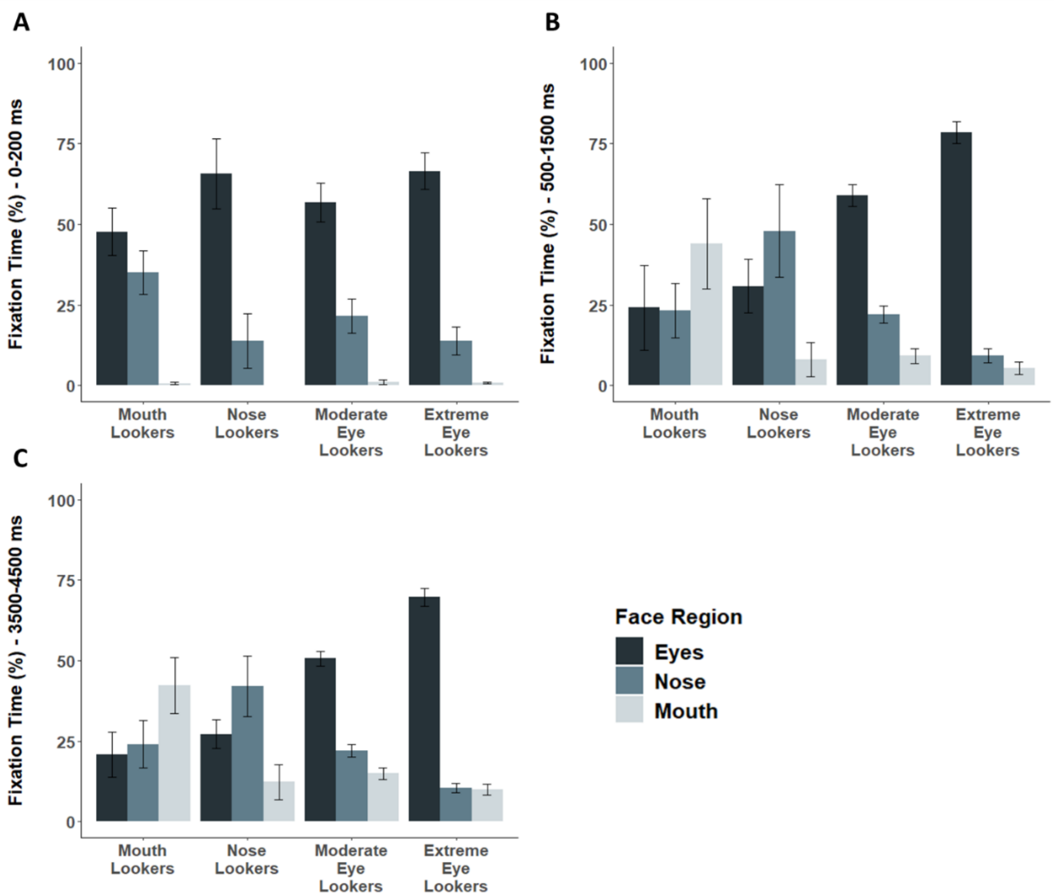

Figure 6. Fixation time (%) of observers with different scanning styles, in samples of 3 time intervals: (A) 0 – 200 ms (B) 500 – 1500 ms (C) 3500 – 4500 ms.

3.2. Fixation Distribution for Different Facial Expressions

Our next analysis aimed to examine how specific emotions determine gaze location in different expressions, beyond individuals. Based on the assumption of specific diagnostic regions for specific emotions, we hypothesized that different emotions selectively attract observers' overt visual attention. Following previous studies (CITATIONs), we defined the "top" half of the face as the eye region, and the "bottom" half of the face as the nose and mouth regions combined. As previously noted, we focused this analysis on the intense ADFES dataset because its facial expressions are based on the FACS prototypical guidelines, and therefore the diagnostic regions of each facial expression are well defined.

A 2-way within-subjects ANOVA was conducted, with face region (top, bottom) and emotion (anger, sadness, fear, disgust and happiness) as factors. The dependent variable was fixation time (%). The main effects for both face region, $F(1,91) = 30.53$, $p < .001$, $\eta_p^2 = .25$, and emotion, $F(4,364) = 9.05$, $p < .001$, $\eta_p^2 = .09$, were significant. The interaction between face region and emotion was also significant, $F(4,364) = 71.34$, $p < .001$, $\eta_p^2 = .44$, indicating that the distribution of fixation time (%) to the top and the bottom halves of the face differed across emotions. See Figure 7 (A).

To examine the driving force of the interaction we further examined this differential gaze allocation, and asked whether the difference between eye viewing and nose + mouth viewing is greater for upper emotions (anger, fear and sadness), compared to lower emotions (disgust and happiness). To this end we conducted a paired t test, comparing the difference between eye viewing and nose + mouth viewing, averaged for the two emotion groups – upper (anger, sadness and fear) and lower (disgust and happiness). This comparison yielded a significant effect, $t(91) = 11.64$, $p < .001$, Cohen's $d = 1.21$, reinforcing the hypothesis that upper emotions are correlated with greater gaze allocation to the upper half of the face, and vice versa for lower emotions.

