Exploring roles for eye movements in face recognition

Maria Haddad, Integrated Program in Neuroscience, McGill University, Montreal

February 2023

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master's of Science.

Table of contents

1. English abstract

The human visual system is highly developed for detecting visual objects, with faces being particularly salient and important objects. Scanning faces with our gaze (i.e. 'free viewing', with eye movements allowed) is thought to be important for our ability to learn new faces, as well as for our facility to recognize familiar ones. Specifically, gaze-tracking research has shown that we tend to make stereotyped eye movements when viewing faces, focusing our gaze on 'internal' facial structures including the eyes, the nose, and the mouth. However, there are disagreements in the field regarding specific roles for saccadic eye movements in face perception, as familiar faces can be recognized under 'passive' viewing (i.e. by fixating on a point). We hypothesize that contradictory results in the field have arisen in part because most studies used faces small enough to be recognized without absolutely requiring eye movements. We thus set out to specifically examine the effect of varying face size on the discrimination of familiar and unfamiliar faces, in fixation and free viewing conditions. We hypothesized that as faces get bigger and the internal facial features fall out of central view, discrimination will not be possible under forced fixation, but will be maintained under free viewing. Interestingly, for both viewing conditions, we find that humans are able to recognize familiar faces and discriminate them from unfamiliar faces over a very large range of sizes (roughly 1-115 degrees in height). Additionally, combined eye and head movements only slightly increased the range over which discrimination between familiar and unfamiliar faces is possible. These results provide insights about the requirement of eye movements in face recognition.

2. French abstract

Le système visuel humain est très développé pour détecter tous les objets visuels, surtout ceux particulièrement saillants et importants tel que les visages. Notre capacité à apprendre de nouveaux visages et à reconnaître des visages familiers semble dépendre de leur observation ‘libre’ (c’est-à-dire, avec mouvements oculaires permis). Plus précisément, nous avons tendance à produire des mouvements oculaires stéréotypés lorsque nous regardons des visages: notre regard est souvent concentré sur les structures faciales dites ‘internes’, comme les yeux, le nez et la bouche. Cependant, il existe des désaccords importants concernant la nécessité absolue de ces mouvements oculaires pour la perception de visages, puisque qu’il est possible de reconnaître un visage familier lors d’une observation ‘passive’ (c’est-à-dire, en fixant un point). Nous supposons que ces contradictions surviennent parce que la plupart des études précédentes utilisaient une taille d’image suffisamment petite ne nécessitant pas de mouvements oculaires pour la reconnaissance des visages. Notre avons donc entrepris d'examiner l'effet de la variation de la taille des visages familiers sur leur discrimination de visages inconnus, sous conditions d’observation ‘passive’ et ‘libre’. Nous avons émis l’hypothèse que quand les visages sont suffisamment grands pour que les traits internes ne sont plus sous la partie centrale du champ visuel, la discrimination de la familiarité des visages ne sera pas possible sous observation passive, mais elle le sera toujours sous observation libre. À travers notre étude, nous constatons que, sous les deux conditions d’observation, les visages familiers peuvent être reconnus et distingués des visages inconnus sur un intervalle de tailles très large (environ 1 à 115 degrés de hauteur). De plus, combiner les mouvements des yeux à des mouvements de tête permet d’étendre davantage cet intervalle de discrimination. Ces résultats élargissent donc nos connaissances concernant l'exigence des mouvements oculaires pour la reconnaissance des visages.

3. Acknowledgements

This project was completed under the supervision of Dr. Daniel Guitton (Professor, Departments of Neurology and Neurosurgery, Physiology, and Biomedical Engineering) and Dr. Stuart Trenholm (Assistant Professor, Department of Neurology and Neurosurgery) at the Montreal Neurological Institute and Hospital. I thank the members of my advisory committee – Dr. Suresh Krishna, Dr. Arjun Krishnaswamy, and Dr. Christopher Pack – for providing useful advice and feedback regarding the project during and outside of yearly meetings, and Dr. Jean-François Poulin for chairing my M.Sc. Thesis Seminar.

Aude Villemain, the Trenholm laboratory manager, helped by preparing all documents submitted to the Research Board Ethics for approval of the study, including participant consent forms, as well as preparing advertisement documents for the study and managing documents pertaining to participants. Ms. Villemain was also in charge of ordering all the materials and equipment required for the study, as well as managing payment for study participation.

