Common structure of saccades and microsaccades in visual perception

Zhenni Wang
Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China
Institute of Psychology and Behavioral Science, Shanghai Jiao Tong University, Shanghai, China

Radha Nila Meghanathan
Department of Experimental Psychology, Otto-von-Guericke University, Magdeburg, Germany

Stefan Pollmann
Department of Experimental Psychology, Otto-von-Guericke University, Magdeburg, Germany
Center for Behavioral Brain Sciences, Magdeburg, Germany

Lihui Wang
Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China
Institute of Psychology and Behavioral Science, Shanghai Jiao Tong University, Shanghai, China

We obtain large amounts of external information through our eyes, a process often considered analogous to picture mapping onto a camera lens. However, our eyes are never as still as a camera lens, with saccades occurring between fixations and microsaccades occurring within a fixation. Although saccades are agreed to be functional for information sampling in visual perception, it remains unknown if microsaccades have a similar function when eye movement is restricted. Here, we demonstrated that saccades and microsaccades share common spatiotemporal structures in viewing visual objects. Twenty-seven adults viewed faces and houses in free-viewing and fixation-controlled conditions. Both saccades and microsaccades showed distinctive spatiotemporal patterns between face and house viewing that could be discriminated by pattern classifications. The classifications based on saccades and microsaccades could also be mutually generalized. Importantly, individuals who showed more distinctive saccadic patterns between faces and houses also showed more distinctive microsaccadic patterns. Moreover, saccades and microsaccades showed a higher structure similarity for face viewing than house viewing and a common orienting preference for the eye region over the mouth region. These findings suggested a common oculomotor program that is used to optimize information sampling during visual object perception.

Introduction

When we look at an object, we carry out eye movements to explore its visual properties. Such eye movements tend to show distinctive paths for different visual categories, which are considered fundamental to visual perception and memory formation (Henderson, Williams, & Falk, 2005; Noton & Stark, 1971). A typical example is that the recognition of faces is often characterized by a T-shape gaze trace; that is, gaze fixations are concentrated in the eye–nose–mouth areas (Arizpe, Kravitz, Yovel, & Baker, 2012). In a recent study, it was shown that the viewing of face images and house images led to distinctive spatiotemporal structures of gaze traces that could be discriminated by multivariate pattern classification, suggesting that gaze traces are critical in the visual discrimination of the object categories (Wang, Baumgartner, Kaule, Hanke, & Pollmann, 2019). Moreover, face- and house-related gaze paths also contributed to distinctive activity patterns in the fusiform face area and the parahippocampal place area, brain areas known to represent face and house images, respectively (Wang et al., 2019).

Although face- and house-related eye movements appear to be a part of the neural representation of

Citation: Wang, Z., Meghanathan, R. N., Pollmann, S., & Wang, L. (2024). Common structure of saccades and microsaccades in visual perception. *Journal of Vision*, *24*(4):20, 1–13, https://doi.org/10.1167/jov.24.4.20.

https://doi.org/10.1167/jov.24.4.20 Received August 16, 2023; published April 24, 2024 ISSN 1534-7362

faces and houses, visual discrimination and distinctive neural activations have been observed even when the eye gaze was maintained at central fixation (Cohen & Tong, 2015; Kanwisher, McDermott, & Chun, 1997). These findings suggested that eye movements may not be necessary for the visual perception of object categories; however, it is well known that our eyes are never still, and microsaccades occur within gaze fixations (Martinez-Conde, Otero-Millan, & Macknik, 2013). It has been shown that the direction of microsaccades is biased to a cued location, suggesting a similar path of microsaccades to that of saccades, though on a smaller scale (Engbert & Kliegl, 2003; Hafed & Clark, 2002). The paths of microsaccades were interpreted as reflecting the shift of covert attention, pointing to a common oculomotor program in visual exploration (Engbert, Mergenthaler, Sinn, & Pikovsky, 2011; Otero-Millan, Macknik, Langston, & Martinez-Conde, 2013).

Although mounting evidence has shown similar paths between saccades and microsaccades during the orienting of spatial attention (Engbert & Kliegl, 2003; Hafed & Clark, 2002; Yuval-Greenberg, Merriam, & Heeger, 2014), it remains largely unknown if there is a common spatiotemporal structure of saccades and microsaccades in visual object perception. This is of significance because the common gaze traces may contribute to the object-related brain activities in the ventral pathway (Haxby, Gobbini, & Nastase, 2020; Liu, Rosenbaum, & Ryan, 2020), which have been observed in both free-viewing and fixation-controlled situations. In free-viewing contexts, it is well known that saccades are carried out to direct the center of gaze toward a specific region so that critical visual information (e.g., eyes/nose/mouth in face perception) can be gathered to achieve recognition (Henderson, 2003; Peterson & Eckstein, 2013; Yarbus, 1967). Similarly, it has been suggested that microsaccades within a fixation bring the visual information of interest close to the locus of fixation (Poletti, Listorti, & Rucci, 2013). Therefore, to recognize the same object or object category, microsaccades in the fixation-controlled situation may have a trace similar to saccades in the free-viewing situation.

Despite the well-documented function of free saccades in information sampling, there is no general agreement concerning the function of microsaccades in visual perception (Poletti & Rucci, 2016; Rolfs, 2009). One suggestion is that microsaccades were generated to ensure visibility and prevent the fading of visual perception (Ditchburn, Fender, & Mayne, 1959; Kelly, 1979; McCamy et al., 2012). Another suggestion is that both saccades and microsaccades reflect an oculomotor sampling strategy by which the visual system can optimize information acquisition (Cunitz & Steinman, 1969; Martinez-Conde, Macknik, Troncoso, & Hubel, 2009; Ko, Poletti, & Rucci, 2010). Although the two points of view are not mutually exclusive, it remains unresolved whether there is a division of function between saccades and microsaccades and whether microsaccades bear cognitive significance similar to that of saccades. Specifically, from the former perspective, microsaccades play a greater biological role in visual perception whereas saccades a more cognitive role. By contrast, the latter suggests an integrated cognitive role of saccades and microsaccades in acquiring visual information, both derived from a continuum of eye movement. By introducing the dichotomy of biological role versus cognitive role, we mean the function "to be able to see" versus the function "to select information to know." For example, ensuring visibility and preventing fading are fundamental for the visual system and are not treated as a visual strategy. By contrast, the acquisition and selection of visual information can be guided by the task goal and treated as a strategy. The two accounts can be assessed by investigating if microsaccades and saccades have a structural similarity beyond the biophysical similarity in visual object perception. The presence of such structural similarity would support the view of an integrated cognitive function of saccades and microsaccades.