After demonstrating different fixation patterns for different facial expressions for the dynamic stimuli as a whole, we next aimed to examine the course of these differences over time. To this end, we sequenced our data into 500-1000 ms bins. For each of the ADFES stimuli, the face was neutral during the first 500 ms, and after that a full-blown facial expression was conveyed until the end of the video. The sequencing of the data revealed an interesting course of gaze allocation to different face regions: at the first second of the facial expression (500 ms – 1500 ms), fixation time (%) for the top or the bottom regions of the face did not substantially differ as a function of the expression presented, and was similar to the baseline of looking at a neutral expression (0 – 500 ms): most of the fixations were directed to the eye region, and a small portion of fixations was directed to the nose and mouth regions. However, at the next second (1500 ms – 2500 ms) this pattern dramatically changed, and different emotions differed in their fixation distribution: for the upper emotions, and especially for anger, observers spent more time fixating on the eyes region, compared to the nose and mouth regions. In contrast, when “lower” emotions were expressed, most of the fixations were directed to the nose and mouth regions. following this stage, this pattern was gradually moderated and fixation distribution for different emotions became more similar to the baseline. Yet, even at the end of the video, “upper” emotions yielded, on average, more fixations to the eye region compared to “lower” emotions. See Figure 7 (B) for more details.

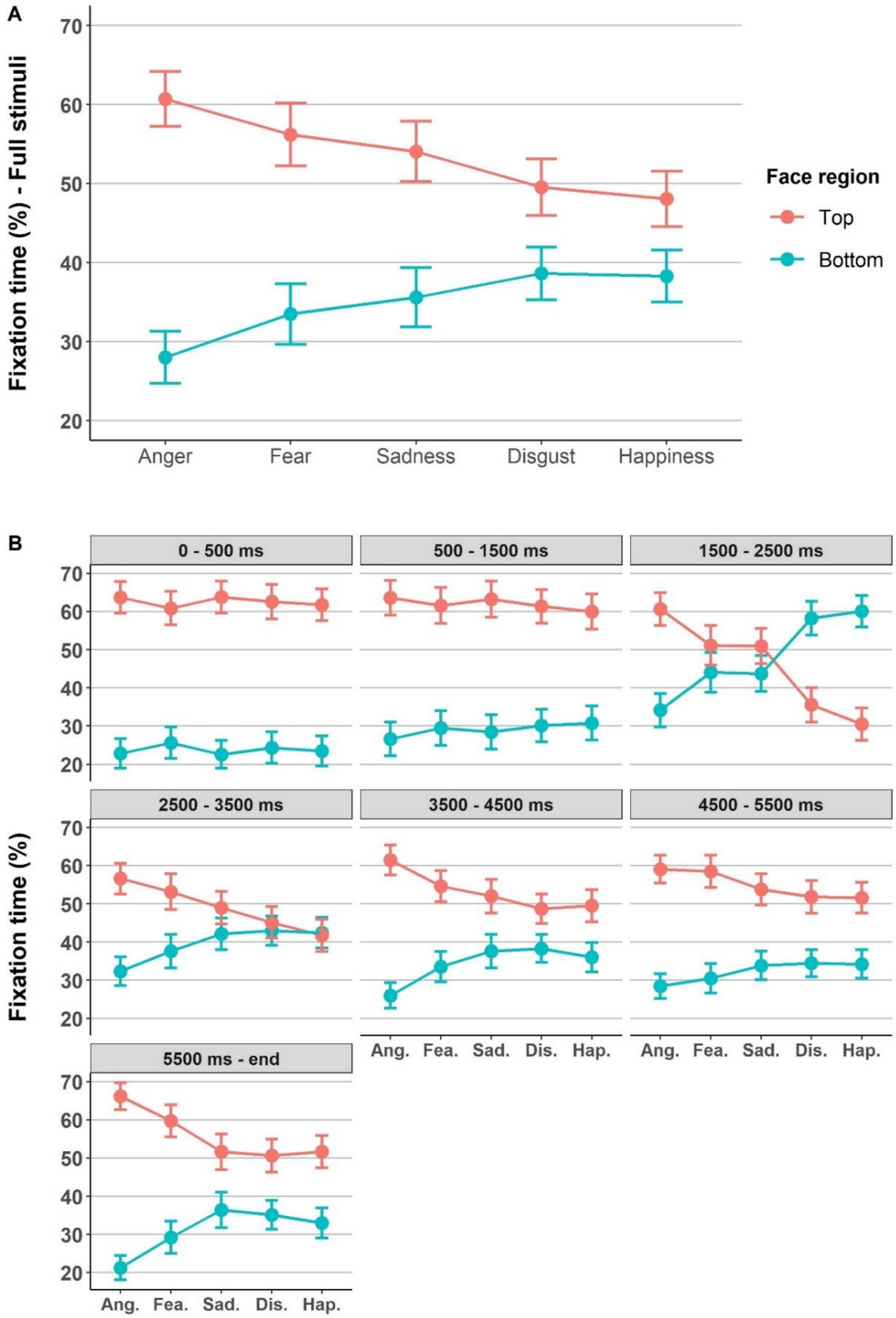

Figure 7. (A) Fixation time (%) for the top half of the face (eye region) and the bottom half of the face (nose + mouth regions), calculated for different emotions from the ADFES (intense stereotypical). Error bars indicate 95% confidence intervals. (B) Time course of Fixation time (%) for the top and bottom half of the face when different emotions were conveyed. The first 500 ms represent the baseline, as at this time facial expression is neutral. After 500 ms a full-blown facial expression in conveyed for the rest of the video. Error bars indicate 95% confidence intervals.