The base virtual reality environment used for the experiments was written and set up by Jonas Lehnert, a Ph.D. candidate in the Krishnaswamy Laboratory (Bellini Life Sciences Complex, McGill University). The original environment was modified to generate new versions with the help of Dr. Rudi Tong, a postdoctoral fellow in the Trenholm Laboratory, and with occasional consultation with Jonas Lehnert.

The experiments using the classic eye tracking system were set up with the help of Asmara Awada, a M.Sc. graduate from the Pack Laboratory (Montreal Neurological Institute, McGill University), Pooya Laamerad, a Ph.D. candidate in the Guitton Laboratory and the Pack Laboratory, and Dr. Matthew Krause, a research associate in the Pack Laboratory.

The project would not have been possible without family members, friends, and colleagues who volunteered to take part in the pilot studies: Lucia Guerra, Angelina Haddad, Gaby Haddad, Pooya Laamerad, Claudia Richa, Tudor Sintu, Kelvin Tian, Dr. Tong, and Ms. Villemain.

Nathan Collett, a previous undergraduate student in the Trenholm Laboratory, helped in the advertisement of the study and the recruitment of participants.

Lucia Guerra helped to collect data by running one experiment session.

Kadjita Asumbisa, a Ph.D. candidate in the Trenholm Laboratory and Peyrache Laboratory (Montreal Neurological Institute, McGill University), Dr. Ronan da Silva, a previous postdoctoral fellow in the Trenholm Laboratory, Dr. Krause, and Dr. Tong provided general help in writing code to analyse the data.

Dr. Jose Correa (the director of the McGill Statistical Consulting Service), Dr. Pedro Gabrielle Vieira (a research associate in the Pack Laboratory), Dr. Krause, Pooya Laamerad, and Dr. Tong provided useful feedback concerning the methods of analysing participant task performance. The use psychometric curves to analyse participant performance was suggested by Dr. Tong. The analysis of eye movement was developed with the help of Dr. Krause, Dr. Golia Shafiei, a Ph.D. graduate from the Misic Laboratory (Montreal Neurological Institute, McGill University), and Dr. Tong.

Finally, I would like to thank people who supported me emotionally throughout my studies, my family: Angelina, Gaby, and Claudia; and friends: Tudor, Rory, Luna, Rudi, Matt, Pooya, Ronan, Amanda, Christina, Airi, Lucia, Kelvin, Shahd, and Lizzie.

4. <u>Contribution of Authors</u>

This thesis was entirely written by Maria Haddad.

Dr. Tong wrote code used in the preparation of the stimuli for the experiment (thin plate spline warping function) and code to generate the foveated images shown in **Figure 3**. Additionally, Dr. Tong and Dr. Krause provided significant amounts of help at every stage of data analysis and discussion of results.

Dr. Daniel Guitton, Dr. Matthew Krause, Dr. Rudi Tong and Dr. Stuart Trenholm provided feedback and comments regarding this document.

5. List of Figures and Tables

Tables:

Figures:

6. List of Abbreviations

- AFC: Alternative forced choice
- ANOVA: analysis of variance
- cm: centimeters
- CVD: colour vision deficiency
- DBSCAN: Density-Based Spatial Clustering of Applications with Noise
- EL: EyeLink 1000 Plus
- FDR: false discovery rate
- H_0: null hypothesis
- HMM: hidden Markov model
- Hz: hertz
- LED: light-emitting diode
- ms: milliseconds
- PC: principal component
- PCA: principal component analysis
- PI-20: 20-Item Prosopagnosia Index
- RBG: red, green, and blue colour model
- ROI: region of interest
- SEM: standard error of the mean
- SDK: software development kit
- UHD: ultra-high-definition
- VR: Virtual reality

7. <u>Introduction</u>

The human visual system is highly developed for detecting objects, with faces being particularly salient objects. Humans have a notable ability to: rapidly recognize familiar faces (Visconti di Oleggio Castello & Gobbini, 2015); to read emotions and social cues from facial expressions (Brecht & Freiwald, 2012); and to even 'see' faces in images in which no faces are present (Liu et al., 2014). Furthermore, impaired facial recognition – referred to as face blindness or prosopagnosia (Corrow et al., 2016) – can result in a significant social handicap.