Here, we investigated if microsaccades during controlled central fixation and normal-scale saccades during free eye movements had common spatiotemporal patterns during the viewing of faces and houses. We answered this question by testing the following predictions. First, we conducted a multivariate pattern classification on the saccadic/microsaccadic features. We predicted that both the saccades made during free eye movements and the microsaccades made during controlled fixation would show distinctive patterns for face versus house viewing. Importantly, a classifier trained with face- and house-related saccades during free eye movements could be generalized to the classification of face- and house-related microsaccades during controlled fixation, and vice versa, implying a common structure of saccade and microsaccade traces. Second, we calculated the representational distance between saccades during free eye movements and microsaccades during controlled fixation to quantify structural dissimilarity. Based on the well-established consistency of gaze patterns for face viewing (Mehoudar, Arizpe, Baker, & Yovel, 2014; Peterson & Eckstein, 2012), we predicted that the saccade–microsaccade representational distance would be lower for face viewing than for house viewing. Third, we quantified the preference for one region over another (e.g., eyes vs. mouth) during face processing and predicted that there would be consistent region preferences between saccades and microsaccades.

Methods

Participants

The sample size was estimated with G*Power 3.0 (Faul, Erdfelder, Lang, & Buchner, 2007) based on the statistics in a recent study that reported the similarity of gaze landing probabilities between foveal and parafoveal scales (Shelchkova, Tang, & Poletti, 2019). Given a correlation coefficient of $r = 0.77$, $\alpha = 0.0005$ reported in this study, and an expected power of 95%, 27 participants were required. Following this criterion, 27 university students (16 females; 18–30 years old; mean ± *SD*, 23.3 ± 6.2 years) participated in the present experiment. All participants reported normal or corrected-to-normal vision, except one participant who reported amblyopia of the right eye. We thus recorded the eye-movement data of his left eye. Written informed consent was obtained from each participant prior to the experiment. The study was conducted in accordance with the tenets of the Declaration of Helsinki and was approved by the local ethics committee. Data and analysis codes have been deposited at OSF (accession code: osf.io/j97uf).

Apparatus and tools

Each participant was seated in a dimly lit and sound-attenuated room at a viewing distance of 75 cm, with their head positioned on a chin-rest. The experiment was programmed with Psychtoolbox and run on MATLAB (MathWorks, Natick, MA). Two screen monitors had to be used due to technical issues. The screen monitor for the first seven participants had a resolution of 1680 pixels (width) × 1050 pixels (height) and a size of 60 × 40 cm. The screen monitor for the last 20 participants had a resolution of 2100 × 1313 pixels and a size of 53 × 30 cm. The sizes of the stimuli were kept the same on the two screens, and the pixel coordinates of the two screens were coregistered based on the stimuli size. For each participant, monocular eye movements (26 right eyes, one left eye) were recorded using the EyeLink 1000 Plus system (SR Research, Ottawa, ON, Canada) at a sampling rate of 1000 Hz. A standard procedure of nine-point calibration and validation was performed at the beginning of the experiment, with a maximum error of 1.0° as the threshold.

Design and procedure

Stimuli were 80 achromatic pictures of faces (20 males, 20 females, all with neutral emotion) and houses ($n = 40$). The size of faces was kept constant at a width of 14.4 degrees of visual angle (dva) × height of 16.7 dva (22 cm × 19 cm on the screen), with an eye-to-mouth distance of 7°. Due to the varying structures, the size of houses was not constant, with a mean width of 19.3° ± 1.0° (25.5 cm ± 1.3 cm on the screen), and a mean height of 13.1° ± 2.4° (17.5 cm ± 3.3 cm on the screen). A green dot (0.27 dva in diameter; RGB = 0, 256, 0) was presented as the central fixation.

Each participant completed two sessions: a central-fixation session and a free-viewing session. In each session, participants were asked to view pictures in a recognition task. In the central-fixation session, participants were required to maintain their eyes on the central fixation during the presentation of the pictures. In the free-viewing session, participants were instructed to look at the pictures with free eye movements. The order of the two sessions was counterbalanced across participants.

At the beginning of each session, six face pictures (three male and three female) and six house pictures were randomly selected from the pool of the 80 pictures. In the first block of each session, these 12 pictures were presented in a random order, with one picture in a specific trial. Participants were asked to memorize the 12 pictures. In each of the following 13 blocks, one to three new pictures were randomly selected from the picture pool and were added to the 12 old pictures. Participants were required to detect the new pictures amidst the old pictures.

Each trial began with a central fixation of a green dot which remained at the center of the screen for 0.6 to 1.0 seconds. Then, a picture was presented that lasted for 1.8 seconds. In the first block, participants only had to memorize the pictures. In each trial in the following 13 blocks, participants were required to press the "J" button on a standard keyboard using the right index finger if the current picture was not seen in the first block. In the central-fixation session, together with the picture, a green dot was presented at the center of the screen. To minimize eye movements, participants were required to maintain their eyes on the green dot until the offset of the picture and the green dot. In the free-viewing session, there was no green dot on top of the picture, and participants were free to move their eyes. After the presentation of the picture, a blank background was presented for 0.8 seconds. Drift checks were applied prior to each trial in the central-fixation session and were applied at the beginning of each block in the free-viewing session. Specifically, a drift check was manually performed before the presentation of the central dot in every single trial. During the drift check, a black dot was presented at the center of the screen. Each trial started and hence the green central dot was presented to replace the black dot only if the participant's gaze position was less than 1° from the black dot for the drift check. At the end of each block,