3.3. The Correlation Between Fixation Patterns and Emotion Recognition

Having established that different expressions modulate individuals' fixation distribution, we next turned to examine whether fixation scanning-styles are correlated with emotion recognition. To test this, we compared the recognition accuracy of the four fixation scanning style groups, for prototypical facial expressions, across the different emotions[4]. Emotion recognition performance data for different fixation scanning-styles and for different emotions are presented in Table 1. Figure 8 describes the the number of participants who attained certain emotion recognition accuracy (%) scores.

Table 1

Emotion recognition performance (% Accuracy) of participants with different scanning profiles across emotions – ADFES stimuli

	% Accuracy				
	Mouth lookers (n = 8) M (SD)	Nose lookers (n = 5) M (SD)	Moderate eye lookers (n = 42) M (SD)	Extreme eye lookers (n = 37) M (SD)	Total (N = 92) M (SD)
Anger	.95 (.09)	.90 (.10)	.96 (.11)	.96 (.08)	.96 (.10)
Disgust	.91 (.22)	.92 (.17)	.88 (.19)	.86 (.20)	.88 (.19)
Fear	.64 (.37)	.75 (.31)	.67 (.24)	.77 (.19)	.71 (.24)
Sadness	.94 (.09)	.87 (.18)	.88 (.15)	.89 (.11)	.89 (.13)

[4] Analyses of fixation patterns and emotion recognition for the JeFEE subtle expressions were also conducted, with similar results. See Appendix D in Supplementary information.

Happiness	1.00 (.00)	1.00 (.00)	.99 (.05)	.99 (.02)	.99 (.03)
Overall	.89 (.13)	.89 (.14)	.88 (.08)	.90 (.06)	.89 (.08)

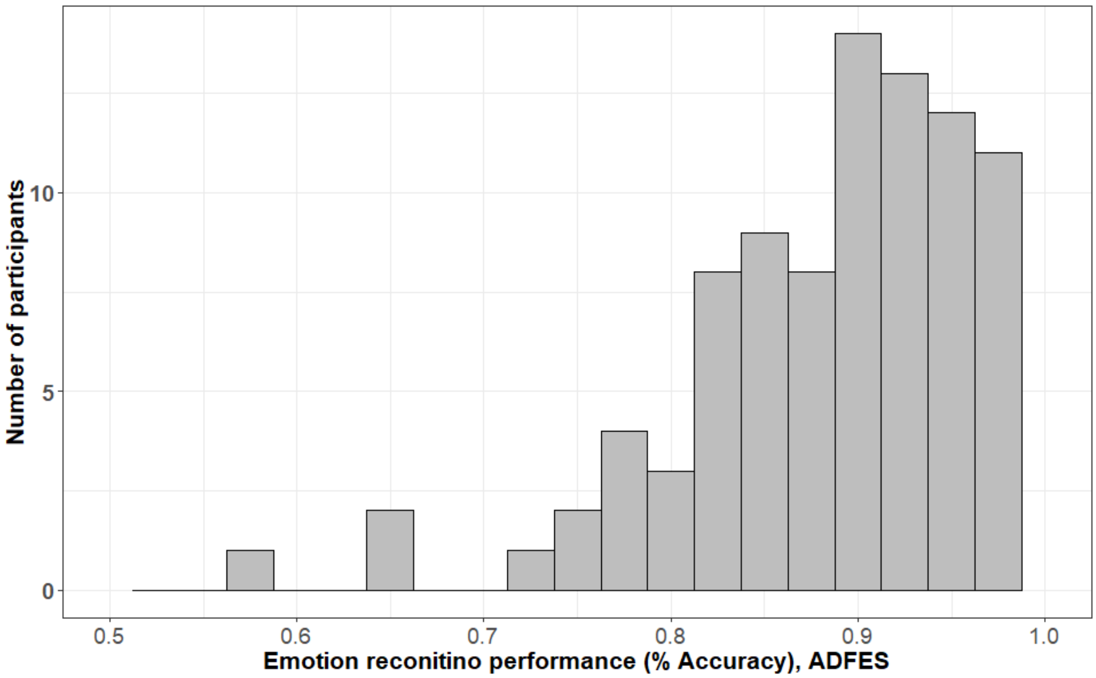

Figure 8. The distribution of emotion recognition accuracy (%) scores for the ADFES stimuli across participant. $N = 92$.

As can be seen in Figure 9 (A), fixation scanning style groups did not differ in their emotion recognition performance for different emotions. To confirm this pattern a 2-way mixed ANOVA with emotion (anger, sadness, fear, disgust and happiness) as within-subjects factor, and fixation scanning style (mouth lookers, nose lookers, moderate eye lookers, extreme eye lookers) as a between-subjects factor, was conducted. The main effect of fixation scanning style was not significant, $F(3,88) = .44$, $p = .72$, suggesting that different fixation profiles do not differ in emotion recognition performance. Therefore, none of the fixation scanning styles is superior over the others in terms of emotion recognition.