A host of human studies that monitored eye movements during tasks relating to face perception have revealed that face images are actively scanned during 'natural' free viewing conditions. Early gaze tracking experiments showed distinctive eye movements patterns and fixational targets when participants were asked to view faces compared to when they were required to do other perceptual tasks (Yarbus, 1967). The role of eye movements during face viewing has been the focus of many subsequent studies (Hsiao, 2010). In brief, it has been shown that during free viewing (i.e. 'active' viewing, while eye movements are allowed) of both familiar and unfamiliar faces, the visual axis, i.e., gaze, is predominantly focused on a subset of 'internal' facial features including the eyes, the nose, and the mouth (Stacey et al., 2005; van Belle, Ramon, et al., 2010). This leads to a commonly described 'T-shape' in the heat maps produced from these scanning eye movements, which is observed during both face learning and recognition paradigms.

Despite the progress listed above, and the fact that eye movements made during face viewing are generally thought to be functional for both learning and recognition, the extent to which they are absolutely required for these tasks, and the extent to which they serve similar or different roles in different tasks has been poorly explored. Indeed, reported results vary greatly between the studies in the literature.

8. Review of the relevant literature

a. Eye movements during face processing

While my project will focus on roles for saccadic eye movements during a recognition paradigm, such eye movements likely serve additional roles in face perception other than for aiding with recognition. Such potential roles include facilitating speech comprehension (Buchan et al., 2007) and attention (by following a speaker's gaze), reading others' emotional states (Peterson & Eckstein, 2012; Schurgin et al., 2014), and the social cues provided in facial expressions (Brecht & Freiwald, 2012), making judgement about personality (Kanan et al., 2015), and more. Clearly, eye movements serve many additional roles beyond recognition in face viewing, but these will not be explored in the present work.

b. Eye movement-dependent recognition of faces

Eye movements made during face viewing are believed to be useful for recognizing familiar faces. Stereotypical patterns seem to arise during recognition-related tasks. In these studies of face recognition, it has been found that different people view the 'internal' features of the faces by utilizing different strategies (defined as 'holistic' versus 'analytic', as outlined below). It has been suggested that individual differences in scanning strategies can lead to differences in recognition performance, and that the distribution of the fixations on the faces differs depending on the observer's degree of familiarity with the faces. This material is reviewed below:

i. Features observed during identification

From the earliest experiments involving recordings of eye movements during free viewing of faces, it was evident that participants look at specific facial features, mostly at the eyes, the lips and the nose (Yarbus, 1967). Since these viewing patterns were specific to face viewing (e.g.

people do not make stereotypic 'T-shape' scanning patterns when viewing most other types of images), many studies have examined whether face-specific gaze patterns are so stereotypical because they are functional, perhaps by permitting the completion of face-processing tasks – such as identification of familiar faces.

One study developed a method to isolate the parts of the face functional for a task (Gosselin & Schyns, 2001). At each trial, an image is shown via a random number of apertures ('Bubbles') of different sizes. The subset of Bubbles allowing correct identification of the face is determined, to create a 'diagnostic' mask. In the study, participants had to learn faces, and then identify them under the Bubbles technique using decomposition at six bands of spatial frequencies. The eyes and the corner of the mouth were the most useful information at the finest spatial frequency, then, at the next band, the eyes, the nose, and the mouth were the most useful. At the third decomposition band, the eyes, the nose, the mouth, and the chin together were the most meaningful, but not independently. The researchers were worried that the results might be due to atypical strategies related to the specific task, due to: presenting information in a sparse way, not setting a limit on the presentation time (unlimited – until a response was received), and restricting the number of faces. So, they presented new images that were filtered via the 'diagnostic' and 'undiagnostic' masks to participants under shorter durations of time. They found that participants performed equally well in recognizing the original faces and the faces filtered with the 'diagnostic' masks, and performed worse when they were shown faces filtered with 'nondiagnostic' masks. Thus, it seems that for recognition, the eyes are primarily the feature that is important at high-resolution.

While the original work by Yarbus (1967) showed that the scan patterns made while looking at the same image could be widely altered by the context (for example, by being given different instructions), these findings were not specific to face processing. Similarly, another study set out to compare the gaze patterns made under three different tasks specific to viewing faces, hypothesizing that the gaze patterns would differ across the tasks (Peterson & Eckstein, 2012). Limiting the analysis to the first fixation, they found that people looked at the point on the bridge of the nose that is aligned between the eyes when they had to determine the face's identity. In comparison, they fixated a lower point on the bridge of the nose for gender identification, and even lower, almost at the tip of the nose, for emotion identification. Then, they validated the functional relevance of these fixation points by comparing task accuracy when participants look at the preferred fixation point (the eyes or the nose) compared to when they fixate on other points (the forehead, the eyes, the nose, or the mouth). Finally, they compared those found preferred fixation points to different models, and they found that the only model to mimic the found preference was a Bayesian observer with a simulated decreasing contrast sensitivity in the periphery. They interpreted that the 'just below the eyes' area was the best location to foveate on since it permits observers to obtain the most information possible about important regions surrounding the fixation point that fell in the peripheral visual field.