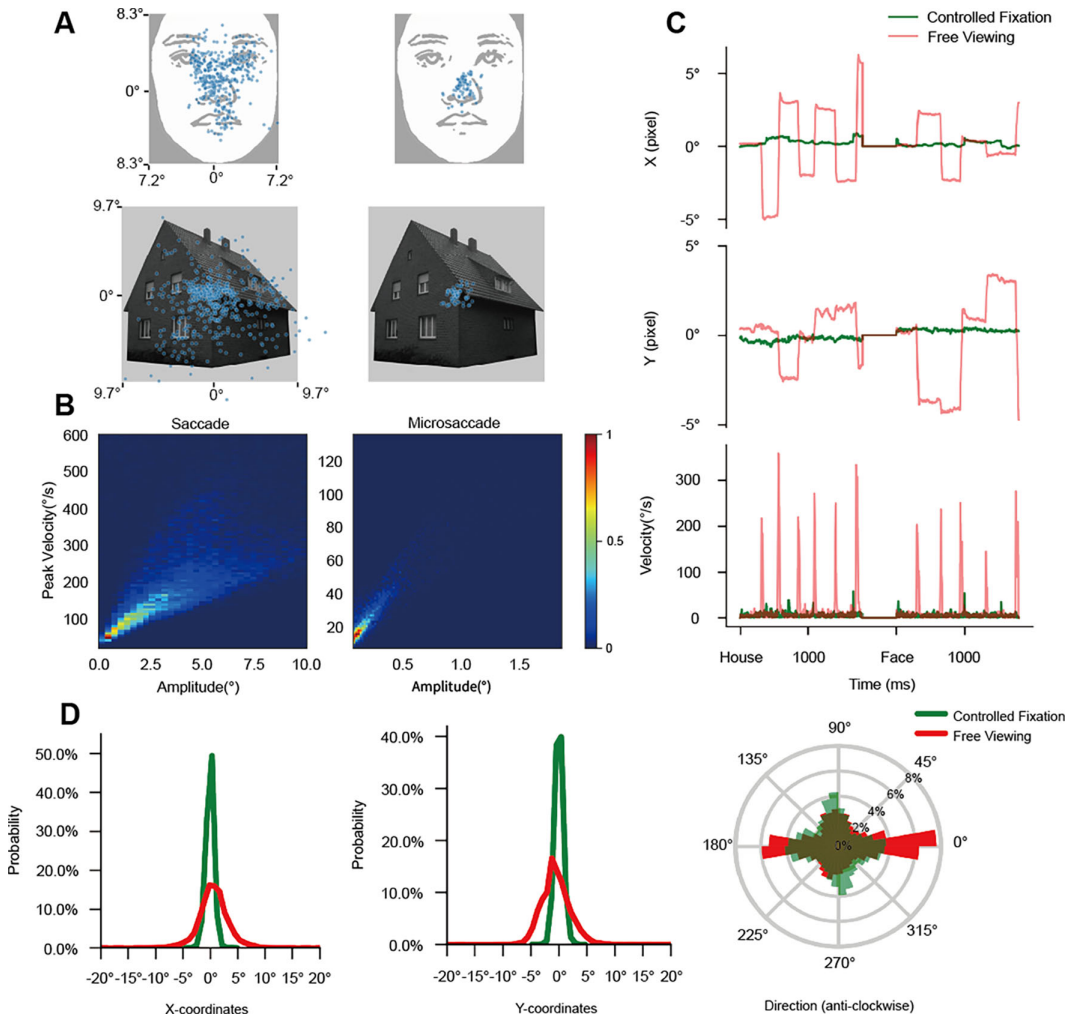

Figure 1. Eye movement patterns in the free-viewing and the central-fixation tasks, respectively. (**A**) The distribution of eye-movement patterns in the free-viewing task (left) and the central-fixation task (right). The examples shown here are the gaze positions for two example pictures pooled over participants. The face images are modified for anonymization. (**B**) Typical correlation between peak velocity and amplitude of the eye movements (main sequence) in the free-viewing task (left) and the central-fixation task (right). (**C**) The eye positions (*x*-coordinates in the upper panel; *y*-coordinates in the middle panel) and saccadic velocity (lower panel) are shown as a function of time in both the free-viewing task and the central-fixation task. The displayed data were obtained from two example trials (a face trial and a house trial) of one participant. The *x*-axis indicates the time relative to the onset of the picture. (**D**) The probability distribution of the gaze positions (left panel: *x*-coordinates in degrees relative to the central dot; middle panel: *y*-coordinates in degrees relative to the central dot) and direction (in degrees, right panel) in both the free-viewing task and the central-fixation task pooled over all trials and all participants. The 0° refers to the horizontal to the right. Note that the saccadic events more than 2° in the central-fixation task were not included in further analysis of microsaccades.

the detection performance of that block was given as feedback.

Detection of saccades and microsaccades

Eye-movement data during the picture presentation (0–1.8 seconds relative to the picture onset) were obtained for data analysis (Figure 1). For the purpose of our research question, we focused on the saccades in the free-viewing task and the microsaccades in the central-fixation task. First, data periods with missing pupil information were discarded. For the free-viewing session, saccade events were identified based on the velocity threshold of 30°/s and the acceleration threshold of 8000°/s^2. For the central-viewing session, a modified version (Engbert & Mergenthaler, 2006) of the algorithm proposed by Engbert and Kliegl (2003) was used to identify microsaccades (https://github.com/dmardanbeigi/Microsaccade-Toolbox-in-Python). Prior to detection, the data 200 ms before and 200 ms after blinks were removed to ensure stable eye

	Saccades				Microsaccades			
	Rate (N/s)	Duration (ms)	Amplitude (°)	Direction (°)	Rate (N/s)	Duration (ms)	Amplitude (°)	Direction (°)
Face								
Mean	2.84	55.11	3.61	162.22	1.59	12.33	0.22	184.32
SD	0.64	68.49	2.36	106.34	0.49	5.65	0.14	99.13
House								
Mean	2.84	61.76	4.23	185.49	1.63	12.11	0.21	187.01
SD	0.67	76.45	3.17	109.52	0.53	5.68	0.13	100.16

Table 1. Characteristics of the saccadic events (saccades in the free-viewing task and microsaccades in the central-fixation task) during face and house viewing.

movements (Krejtz, Duchowski, Niedzielska, Biele, & Krejtz, 2018). Following Engbert and Kliegl (2003), horizontal and vertical velocities were respectively computed across 11 consecutive data samples. A velocity threshold of 6 *SD* was set to extract saccadic events. In addition, a minimum intersaccadic interval of 20 ms was adopted to avoid the misidentification of potential overshoot corrections as new saccades (Siegenthaler et al., 2014). Microsaccades for further analysis were defined as saccades with a velocity range of 8°/s to 150°/s, an amplitude range of 0.08° to 2°, and a duration range of 5 to 40 ms. The descriptive characteristics of saccades and microsaccades are shown in Table 1. The typical correlations between amplitudes and peak velocities (Bahill, Clark, & Stark, 1975) are shown in Figure 1B. The distributions of the gaze positions in the two tasks are shown in Figure 1D. One may expect stronger horizontal distributions as shown in previous studies (e.g., Engbert & Kliegl, 2003). However, in contrast to previous studies, the pictures (especially faces) shown to the observers are of a vertical configuration in the present study. To successfully recognize this vertical configuration while free eye movements were constrained, the microsaccades may also have been biased for the vertical direction (Figure 1D, right panel).

Note here we chose a relatively loose criterion (i.e., amplitude of 0.08°–2°) to ensure there were enough microsaccades for data analysis. The results still held when a conservative criterion was used (0.08°–1°) (Supplementary Table S1 and Supplementary Figure S1), which led to a 6.4% loss of the microsaccades.

Multivariate classification

To investigate if saccades and microsaccades had a common spatiotemporal pattern during visual perception, we conducted classification analysis with the expectation that the classifier trained by the category of saccades (face-related saccades vs. house-related saccades) could predict the category of microsaccades (face-related microsaccades vs. house-related microsaccades), and vice versa.