Our next question was whether observers with different fixation scanning styles have selective "expertise" for specific emotion categories, rather than overall emotion recognition performance. As described above, the diagnostic cues of the prototypical displays of anger, sadness and fear ("upper emotions") are located at the top half of the face, while the diagnostic cues of the prototypical displays of disgust and happiness ("lower emotions") are located at the bottom half of the face. Therefore, nose and mouth lookers who spend relatively more time fixating at the nose and mouth regions of the face may be expected to exhibit superior emotion recognition of the lower emotions, while moderate and extreme eye lookers who spend more time fixating at the eye region may be expected to exhibit superior recognition of upper emotions.

To test this hypothesis we analyzed the recognition of different emotions rather than the overall recognition performance. The emotion x fixation scanning style interaction was not significant, $F(12,352) = 1.06$, $p = .39$, indicating that different scanning styles do not differ in their emotion recognition of different emotions. Next we combined the upper and lower emotions into two distinct categories (see Figure 9 (B)), and conducted a new 2-way mixed ANOVA with emotion type (upper emotions, lower emotions) as within-subjects factor, and fixation scanning style (mouth lookers, nose lookers, moderate eye lookers, extreme eye lookers) as a between-subjects factor. The interaction between fixation scanning style and emotion type was not significant, $F(3,88) = .01$, $p = .43$. The lack of statistical differences in emotion recognition among observers with different fixation scanning styles, even when taking into account different emotion categories, also suggests that fixation scanning patterns and emotion recognition are not related.

Following the rejection of the alternative hypotheses with the ANOVA analyses, we aimed to quantify the evidence that the data provide in favor of the null hypothesis, i.e., that fixation scanning style is not correlated with emotion recognition performance. Therefore, we conducted Bayesian repeated measures ANOVA[5] with emotion (anger, sadness, fear, disgust and happiness) as within-subjects factor, and fixation scanning style (mouth lookers, nose lookers, moderate eye lookers, extreme eye lookers) as a between-subjects factor. In order to examine the main effect of fixation scanning style and the interaction of fixation scanning style with emotion (similarly to the ANOVA analyses), and given the a priori knowledge that accuracy rates differ across different emotions, we added the main effect of emotion to the null model. For the model including scanning style as a main effect, we found strong evidence for the data to have occurred under H_0, rather than under H_1, BF_{10} = .05, default prior P(M) = .33. When including both the main effect of scanning style and the interaction of scanning style X emotion, we found very strong evidence in favor of the null hypothesis, BF_{10} = .005, default prior P(M) = .33. These results strongly reinforce the null hypothesis, which states that fixation scanning pattern profiles are not related to emotion recognition performance for different emotions, and observers with different fixation scanning styles do not have selective "expertise" for specific emotions.

[5] All Bayesian analyses were performed using JASP software: JASP Team (2017). JASP (Version 0.8.4) [Computer software]. URL https://jasp-stats.org.