Although most studies have turned their attention to the features mentioned above, mainly the eyes and the mouth, another study emphasized the role of eyebrows in face recognition (Sadr et al., 2003). Participants viewed familiar faces that were intact, or for which either the eyes or the eyebrows had been removed. They were tasked to press on a button as soon as they recognized the presented face, and to identify the celebrity seen by name. Interestingly, recognition performance dropped significantly more when the eyebrows were absent compared to when the eyes were absent. This result was surprising because it could not be explained by low-level image differences, i.e. the absolute and squared differences between the faces lacking eyebrows and the original faces were significantly smaller than those between the faces lacking eyes and the original faces.

ii. ‘Analytic’ and ‘holistic’ viewing patterns

Generally, there are two diverging ideas regarding the types of scanning patterns that allow recognition of faces. Some studies argue that faces are scanned ‘analytically’, whereas others suggest they are processed in a ‘holistic’ manner. Although the exact definitions of the two pattern types somewhat vary across the literature, what is generally meant by an analytic scanning pattern is that the local facial features are scanned, one after the other. Here the idea is that a face is recognized via information pooling across the multiple individual features, but the exact arrangement of the features (the distance between them, their placement, etc.) does not matter.

Holistic scanning, on the other hand, generally means that all features of a face are processed in a configural way (meaning that two or more features are processed at once instead of independently), or at least that the spatial relationships between certain facial features are essential for its recognition (Tanaka & Simonyi, 2016). Behavioural studies have provided indirect evidence for holistic processing, for example, by showing that recognition of facial features depends on the relative position and identity of the facial features (Tanaka & Sengco, 1997). By comparison, others have provided evidence that this is mostly not the case when recognizing familiar faces, and argue against such configurational processing (Burton et al., 2015). Indeed, the recognition of familiar faces can occur despite configural modifications, can be harmed by non-configural changes, and, finally, in studies that separate face ‘shape’ from face ‘texture’, recognition seems to be heavily dependent on the shape-free textures rather than the texture-free spatial relationships between features (Burton et al., 2015).

Some studies have tried to elucidate the origins of the two different types of scanning patterns, analytic and holistic, that were identified during face viewing. A few have investigated a potential link between an individual’s cultural background and scanning pattern type used, but no result seems conclusive. Some argue for a clear link between cultural background and scanning pattern (Kelly et al., 2011; Miellet et al., 2013), whereas others have pointed out different results. One study used hidden Markov models (HMMs) to classify East Asian participants' eye movements as holistic versus analytic and found individual differences in the preferred pattern (Chuk et al., 2014). However, these individual differences had no effect on recognition accuracy. Another study compared the type of patterns preferred by East Asian (Chinese) and Caucasian (Australian) participants, and, although they did not find a significantly different preference between the two groups, they found that participants using analytic patterns with a left-eye bias performed better than other participants (Chuk, Crookes, et al., 2017). Another study even found that the location that was gazed at initially does not vary between cultures, except for some small differences horizontally (Or et al., 2015). In fact, the first fixation reflects the optimal fixation locations on the observed face: participants looked at the point between the eyes and the nose, fixating slightly higher when observing Asian faces compared to Caucasian faces, regardless of the participants' own cultural background.

Others have investigated the contribution of the two brain hemispheres to each type of scanning pattern. While the precise experimental tasks and results varied across the studies, there is agreement that the right hemisphere is mainly responsible for holistic processing, whereas the left hemisphere relies on analytic processing. Studies used upright and inverted faces, either whole faces in separate visual fields (Cattaneo et al., 2014; Hillger & Koenig, 1991), or chimeric faces (Harrison & Strother, 2020), as they permit to test if the processing is holistic or not. These studies generally found a right hemisphere superiority for upright faces but no superiority for inverted faces, this latter finding being a sign of holistic processing. A study using composite faces, stimuli which are known to lead to decreases in performance under holistic processing, also found the right hemisphere to be more responsible for holistic processing of familiar faces than the left hemisphere (Ramon & Rossion, 2012).