For each of the detected saccadic events in each trial, the following features were obtained: index within a specific trial (in numerical order), duration (in milliseconds), x-amplitude (in degrees), y-amplitude (in degrees), direction (relative to the x-axis, in degrees), the (x, y) coordinates of the starting point (relative to the center of the screen, in pixels), and the (x, y) coordinates of the ending point (relative to the center of the screen, in pixels). To avoid over-fitting, feature selection was conducted in the following steps. First, given the dependency among the starting point, the amplitude, and the ending point, only the last two of these three features were kept. Second, an independent t-test was performed on each of the features (except the index) between the face trials and the house trials. For saccades, there were significant differences between the face trials and the house trials in duration, direction, y-amplitude, x-coordinate of the end point (x-end), and y-coordinate of the end point (y-end) (Supplementary Table S2). For microsaccades, there were significant differences in y-amplitude, x-end, and y-end. The classification analysis was then based on these significant features and the index, with the significant features capturing the properties of single saccadic events and the index capturing the temporal relation of the saccadic events. We termed these features "saccadic features" and "microsaccadic features," respectively.

The classification analysis was performed using the scikit-learn package (http://github.com/scikit-learn) (Pedregosa et al., 2011). Specifically, two directions of classification were performed. In one direction, a linear support vector machine (SVM) classifier was trained and cross-validated based on the saccadic features of the two categories (face vs. house) in the free-viewing task. To investigate pattern similarity between the saccades and microsaccades, the performance of the classifier was tested in predicting the two categories in the central-fixation task. In the other direction, a

classifier was trained and cross-validated based on the microsaccadic features in the central-fixation task and was tested in the free-viewing task. To normalize the different scales of saccades and microsaccades, the features within each saccadic type were transformed into *Z*-scores for the classification analysis.

Permutation-based testing was conducted to assess the statistical significance. For each participant, a 5-fold leave-one-out cross-validation procedure was conducted within a saccadic type (saccades or microsaccades). The 5-fold leave-one-out cross-validation was performed 50 times by grouping different trials into different folds (without repetition), and the median of the 250 predicting accuracies was used to represent the within-type accuracy. Here, we used the median as the sample estimate because bimodal distributions of the predicting accuracies were observed in some participants. Then a permutation-based accuracy was calculated by shuffling the labels of the two categories with the same 5-fold cross-validation procedure. This permutation was repeated 1000 times, resulting in 1000 chance accuracies for each participant. For group-level statistical testing, one chance accuracy was randomly selected from the set of chance accuracies, and the median of these individual chance accuracies was used to represent the group chance accuracy. This procedure was repeated 1000 times with the replacement of the individual accuracy, resulting in a distribution of 1000 group chance accuracies. Significance testing was performed by calculating the probability of the unpermuted median accuracy across participants in the distribution of the permuted group chance accuracies (one-tailed). The same permutation procedures were conducted to test the cross-type predicting accuracies, except that one saccadic type was included as the training set and the other type was included as the test set.

Representational distance

To quantify the similarity of the eye movement patterns among the different experimental conditions, we calculated the representational distance based on three features: the *x*- and *y*-coordinates (in pixels) of the landing position and the direction (in degrees) of the saccades/microsaccades (Figure 2C). We chose these three features because of their geometrical independence, which satisfies the assumption of Euclidean distance. The representational distance was then quantified by calculating the Euclidean distance in the three-dimensional space with the scipy package (https://scipy.org/), assuming that a lower distance indicates a higher representational similarity (Kriegeskorte, Mur, & Bandettini, 2008).

The first prediction concerning the representational distance was that individuals who showed larger saccade-based representational distance (SRD) between face and house would also show larger microsaccade-based representational distance (MRD). For each participant, we calculated the SRD matrix between face trials and house trials in the free-viewing task, and the MRD matrix between face trials and house trials in the central-fixation task. The SRD/MRD values in the corresponding matrices for each participant were averaged into a mean SRD/MRD. A Pearson correlation (one-tailed) was then conducted between the individual SRD (face vs. house) and the individual MRD (face vs. house).

The second prediction was that the representational distance between saccades and microsaccades (SMRD) would be lower for face than for house, because the highly structured face images would be accompanied by highly structured eye movements, and the SMRD would also be lower for face than for cross-category (i.e., face–house) SMRD. For each participant, we calculated the SMRD matrix among face trials (face–face SMRD), among house trials (house–house SMRD), and between face and house trials in the two tasks, respectively. The SMRD values of each of the three categories were averaged within each observer. A repeated-measures analysis of variance (ANOVA) was conducted to compare the face–face, house–house, and face–house SMRDs. Pairwise comparisons with Bonferroni corrections were performed following a significant effect.

Orienting-position preference

The eye region and the mouth region were defined by facial landmarks (Sagonas, Tzimiropoulos, Zafeiriou, & Pantic, 2013). For each face image, the facial landmarks were detected with Dlib (https://github.com/davisking/dlib) and a trained dataset (http://dlib.net/files/shape_predictor_68_face_landmarks.dat.bz2). The eye region was defined as the triangle area where the central fixation point (nose, the 31st landmark in Sagnosa et al., 2013) was first set as one vertex (Figure 3A, left). Then, one side of the triangle was defined as the line from the central fixation point to the tail of the left eye (the 37th landmark), and the second side was defined as the line from the central fixation point to the tail of the right eye (the 46th landmark). The third side was defined as the horizontal line going through the top edge of the eyebrows (the 20th or 25th landmark) and vertical to the central fixation. The meeting points of the horizontal side and the other two sides were defined as the vertexes. Similarly, the mouth region was defined as the triangle area where the central fixation point was first set as one vertex. Then, one side of the triangle was defined as the line from the central fixation point to the left corner of the mouth (the 49th landmark), and the second side was the line from the central fixation point