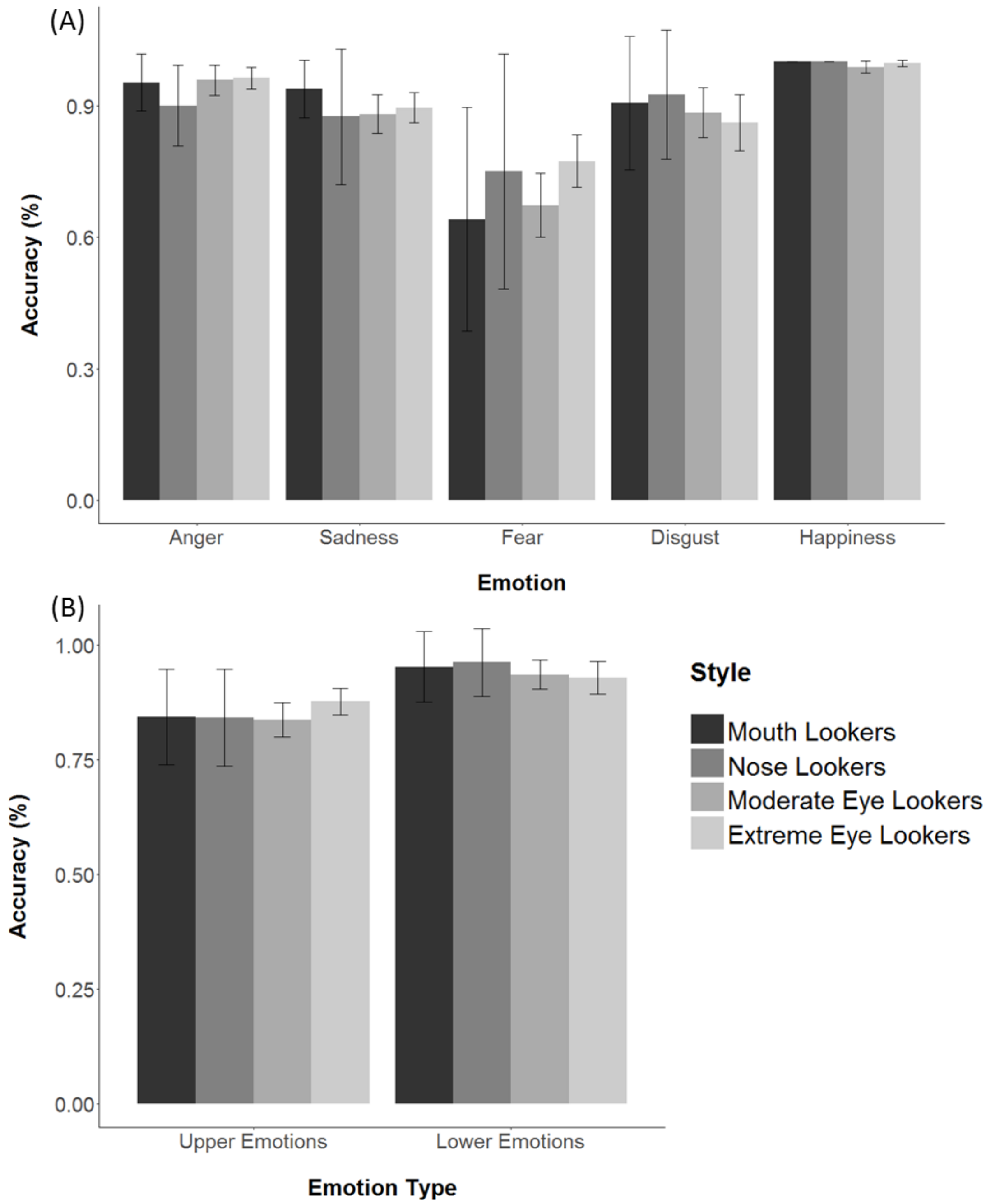

Figure 9. (A) Emotion recognition performance (Accuracy) of viewers with different scanning styles, across different emotions. (B) Emotion recognition performance of viewers with different scanning styles, across different emotion categories: upper emotions (anger, sadness and fear) and lower emotions (disgust and happiness). In both cases, scanning style groups were not significantly different in emotion recognition performance. Error bars indicate 95% confidence intervals.

After demonstrating that fixation scanning style does not predict emotion recognition of prototypical facial expressions at the group level, neither in general nor for specific emotions, we further examined the correlation between fixation distribution and emotion recognition performance at the trial level. To this end, we conducted logistic regressions with emotion recognition accuracy (1 = hit, 0 = miss) as the predicted variable. The fixation time (%) on the eye region was the predictive variable for the recognition of upper emotions (anger, sadness and fear), while the fixation time (%) on the nose + mouth regions was the predictive variable for the recognition of lower emotions (disgust and happiness).

For 4 out of 5 emotions (anger, sadness, fear and happiness), the regression results were consistent with our previous findings – longer viewing on diagnostic regions were not predictive for emotion recognition performance. The only exception was observed for facial expressions of disgust, which yielded that longer time spent on the mouth and nose regions, increased the probability of correct disgust labelling. Hence, focusing on the diagnostic region of disgust was predictive for emotion recognition of these expressions, at the trial level. This result is an exception among the other emotions that were examined and should therefore be interpreted with caution, see Table 2 for more details.

Table 2

L*ogistic regressions of fixation time (%) on the eye / nose + mouth regions on emotion recognition accuracy (1 = hit, 0 = miss) for different emotions*

Emotion	Predictor	Odds ratio	*z*	*P*
Anger	Fixation time (%) eyes	.99	- .77	.44

Sadness	Fixation time (%) eyes	.95	-.88	.38
Fear	Fixation time (%) eyes	1.00	1.01	.31
Disgust	Fixation time (%) nose + mouth	1.01	2.00	.046 *
Happiness	Fixation time (%) nose + mouth	1.01	.50	.61

* $p < .05$

4. Discussion

4.1. Summary of main findings

In the current study we established and characterized individual differences in fixation distribution during emotion recognition of facial expressions. We revealed 4 groups of observers that differ in their fixation scanning styles of facial expressions - mouth lookers, nose lookers, moderate eye lookers and extreme eye lookers. These scanning styles were highly stable over time, even up to 12 months. These findings add to the evidence of individual differences in the way people scan faces during identity recognition (Mehoudar et al., 2014; Peterson & Eckstein, 2013), and reinforce the notion that one's personal strategy for fixating on faces is an intrinsic individual trait, elicited across different tasks and stimuli. Thus, the observer, and not only the observed face, is a strong predictor for the fixation pattern.

We also demonstrated that specific emotions systematically alter observers' distribution of fixations over the face, consistently with the regions conveying diagnostic information. While this finding largely replicates previous findings using static facial stimuli (Beaudry et al., 2014; Eisenbarth & Alpers, 2011; Schurgin et al., 2014), only a few studies have tested gaze allocation to diagnostic regions of dynamic facial expressions which are arguably more ecological and representative of real-life expressions (Blais et al., 2017; Calvo, Fernández-Martín, Gutiérrez-García & Lundqvist, 2018). These two studies lead to apparently inconsistent results, in which the current study resolves. Intriguingly, the effect of different fixation patterns across emotions is not constant through time. When isolating the first second of the facial expression (500 ms – 1500 ms of stimulus presentation), there were no apparent difference in fixation distributions across emotions. This finding is consistent with the result reported by Blais et al., (2017), who used short stimulus duration (500 ms). However, at the next second (1500 ms – 2500 ms of stimulus presentation) a dramatic change in fixation distributions across emotions is evident. This finding is consistent with Calvo et al., 2018, who used longer stimulus duration, of 1,033 ms. This novel finding of changes over the course of dynamic facial expression stimuli may suggest an elucidation to the controversial findings in previous studies.

Having established individual differences in fixation distribution and differential fixation patterns to specific emotion categories, we continued to examine how these factors contribute to emotion recognition. As described in the introduction, a multitude of studies from diverse fields in psychology couple fixation distribution and emotion recognition such that some patterns of fixation distribution are superior for emotion recognition of specific emotions. Our results suggest that this link is far from straightforward.