Furthermore, many studies have compared analytic and holistic face viewing patterns, trying to find if either pattern confers an advantage for recognition performance. In one study, eye movements made by participants were analysed using HMMs in order to study the effect of the pattern used on the task performance (Chuk et al., 2014). The HMM analysis was set up to cluster the participants into two groups, which were then manually labelled as holistic or analytic, and no difference was found in the performance accuracy of the two groups. In another study, this issue was approached by comparing the scan pattern types across learning and recognition tasks to understand how a difference or similarity between the two tasks relates to recognition accuracy (Chuk, Chan, et al., 2017). The authors thought it was possible that using one specific pattern type out of the two during learning of new faces might subsequently facilitate recognition of those learned faces, or perhaps that the stability or similarity of the scan patterns across learning and recognition of a same face might affect task performance. They found that holistic patterns were employed slightly more for learning than recognition and that analytic patterns led to slightly better recognition performance. Additionally, about half of the participants did not use the same patterns between learning and recognition, but the extent of pattern similarity or dissimilarity across learning and recognition was not related to task performance. Finally, results from another study indicate that normal observers use holistic processing to recognize faces, whereas individuals with prosopagnosia lack holistic face perception and use analytic processing for this purpose (Van Belle, De Graef, et al., 2010).

Thus, the two scanning patterns, holistic and analytic, remain poorly defined. Consequently, despite significant work on this topic, it is unclear what exact role the two patterns play in face recognition and how these patterns link to task performance. However, what is clear is that people are capable to employ both these strategies in different circumstances.

iii. Viewing 'familiar' versus 'unfamiliar' faces

In our daily life, we encounter faces that range widely in familiarity. As mentioned in the Introduction, the 'internal' features of the faces (e.g. eyes, nose, mouth) are predominantly looked at during free viewing more than the 'external' parts, and this is the case regardless of face familiarity (Stacey et al., 2005; van Belle, Ramon, et al., 2010). Some studies have however contrasted the eye movements made when participants were tasked to recognize faces of different levels of familiarity, or to tell apart 'familiar' faces from 'unfamiliar' faces, and these studies have found varying results. Interestingly, as detailed below, some differences are noted in terms of the proportion of time viewing different features, as well as the number of fixations made on them.

Stacey et al. (2005) compared the eye movements made across three different tasks, testing for differential sampling of internal features when faces of various familiarity levels were looked at. In one of the tasks, participants were presented with two images simultaneously and they were asked to indicate whether the two images correspond to a same face or not. The images were of uncropped faces belonging either to unfamiliar individuals (i.e. faces that had never been seen before by the participant), or familiar individuals (i.e. faces taken from television and magazines that the participant had been exposed to previously). When analysing the eye movements made during this task only, there was more sampling of internal features rather than external features for familiar faces compared to unfamiliar faces. No differences were observed during the other two tasks, in which participants were shown one image at a time and had to indicate familiarity ('yes' or 'no') to the face seen. Nevertheless, other differences have been observed when images of faces were cropped to keep the internal facial features only, such as in the studies detailed below.

Another study recorded the eye movements made by participants when they had to tell apart familiar famous faces from newly learned faces (Barton et al., 2006). When viewing newly learned faces, participants looked at the upper face region (including the eyes) more than when they were viewing already familiar faces. In addition, the global scanning sequences made when viewing new faces were more stereotyped, whereas the sequences were more idiosyncratic for famous faces. The authors interpreted this finding as an influence of previous experience viewing familiar faces such that the participants had internal representations of those faces.

Taking another approach, (van Belle, Ramon, et al., 2010) looked at the differences in eye movements not on the level of trials, but rather on a fixation-by-fixation basis during a task. In this study, the participants had to identify images of personally familiar classmates from unfamiliar people. Although the first fixation was essentially the same, small differences became increasingly visible after the second fixation until the last fixation before a response (familiar or unfamiliar) was received. Certain individual facial features (the right eye and the mouth) were fixated more for familiar faces; whereas for unfamiliar faces, the central area between the eyes, which as the center of mass of the face represents the most salient region of the face, was fixated the most.

Thus, while participants seem to make more idiosyncratic eye movements towards the internal facial features when observing familiar versus unfamiliar faces, it is still unclear what the exact differences are, and if and how these differences are functional.