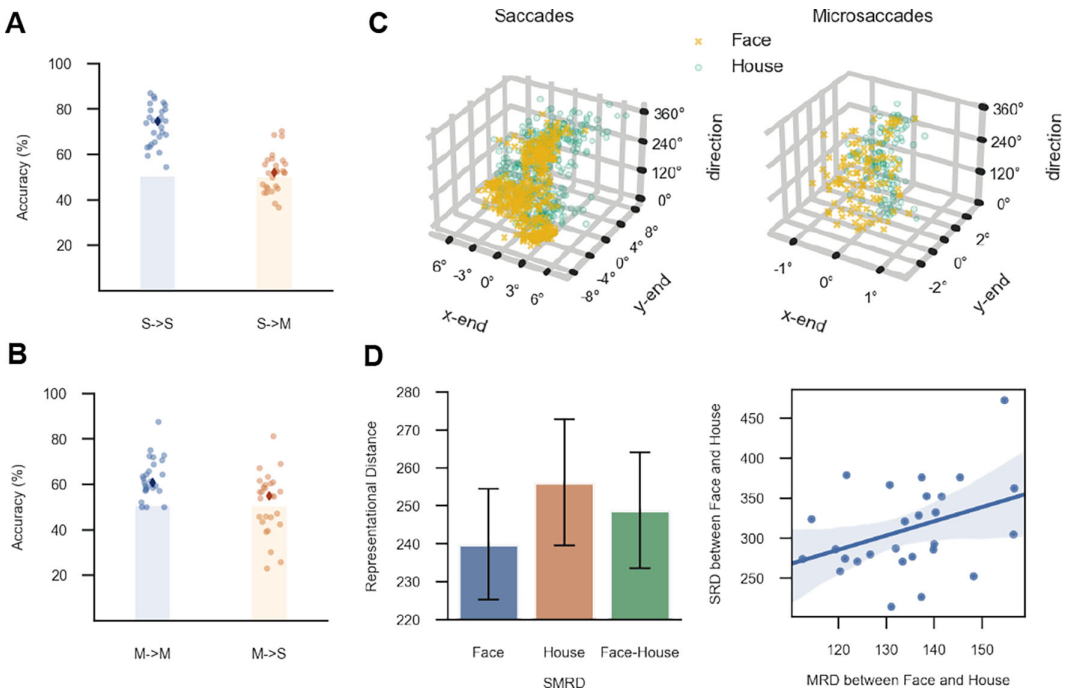

Figure 2. (**A**) The accuracies of the classifier trained with the saccadic features in the free-viewing task in predicting the categories (face vs. house) in the free-viewing task (S->S) and the central-fixation task (S->M). The dots indicate the individual accuracies, and the diamonds indicate the median accuracy. The shaded areas indicate accuracies below the 95th percentile of the null distribution obtained from the permutations. (**B**) The accuracies of the classifier trained with the microsaccadic features in the central-fixation task in predicting the categories (face vs. house) in the central-fixation task (M->M) and the free-viewing task (M->S). (**C**) The three-dimensional (*x*- and *y*-coordinates of the landing position and direction) representational space of the saccadic parameters (left, saccades; right, microsaccades). The data points are the saccadic parameters from one participant (yellow for face and cyan for house). (**D**) The representational distance between saccades and microsaccades for face–face, house–house, and face–house pairs (left panel, error bars indicate 95% CIs). The scatterplot (with best-fitting line, right panel) illustrates the individual SRD between face and house as a function of the individual MRD.

to the right corner of the mouth (the 55th landmark). The third side was defined as the horizontal line going through the other two lines, with the constraint that the formed triangle had the same size in area as the eye region. This was to ensure that the potential orienting probability difference between the two regions could not be simply due to the size difference (e.g., larger size leads to higher orienting probability).

For each participant, we calculated the orienting position probability of the saccades/microsaccades in the above-defined eye region and mouth region, respectively. To test if there was an overall preference for one region over another, a 2 (saccade vs. microsaccade) $\times$ 2 (eye region vs. mouth region) ANOVA was conducted on the probabilities.

For each participant, we also calculated the preference for the eye region by subtracting the probability in the mouth region from the probability in the eye region. To test the consistency of the preference, a Pearson correlation was then performed on the preference scores between saccades and microsaccades.

Results

Behavioral performance

The recognition task was performed equally well in the free-viewing task and the central-fixation task ($t < 1$), with mean accuracy values of 98.4% $\pm$ 1.2% in the free-viewing task and 98.2% $\pm$ 1.3% in the central-fixation task.

Classification

The classification analysis revealed significant accuracies for both within-type prediction and cross-type prediction. Specifically, the classification based on the saccadic features showed significantly above-chance accuracies (74.6%) in predicting the categories (face vs. house) within the free-viewing task ($p < 0.001$), as well as in the central-fixation task (52.0%; $p < 0.001$) (Figure 2A). In the other direction,