Within a sample of typical participants, considerably distinctive patterns of fixations were related to comparable recognition accuracy of specific emotions. This finding suggests that extremely diverse fixation strategies can lead to comparable levels of performance. When examining the fixation patterns at the trial level, this was true for all emotions (anger, sadness, fear and happiness), with the exception of disgust. The notion that emotion recognition from facial expressions can be achieved with diverse gaze patterns echoes findings in the area of face identity perception indicating that differences in fixation preferences are not correlated with recognition performance (Caldara, 2017; Mehoudar et al., 2014; Peterson & Eckstein, 2013).

While it is clear that different stereotypical categories of facial emotion display distinct diagnostic regions our results suggest that diverse fixation patterns are equally efficient in the extraction of critical facial features. These findings strongly suggest that rather than direct fixations being the sole key factor, diagnostic facial features can be extracted by extra-foveal vision. Recording eye movements during free viewing cannot quantify the information that is sampled extra-foveally, and therefore fixation distribution is not a direct measure of all the visual information being extracted and processed. Hence, observers with different scanning styles may differ in the amount of information they extract from extra-foveal visual regions (Henderson & Ferreira, 2013; Peterson & Eckstein, 2013) but not in the overall visual signals that were extracted from the observed face.

Several explanations can be raised for the discrepancy between the overall lack of correlation between fixation patterns and emotion recognition demonstrated in the present study, and the significant correlation shown in previous studies. The first explanation may relate to experimental confounds in previous studies. For example, it has previously been shown that there are cultural differences in emotion recognition performance (Russell, 1994). It has also been demonstrated that observers from different cultures rely on different diagnostic cues when recognizing facial expressions. For example, Yuki, Maddux & Masuda (2007) showed that Japanese individuals focus more strongly on the eyes than the mouth when interpreting others' emotions, while American individuals focus relatively more on the mouth when interpreting facial expressions.

The correlation between these two factors may result from a common factor – one's culture. Examination of scanning patterns and emotion recognition in a cross cultural study confounds the independent variable of gaze pattern with the participants' culture. The present study, however, was conducted within a fairly homogenous and well-matched group of individuals, such that this confound was neutralized. In the same manner, confounds resulting from aging (Wong et al., 2005) or neurological disorders (Adolphs et al, 2005) may have obscured the interpretation of previous results.

Methodological differences across studies may have also played a role. In the present study we used dynamic facial expressions. By contrast, most of the previous studies of scanning patterns during emotion recognition used static images (Adolphs et al, 2005; Jack et al., 2009; Vaidya et al., 2014; Wong et al., 2005). These static stimuli may lack the richness of diagnostic cues that arise in dynamic stimuli, and therefore they may have influenced the correlation between extraction of visual information and task performance.

4.2. Limitations

Several limitations of the current study should be noted. First, observers' eye movements were recorded during long viewing durations (video clip duration of 6 seconds for the ADFES stimuli), and reporting the perceived emotion was allowed only after the full video clip was presented. However, humans identify facial expressions very rapidly, often within less than a second, and even a short glance at a diagnostic region may be enough to extract the essence of the facial activity and recognize the emotion (Tracy & Robins, 2008).

Therefore, it cannot be ruled out that considerable eye movement data that were analyzed in our study were collected after identification occurred. These later fixations may not be necessary for emotion recognition, and therefore may not represent observer’s internal scanning strategy for task performance but possibly idiosyncratic preference or other behaviors, such as social maintenance of eye contact (Peterson & Eckstein, 2013), with minimal influence on performance. Hence, in long presentation durations, diverse fixation patterns can lead to accurate emotion recognition if they share even one or two fixations to a diagnostic feature, which is sufficient for emotion recognition. Relying on the distribution of fixations that were recorded beyond this “critical glance” may mask a putative “pure” scanning pattern that is sufficient for this task.

As reported, we addressed this possible limitation by analyzing participants’ gaze allocation to different face regions along the trials (Figure 5). Our findings suggested that the use of dynamic stimuli with long duration did not bias observers' scanning preferences, which remained considerably stable over the trials. Yet, additional research is warranted. For example, future work may limit the duration of the stimuli to prevent redundant eye movements beyond the minimal fixations that are required for sufficient recognition. Such work may reinforce the conclusion that fixation distribution is not predictive of emotional expression discrimination, however, it is beyond the scope of the current investigation.

Another aspect of data collection that needs to be considered is that the calculation of scanning patterns in the present study does not take into account the specific sequence of visually sampling across facial regions over time. Rather, only the spatial aspect of scanning was taken into account, i.e. the “fixation distribution” – the proportion of time gaze was directed to each of the three facial regions of interest, across the full trial. The complex analysis of temporal sequence is beyond the scope of the current report, its role in emotion recognition is not well-established and we purposefully aimed to examine the more commonly tested and established spatial account. Nevertheless, it is possible that the detection of specific visual cues in specific time points during each trial may predict emotion recognition. Thus, our study calls into question the alleged link between fixation distribution and emotion recognition, but it does not address the more general link between scanning patterns (that include the temporal sequence of fixations) and emotion recognition.