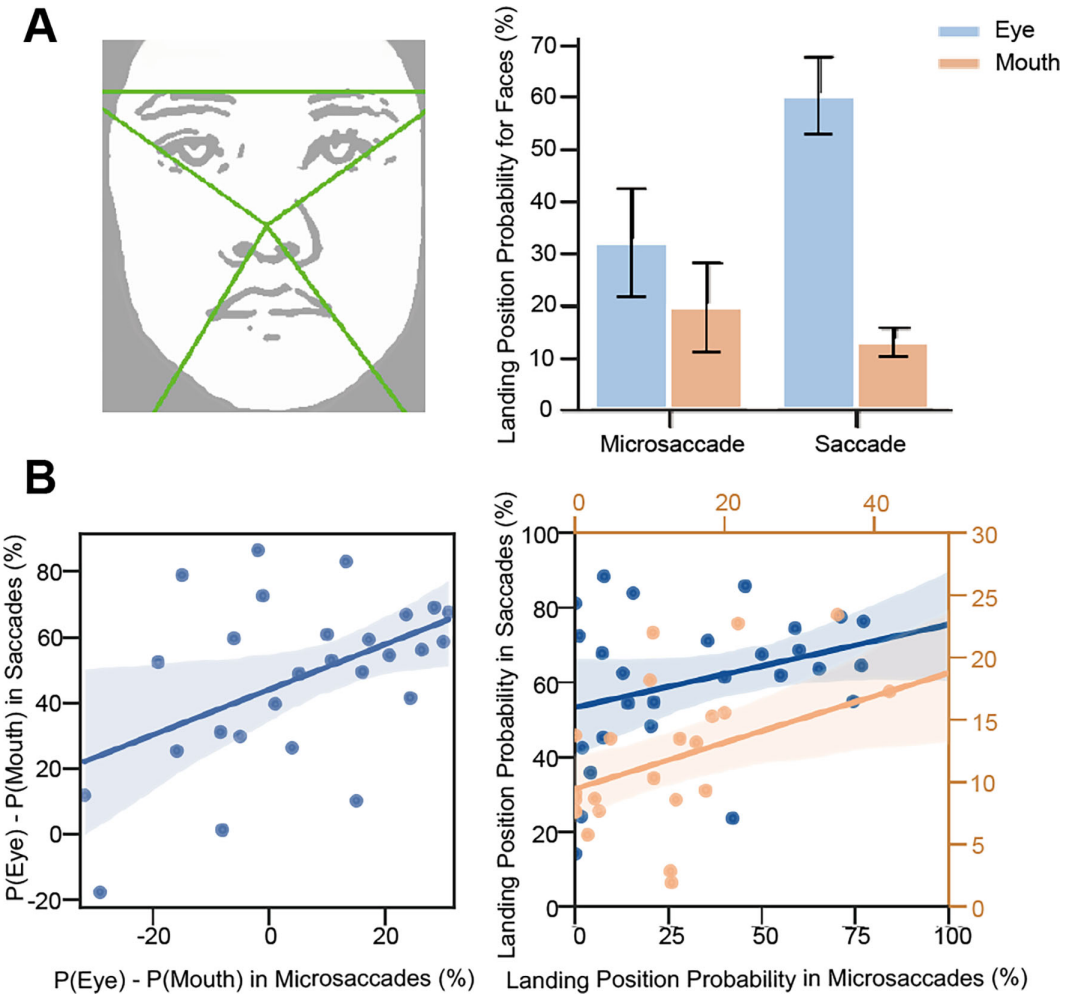

Figure 3. (**A**, left) The eye region (upper triangle) and the mouth region (lower triangle) used to calculate the orienting probabilities of the saccades and microsaccades. The face image is modified for anonymization. (**A**, right) The orienting probabilities are shown as a function of saccadic type and region. Error bars indicate 95% CIs. (**B**, left) The scatterplot (with best-fitting line) illustrates the individual preference scores of saccades for the eye region over the mouth region (probability difference) as a function of the individual preference scores of microsaccades. (**B**, right) The scatterplots (with best-fitting lines) illustrate the individual orienting probabilities of saccades (blue for the eye region and orange for the mouth region) as a function of the individual orienting probabilities of microsaccades. These two scatterplots are included to show that the correlation patterns are consistent for both the eye region and the mouth region.

the classification based on the microsaccadic features showed significantly above-chance accuracies (60.9%) in predicting the categories within the central-fixation task ($p < 0.001$), as well as in the free-viewing task (53.6%; $p < 0.001$) (Figure 2B). These results suggest a common spatiotemporal structure between saccades and microsaccades during visual perception of faces and houses. One might be concerned that the saccadic events detected in the free-viewing task contained small-scale saccades that shared the parameters with the microsaccades detected in the central-fixation task and hence confounded with the cross-type classification. To exclude this alternative, we removed the small-scale saccades in the free-viewing task (the ones shared the amplitude and duration ranges of the microsaccades in the central-fixation task, 23.1% of all saccades). The classifications showed the same pattern of results (Supplementary Table S1).

Representational distance

We have assumed that the representational distance reflects the structural dissimilarity of eye movements. As an assumption check, we first tested the measurement with the saccade data in the free-viewing task. The results showed that the face–face SRD was lower than house–house SRD and face–house SRD (Supplementary Figure S2), confirming our assumption that the stimuli with high structure similarity (e.g.,

faces) would be accompanied by a low representational distance of eye movements.

Importantly, there was a positive correlation between the SRD (face vs. house) and the MRD (face vs. house), with Pearson $r = 0.379$ and $p = 0.026$ (Figure 2D, right), suggesting that individuals who had distinctive saccade structures between face and house also had distinctive microsaccade structures between face and house. The ANOVA on the SMRD (face–face, house–house vs. face–house) showed a significant main effect, $F(2, 52) = 11.60$, $p < 0.001$, $\eta_p^2 = 0.309$ (Figure 2D, left). Further pairwise comparisons showed that the face–face SMRD was lower than both the house–house SMRD ($p < 0.001$) and the face–house SMRD ($p = 0.033$), whereas the difference between the latter two did not reach significance ($p = 0.101$, Bonferroni corrected). These results indicate a higher structural similarity between saccades and microsaccades during face viewing.

Orienting-position preference

The 2 (saccade vs. microsaccade) × 2 (eye region vs. mouth region) ANOVA showed a main effect of orienting region, $F(1, 26) = 25.50$, $p < 0.001$, $\eta_p^2 = 0.495$, with significantly higher probabilities in the eye region (46.3%) than in the mouth region (16.4%) (Figure 3A, right). This result suggested an overall preference for the eye region over the mouth region in face perception. There was a main effect of saccadic type, $F(1, 26) = 15.51$, $p < 0.001$, $\eta_p^2 = 0.374$, indicating higher probabilities of saccades (36.7%) in orienting to the combined eye–mouth region than microsaccades (26.0%). The interaction between saccadic type and orienting region was also significant, $F(1, 26) = 21.01$, $p < 0.001$, $\eta_p^2 = 0.448$, which was due to the more pronounced preference for the eye region in saccades (probability difference eye vs. mouth: 47.3%) than in microsaccades (probability difference eye vs. mouth: 12.4%). Further pairwise tests on the simple effects showed a significant region difference for saccades, $t(26) = 9.60$, $p < 0.001$, but the difference for microsaccades did not reach significance, $t(26) = 1.44$, $p = 0.162$. However, this insignificant difference should not be taken as "there was no region preference in microsaccades" (i.e., the null hypothesis). To evaluate the null hypothesis, we performed the Bayes factor analysis, which showed that the alternative hypothesis "there was a preference for the eye region over the mouth region" is 1.953 times more likely to be true than the null hypothesis (i.e., BF = 1.953). The results suggest that the null hypothesis should not be accepted and that the main effect revealed by the ANOVA was not entirely driven by saccades. Instead, it was also driven, although to a lesser extent, by microsaccades (which can also be seen by the mean difference of 12%). Moreover, the preference for the eye region over the mouth region, as quantified by the difference in orienting probabilities, showed a positive correlation between saccades and microsaccades (Pearson $r = 0.483$; $p = 0.011$; 95% confidence interval [CI], 0.127–0.729) (Figure 3B, left), suggesting a consistent orienting preference across saccades and microsaccades.

As a sanity check, we defined the mouth region as the triangle first and then defined the eye region according to the size of the mouth region. The same pattern of results was observed (see Supplementary Figure S3). One potential alternative account for the higher orienting probabilities in the eye region than in the mouth region could be that the observers had an overall preference for the upper visual field over the lower visual field. To test this alternative account, we calculated the orienting probabilities for house images. Specifically, for each house image, we defined the upper area as the average area of the eye regions across the face images and the lower area as the average area of the mouth regions across the face images. The 2 (saccade vs. microsaccade) × 2 (eye region vs. mouth region) ANOVA showed only a main effect of saccadic type, $F(1, 26) = 12.30$, $p = 0.002$, $\eta_p^2 = 0.321$, with higher probabilities of saccades (24.3%) orienting to the combined region than microsaccades (16.5%). However, neither the main effect of the orienting region, $F(1, 26) = 1.90$, $p = 0.180$, nor the interaction ($F < 1$) reached significance. These results suggest that the preference for the eye region over the mouth region observed in both saccades and microsaccades cannot be simply due to a preference for the upper visual field over the lower visual field.