Moreover, the division of our sample into 4 clusters of observers that differ in their fixation scanning styles, based on a data-driven cluster analysis approach, yielded unequal number of participants in each group of observer. Specifically, we found a strong bias for the preference of eye looking over nose and mouth looking (79 extreme + moderate eye lookers, vs 13 nose + mouth lookers). The main concern regarding the low frequency of nose and mouth lookers in our sample is that our follow-up statistical analyses, in which we divided our data to groups based on this cluster analysis, may be insensitive to find an effect in case that such an effect exists. We addressed this issue by supporting our ANOVA analyses with Bayesian statistics, which reinforced our interpretation that the data are sensitive enough, and provide strong evidence in favor of the null hypothesis. In addition, we conducted a logistic-regression analysis that complements the ANOVA analyses on the trial level, rather than on the group level, and ignores the clustering of the sample. Overall in our sample, for the ADFES stimuli, the proportion of direct fixation time (%) on the eyes region was 55%, and for the nose + mouth region was 34%. Thus, the strong bias in favor of the eye region that was obtained in the cluster analysis was moderated in this complementary regression analysis. Another concern may be that it could be challenging to replicate such clustering in a new sample since some of the rare clusters (nose and mouth lookers) may be missing. Future studies with large samples should try to replicate these results and to evaluate the frequency of visual preference patterns in the population.

4.3. Implications and Future Directions

The current novel findings regarding individual differences in fixation patterns on facial expressions, and the weak contribution of these fixation patterns to emotion recognition may have implications in diverse areas of research. For example, from the clinical perspective, studies frequently examine possible mechanisms that underlie emotion recognition deficits in neuropsychological patients, and fixation distribution is a common and prominent candidate for such a mechanism (e.g., Clark, Neargarder & Cronin-Golomb, 2010; Pelphrey, Sasson, Reznick, Paul, Goldman & Piven, 2002; Streit, Wolwer & Gaebel, 1997; van Asselen et al., 2012). The establishment of different scanning styles in typical observers, and the notion that fixation distribution is not a strong predictor for emotion recognition in healthy individuals, is crucial for the interpretation of findings in clinical patients.

Individual differences in fixation patterns can also spark new directions in future work involving emotion recognition tasks and eye movements. One of the most interesting questions arising from this study is, what is the connection between visual scanning patterns and emotion recognition performance? The findings of the present study suggest that emotion recognition from faces may be achieved with diverse patterns of fixation distribution over the face due to differential reliance of observers on extra-foveal information sampling. Thus, observers with different fixation distribution may extract the same diagnostic information that is critical for facial expression recognition. However, this hypothesis was not empirically tested in this study. Future studies that measure extra-foveal perception and examine it’s correlation with emotion recognition performance and it’s relation to fixation distributions can shed light on the nature of these individual differences in the level of visual information extraction rather than gaze patterns.

4.4. Conclusion

Observers differ in their scanning styles of facial expressions, in a robust and consistent manner. These individual differences can be clustered into distinct scanning style profiles. However, while previous work suggested that successful facial emotion recognition relies on “ideal” distribution of fixations over the face, the present study demonstrates that individuals who vastly differ in terms of fixation distribution profiles, accomplish comparably successful emotion recognition. These findings suggest that diverse patterns of facial scanning can be efficient for successful emotion recognition.

Acknowledgements

This work was supported by an Israel Science Foundation [ISF#259/18] grant to Hillel Aviezer and by a BSF grant (2013028) to Hillel Aviezer.

References

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, 433(7021), 68-72.

Asendorpf, J. B. (1992). Beyond stability: Predicting inter-individual differences in intra-individual change. *European Journal of Personality*, 6(2), 103-117.

Aviezer, H., Hassin, R. R., Perry, A., Dudarev, V., & Bentin, S. (2012). The right place at the right time: Priming facial expressions with emotional face components in developmental visual agnosia. *Neuropsychologia*, 50(5), 949-957.

Beaudry, O., Roy-Charland, A., Perron, M., Cormier, I., & Tapp, R. (2014). Featural processing in recognition of emotional facial expressions. Cognition & emotion, 28(3), 416-432.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

Blais, C., Fiset, D., Roy, C., Saumure Régimbald, C., & Gosselin, F. (2017). Eye fixation patterns for categorizing static and dynamic facial expressions. *Emotion*, *17*(7), 1107.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PloS one*, 3(8), e3022.

Blais, C., Roy, C., Fiset, D., Arguin, M., & Gosselin, F. (2012). The eyes are not the window to basic emotions. *Neuropsychologia*, *50*(12), 2830-2838.

Caldara, R. (2017). Culture reveals a flexible system for face processing. *Current Directions in Psychological Science*, *26*(3), 249-255.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, 6(8), 641-651.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human perception and performance*, 26(2), 527.

Calvo, M. G., Fernández-Martín, A., Gutiérrez-García, A., & Lundqvist, D. (2018). Selective eye fixations on diagnostic face regions of dynamic emotional expressions: KDEF-dyn database. *Scientific reports*, *8*(1), 17039.

Clark, U. S., Neargarder, S., & Cronin-Golomb, A. (2010). Visual exploration of emotional facial expressions in Parkinson's disease. *Neuropsychologia*, *48*(7), 1901-1913.

Cohn, J. F., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the Facial Action Coding System. In J. A. Coan & J. B. Allen (Eds.), *The handbook of emotion elicitation and assessment* (pp. 203–221). New York, NY: Oxford University Press.