Discussion

Previous studies have shown similarities between saccades and microsaccades during high-level visual perception (Mergenthaler & Engbert, 2010; Otero-Millan, Troncoso, Macknik, Serrano-Pedraza, & Martinez-Conde, 2008). For example, the rates of both saccades and microsaccades increased as a function of the increasing demand for visual exploration, and the two kinds of saccades had comparable intersaccadic intervals regardless of different visual stimulation (Otero-Millan et al., 2008). During scene perception, the saccade rates in the free-viewing task showed a positive correlation with the microsaccade rates in the fixation-controlled task across individuals (Mergenthaler & Engbert, 2010). In an extension of these findings, the current results showed that saccades and microsaccades shared geometrical properties that related to a specific object category. Specifically, both saccades during free viewing and microsaccades during controlled fixation showed distinctive spatiotemporal structures between face and

house viewing. Importantly, the classification model trained with the saccadic patterns in discriminating face versus house viewing can be generalized to classifying the face- versus house-related microsaccadic patterns, and vice versa. Moreover, individuals who had more distinctive saccadic patterns between face and house viewing also had more distinctive microsaccadic patterns. The category-distinguishing patterns shared between saccades and microsaccades suggest that a common oculomotor program was used to explore a specific visual category.

A further prediction of a common oculomotor program in visual perception is that the structural similarity between saccades and microsaccades would be more pronounced for visual stimuli with a consistent structure such as faces. In free-viewing contexts, it has been shown that the saccade patterns within an observer were consistent across face exemplars and task sets (Mehoudar et al., 2014; Peterson & Eckstein, 2012). In agreement with our prediction, the results of representational distance analysis showed that the structural similarity between saccades and microsaccades for faces was higher than the structural similarity for houses. As a common oculomotor program, the shared structure of gaze traces may act to optimize the information sampling during visual exploration (Martinez-Conde et al., 2009). Taking face perception as an example, saccades and microsaccades would both be tuned to actively sample the critical information (eyes/ nose/mouth) to achieve recognition. Although saccades directly land in the critical regions to optimize the information sampling, microsaccades may be tuned to the direction of these regions, as the direction of microsaccades has been shown to reflect the shift of covert attention (Engbert & Kliegl, 2003; Hafed & Clark, 2002). The current results indicate that there were higher orienting probabilities in the eye region than in the mouth region for both saccades and microsaccades, and the individual preference for the eye region over the mouth region showed a positive correlation between the two kinds of saccades.

A similar pattern between saccades and microsaccades during face perception was also observed in a recent study (Shelchkova et al., 2019). In this study, saccade patterns were examined when faces were presented in a size extended into parafoveal vision (around 4° of height), and microsaccade patterns were examined in fine spatial resolution when faces were presented within foveal vision (around 1° of height). Despite the different spatial scales of the faces, a shared preference for the landing region (eyes/mouth/nose) was observed for saccades and microsaccades, and the two kinds of saccades had a similar temporal course in each of the regions. Although both the results of Shelchkova et al. (2019) and the current results support a common structure across saccades and microsaccades, different mechanisms should be noted during the microsaccadic processing. In Shelchkova et al. (2019), the microsaccades were mainly involved in bringing the interested information into the fixation locus within foveal vision where all information was available. In the present study, however, the pictures were presented in a size extending to peripheral vision while the central fixation was controlled. The microsaccades may instead be tuned in a way that the interested information could be better, although still suboptimally, utilized by peripheral vision. The function of microsaccades in visual perception may be not only to optimize the spatial tuning within foveal vision (Poletti et al., 2013; Rucci, Iovin, Poletti, & Santini, 2007) but also to improve the quality of peripheral vision (McCamy et al., 2012).

It should be noted that the preference of saccades and microsaccades for a specific region cannot be equated to the preference shown in previous studies (Peterson & Eckstein, 2012; Peterson & Eckstein, 2013). In these studies, an unpredictable fixation point was presented at a peripheral location prior to the face presentation. Due to the lack of prior information, the saccade preference for a specific region was thus more idiosyncratic among the observers. In the current study, to achieve a fair comparison between saccades and microsaccades, participants were required to fix their gaze on the nose (i.e., where the central fixation was presented) in both the free-viewing task and the central-fixation task. The initial sweep of the face images during the starting gaze position may provide prior information to guide the following gaze traces, leading to an overall preference for the eye region over the mouth region among the observers.

The structural similarity between saccades and microsaccades in visual object perception can be understood in the framework of spatial attention. Although overt spatial attention is coupled with saccades, covert attention is marked by microsaccades (Engbert & Kliegl, 2003; Hafed & Clark, 2002; Yuval-Greenberg et al., 2014). To form an object percept and achieve recognition, the information on the space covered by the object has to be sampled and accumulated. In free-viewing contexts, the information sampling is implemented by the normal-scale saccades (i.e., overt attention). In fixation-constrained contexts such as the central-fixation condition here, covert attention has to be paid to the critical areas of the visual object, which leads to the microsaccades being biased toward the attended areas. Moreover, the similarity between saccades and microsaccades covers a broader scope than the domains of spatial attention and object perception. It has been suggested that saccades and microsaccades are generated and controlled by the same oculomotor system (Hafed, 2011; Hafed & Krauzlis, 2012; Rolfs, Kliegl, & Engbert, 2008; Zuber, Stark,

& Cook, 1965) and similarly modulate early visual processing (McFarland, Bondy, Saunders, Cumming, & Butts, 2015; Scholes, McGraw, & Roach, 2018). Taken together, the evidence points to a continuum of eye movement rather than a boundary between saccades and microsaccades.

The current findings not only suggest a common cognitive role of microsaccades and saccades in visual perception but also provide important insights into how eye movements should be treated in vision studies. For example, in studying the neural mechanism of visual perception, the eye gaze is often maintained at the central fixation to ensure that the observed neural activities cannot be due to the eye movements. However, even during the maintained central fixation, as shown here, there were structured microsaccades. These structured microsaccades might have a contribution to the neural activities similar to that of the saccades (Stacchi, Ramon, Lao, & Caldara, 2019; Wang et al., 2019), and hence should not be assumed to be irrelevant to the object-related neural representations. Nevertheless, so far we have only provided eye-movement evidence based on two visual categories; generalization to other categories and the corresponding contributions to the neural representations must be verified by future studies.

Conclusions

In summary, our results showed a common structure of gaze traces between saccades and microsaccades during visual object perception. These object-related gaze traces may derive from a common oculomotor program that is used to optimize information sampling for the forming of visual perception.

Keywords: eye movements, microsaccades, face perception, representational similarity

Acknowledgments

The authors thank Xu Liu, PhD, for his suggestions and technical support in data analysis, and Yiwen Zhu for her assistance in data analysis.

Funded by grants from the National Natural Science Foundation of China (32271086), the Shanghai Sailing Program (20YF1422100), and the Deutsche Forschungsgemeinschaft (PO548/18-1). L.W. is supported by a Mercator Fellowship of the Deutsche Forschungsgemeinschaft (450600965).

Commercial relationships: none.
Corresponding author: Lihui Wang.
Email: lihui.wang@sjtu.edu.cn.
Address: Institute of Psychology and Behavioral Science, Shanghai Jiao Tong University, Shanghai 200240, China.

References

Arizpe, J., Kravitz, D. J., Yovel, G., & Baker, C. I. (2012). Start position influences fixation patterns during face processing: Difficulties with eye movements as a measure of information use. *PLoS One, 7*, e31106.

Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences, 24*, 191–204.

Cohen, E. H., & Tong, F. (2015). Neural mechanisms of object-based attention. *Cerebral Cortex, 25*, 1080–1092.

Cunitz, R. J., & Steinman, R. M. (1969). Comparison of saccadic eye movements during fixation and reading. *Vision Research, 9*, 683–693.

Ditchburn, R. W., Fender, D. H., & Mayne, S. (1959). Vision with controlled movements of the retinal image. *Journal of Physiology, 145*, 98–107.

Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research, 43*, 1035–1045.

Engbert, R., & Mergenthaler, K. (2006). Microsaccades are triggered by low retinal image slip. *Proceedings of National Academy of Sciences, USA, 103*, 7192–7197.

Engbert, R., Mergenthaler, K., Sinn, P., & Pikovsky, A. (2011). An integrated model of fixational eye movements and microsaccades. *Proceedings of National Academy of Sciences, USA, 108*, E765–E770.

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.

Hafed, Z. M. (2011). Mechanisms for generating and compensating for the smallest possible saccades. *European Journal of Neuroscience, 33*, 2101–2113.

Hafed, Z. M., & Clark, J. J. (2002). Microsaccades as an overt measure of covert attention shifts. *Vision Research, 22*, 2533–2545.

Hafed, Z. M., & Krauzlis, R. J. (2012). Similarity of superior colliculus involvement in microsaccade and saccade generation. *Journal of Neurophysiology, 107*, 1904–1916.

Haxby, J. V., Gobbini, M. I., & Nastase, S. A. (2020). Naturalistic stimuli reveal a dominant role for agent

action in visual representation. *NeuroImage, 216*, 116561.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7*, 498–504.

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition, 33*, 98–106.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*, 4302–4311.

Kelly, D. H. (1979). Motion and vision. I. Stabilized images of stationary gratings. *Journal of Optical Society of America, 62*, 685–689.

Ko, H. K., Poletti, M., & Rucci, M. (2010). Microsaccades precisely relocate gaze in a high visual acuity task. *Nature Neuroscience, 13*, 1549–1553.

Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., & Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLoS One, 13*, e0203629.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2*, 4.

Liu, Z., Rosenbaum, R. S., & Ryan, J. D. (2020). Restricting visual exploration directly impedes neural activity, functional connectivity, and memory. *Cerebral Cortex Communications, 1*, 1–15.

Martinez-Conde, S., Macknik, S.L., Troncoso, X., & Hubel, D. H. (2009). Microsaccades: A neurophysiological analysis. *Trends in Neurosciences, 32*, 463–475.

Martinez-Conde, S., Otero-Millan, J., & Macknik, S. L. (2013). The impact of microsaccades on vision: Towards a unified theory of saccadic function. *Nature Reviews Neuroscience, 14*, 83–96.

McCamy, M. B., Otero-Millan, J., Macknik, S. L., Yang, Y., Troncoso, X. G., & Baer, S. M., ...Martinez-Conde, S. (2012). Microsaccadic efficacy and contribution to foveal and peripheral vision. *Journal of Neuroscience, 32*, 9194–9204.

McFarland, J. M., Bondy, A. G., Saunders, R. C., Cumming, B. G., & Butts, D. A. (2015). Saccadic modulation of stimulus processing in primary visual cortex. *Nature Communications, 6*, 8110.

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 Vision, 14*(7):6, 1–11, https://doil.org/10.1167/14.7.6.

Mergenthaler, K., & Engbert, R. (2010). Microsaccades are different from saccades in scene perception. *Experimental Brain Research, 203*, 753–757.

Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research, 11*, 929–942.

Otero-Millan, J., Macknik, S. L., Langston, R. E., & Martinez-Conde, S. (2013). An oculomotor continuum from exploration to fixation. *Proceedings of the National Academy of Sciences, USA, 110*, 6175–6180.

Otero-Millan, J., Troncoso, X. G., Macknik, S. L., Serrano-Pedraza, I., & Martinez-Conde, S. (2008). Saccades and microsaccades during visual fixation, exploration, and search: Foundations for a common saccadic generator. *Journal of Vision, 8*(14):21, 1–18, https://doi.org/10.1167/8.14.21.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., & Grisel, O., ...Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences, USA, 109*, 3314–3323.

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*, 1216–1225.

Poletti, M., Listorti, C., & Rucci, M. (2013). Microscopic eye movements compensate for nonhomogeneous vision within the fovea. *Current Biology, 23*, 1691–1695.

Poletti, M., & Rucci, M. (2016). A compact field guide to the study of microsaccades: Challenges and functions. *Vision Research, 118*, 83–97.

Rolfs, M. (2009). Microsaccades: Small steps on a long way. *Vision Research, 49*, 2415–2441.

Rolfs, M., Kliegl, R., & Engbert, R. (2008). Toward a model of microsaccade generation: The case of microsaccadic inhibition. *Journal of Vision, 8*(11):5, 1–23, https://doi.org/10.1167/8.11.5.

Rucci, M., Iovin, R., Poletti, M., & Santini, F. (2007). Miniature eye movements enhance fine spatial detail. *Nature, 447*, 851–854.

Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., & Pantic, M. (2013). 300 faces in-the-wild challenge: The first facial landmark localization challenge. In C. Sagonas, G. Tzimiropoulos, S. Zaveiriou & M. Pantic (Eds.), *Proceedings of the 2013 IEEE International Conference on Computer Vision*

Workshops (pp. 397–403). Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Scholes, C., McGraw, P. V., & Roach, N. W. (2018). Selective modulation of visual sensitivity during fixation. *Journal of Neurophysiology, 119*, 2059–2067.

Shelchkova, N., Tang, C., & Poletti, M. (2019). Task-driven visual exploration at the foveal scale. *Proceedings of the National Academy of Sciences, USA, 116*, 5811–5818.

Siegenthaler, E., Costela, F. M., McCamy, M. B., Di Stasi, L. L., Otero-Millan, J., & Sonderegger, A., ...Martinez-Conde, S. (2014). Task difficulty in mental arithmetic affects microsaccadic rates and magnitudes. *European Journal of Neuroscience, 39*, 287–294.

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

Wang, L., Baumgartner, F., Kaule, F. R., Hanke, M., & Pollmann, S. (2019). Individual face and house-related eye movement patterns distinctively activate FFA and PPA. *Nature Communications, 10*, 5532.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.

Yuval-Greenberg, S., Merriam, E. P., & Heeger, D. J. (2014). Spontaneous microsaccades reflect shifts in covert attention. *Journal of Neuroscience, 34*, 13693–13700.

Zuber, B. L., Stark, L., & Cook, G. (1965). Microsaccades and the velocity-amplitude relationship for saccadic eye movements. *Science, 150*, 1459–1460.