Eisenbarth, H., & Alpers, G. W. (2011). Happy mouth and sad eyes: scanning emotional facial expressions. *Emotion*, *11*(4), 860.

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.

Fiset, D., Blais, C., Royer, J., Richoz, A. R., Dugas, G., & Caldara, R. (2017). Mapping the impairment in decoding static facial expressions of emotion in prosopagnosia. *Social cognitive and affective neuroscience*, *12*(8), 1334-1341.

Friesen, W. V., & Ekman, P. (1992). *Changes in FACS scoring* (pp. 1–19). San Francisco, CA: University of California, San Francisco.

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.

Henderson, J., & Ferreira, F. (2013). *The interface of language, vision, and action: Eye movements and the visual world*. Psychology Press.

Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology*, 19(18), 1543-1548.

Jack, R. E., Garrod, O. G., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current biology*, *24*(2), 187-192.

Janik, S. W., Wellens, A. R., Goldberg, M. L., & Dell'Osso, L. F. (1978). Eyes as the center of focus in the visual examination of human faces. *Perceptual and motor skills*, 47(3), 857-858.

Ketchen Jr, D. J., & Shook, C. L. (1996). The application of cluster analysis in strategic management research: an analysis and critique. *Strategic management journal*, 441-458.

Mehoudar, E., Arizpe, J., Baker, C. I., & Yovel, G. (2014). Faces in the eye of the beholder: Unique and stable eye scanning patterns of individual observers. *Journal of visio*n, 14(7), 6-6.

Miellet, S. R., He, L., Zhou, X., Lao, J., & Caldara, R. (2012). When East meets West: gaze-contingent Blindspots abolish cultural diversity in eye movements for faces. *Journal of Eye Movement Research*, 5 (2), 1-12.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of autism and developmental disorders*, *32*(4), 249-261.

Peterson, M. F., & Eckstein, M. P. (2013). Individual differences in eye movements during face identification reflect observer-specific optimal points of fixation. *Psychological science*, 24(7), 1216-1225.

Peterson, M. F., Lin, J., Zaun, I., & Kanwisher, N. (2016). Individual differences in face-looking behavior generalize from the lab to the world. *Journal of vision*, 16(7), 12-12.

Peterson, M. F., Zaun, I., Hoke, H., Jiahui, G., Duchaine, B., & Kanwisher, N. (2019). Eye movements and retinotopic tuning in developmental prosopagnosia. *Journal of vision*, *19*(9), 7-7.

Richoz, A. R., Jack, R. E., Garrod, O. G., Schyns, P. G., & Caldara, R. (2015). Reconstructing dynamic mental models of facial expressions in prosopagnosia reveals distinct representations for identity and expression. *Cortex*, *65*, 50-64.

Risko, E. F., Laidlaw, K. E., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with real versus reel stimuli: toward an empirical approach to concerns about ecological validity. *Frontiers in human neuroscience*, *6*.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological bulletin*, *115*(1), 102.

Schurgin, M. W., Nelson, J., Iida, S., Ohira, H., Chiao, J. Y., & Franconeri, S. L. (2014). Eye movements during emotion recognition in faces. *Journal of vision*, 14(13), 14-14.

Schyns, P. G., Petro, L. S., & Smith, M. L. (2007). Dynamics of visual information integration in the brain for categorizing facial expressions. *Current Biology*, 17(18), 1580-1585.

Sekiguchi, T. (2011). Individual differences in face memory and eye fixation patterns during face learning. *Acta psychologica*, 137(1), 1-9.

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological science*, 16(3), 184-189.

Stacchi, L., Ramon, M., Lao, J., & Caldara, R. (2019). Neural representations of faces are tuned to eye movements. *Journal of Neuroscience*, *39*(21), 4113-4123.

Streit, M., Wölwer, W., & Gaebel, W. (1997). Facial-affect recognition and visual scanning behaviour in the course of schizophrenia. *Schizophrenia research*, 24(3), 311-317.

Tracy, J. L., & Robins, R. W. (2008). The automaticity of emotion recognition. *Emotion*, *8*(1), 81.

Vaidya, A. R., Jin, C., & Fellows, L. K. (2014). Eye spy: The predictive value of fixation patterns in detecting subtle and extreme emotions from faces. *Cognition*, 133(2), 443-456.

van Asselen, M., Júlio, F., Januário, C., Campos, E. B., Almeida, I., Cavaco, S., & Castelo-Branco, M. (2012). Scanning patterns of faces do not explain impaired emotion recognition in Huntington disease: evidence for a high level mechanism. *Frontiers in psychology*, *3*.

Van Der Schalk, J., Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion*, *11*(4), 907.

Wong, B., Cronin-Golomb, A., & Neargarder, S. (2005). Patterns of visual scanning as predictors of emotion identification in normal aging. *Neuropsychology*, 19(6), 739.

Yarbus A. L. (1967). *Eye movements and vision*. New York: Springer.

Yitzhak, N., Giladi, N., Gurevich, T., Messinger, D. S., Prince, E. B., Martin, K., & Aviezer, H. (2017). Gently does it: Humans outperform a software classifier in recognizing subtle, nonstereotypical facial expressions. *Emotion*, *17*(8), 1187.

Yuki, M., Maddux, W. W., & Masuda, T. (2007). Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *Journal of Experimental Social Psychology*, 43(2), 303-311.