Therefore, despite much research examining roles for eye movements in face recognition, it is still unclear what type of viewing strategy is most useful for recognition, and if the viewing scan paths made are functional. It is possible that eye movements mostly serve to bring our gaze to a singular position that is most informative for recognition, as most studies cited above have found participants to look at internal facial features of familiar faces the most. As such, it remains unclear to what extent eye movements are required or not required to recognize familiar faces.

c. Eye movement-independent recognition of faces

Despite the abundance of studies examining roles for eye movements in face recognition, several results question the necessity of eye movements in face recognition. For instance, in one study, faces were presented for a short duration which allowed participants to only make one or two saccades (Peterson & Eckstein, 2013). Still, this paradigm resulted in high recognition success rates. Another study found that recognition performance saturated when the participants were allowed to make two fixations at different locations on a face, with additional fixations at different locations not increasing task performance (Hsiao & Cottrell, 2008). Finally, while not explicitly related to recognizing one specific face from another, it has been shown that humans can recognize face from non-face images on very short timescales – on the order of 140 milliseconds (ms) – which would only permit them to make a single saccade at maximum (Crouzet et al., 2010).

Additionally, some studies have related the eye movement patterns made during free viewing to recognition performance when fixation was enforced at different parts of the face. The preferred regions observed when faces were freely viewed were found to correspond with points for which recognition performance is optimal under forced fixation (Peterson & Eckstein, 2013), and with the parts of the face that elicit stronger neural face discrimination responses during forced fixation (Stacchi et al., 2019). For both studies, the location of the preferred fixation was highly idiosyncratic across participants, meaning that the effects observed on the recognition performance and the neural responses was independent of the actual location preferred.

d. Learning of faces

i. Eye movement-dependent learning of faces

Beyond the roles cited thus far, eye movements seem to significantly aid in learning new faces. Although many studies have addressed issues related to face familiarization (Johnston & Edmonds, 2009), only a few have studied the contribution of eye movements to this process. As detailed below, there is evidence that recognition performance can be enhanced by free viewing faces during learning, compared to fixating on a unique location (generally, between the eyes).

In one study in particular, participants were required to ‘learn’ previously unfamiliar faces for ten seconds, after which they were shown these learned faces along with new ones, and they were asked to indicate whether each face was familiar. The faces were 10 degrees in height (size from forehead to chin), and were shown one at a time, in a random order (Henderson et al., 2005). Only one single view was presented per face to be learned, and this same view was shown during the testing phases. One group of participants was allowed to make eye movements while learning the face images whereas the other group was required to fixate on a dot while learning the same images. Although participants who fixated during learning performed above chance, those who were allowed to freely view the faces performed significantly better. This suggests that eye movements are indeed functional during both learning and recognition, but that, similar to face recognition, many details remain unclear regarding their role in learning of faces.

ii. Eye movements during learning versus recognition of faces

Since there is evidence that eye movements are important for both learning and recognizing faces, it follows that looking at the differences in eye movements between these two types of tasks or, at how eye movements change with increased familiarity to faces, might help elucidate how gaze patterns are functional for learning faces. Below, I discuss some studies that have looked at the differences under those contexts and have generally found that making eye movements widely spread all over a certain face becomes less necessary with increased exposure to this same face.

One study analysed how patterns of eye movements changed during the process of learning faces (Heisz & Shore, 2008). Participants learned thirty new faces during three separate sessions, and, at each session, had to correctly identify the newly and the previously learned faces. As such, they were exposed to faces from the first session more times than those from subsequent sessions, and so on. During a fourth session, they were tested in recall and recognition tasks. As a face became more and more familiar, the number of fixations made and the nature of the eye movement patterns changed. During both types of tests, participants made fewer fixations on the face; and during recall tasks, they fixated more on the eyes, and less on the rest of the facial features.

These results are in agreement with another face learning study which aimed to contrast the rapid 'two-fixation' recognition (mentioned above, see Hsiao and Cottrell (2008)) to the process of learning faces (Arizpe et al., 2019). They found that providing longer times of observation during the learning stage improved later recognition, but that providing longer times during the recognition stage did not improve recognition. They concluded that, during learning, eye movements serve the purpose of integrating information from several facial features together to create an internal holistic representation of the learned face, to then permit rapid recognition of faces, even when only a limited number of fixations is allowed for this task.

Finally, one of the studies which was already mentioned above examined the pattern of eye movements made by the same research participants during face learning and recognition (Chuk, Chan, et al., 2017). The researchers found mixed results: both pattern types, analytic and holistic were used for the two tasks, and participants did not always use the same pattern type across tasks.

Some participants always used holistic scan patterns, some always used analytic patterns, and others switched patterns between learning and recognition. In addition, keeping patterns across the two tasks versus switching between them did not affect performance, meaning that the exact impact of a given scan strategy for a given person remained unclear.

e. Limitations of previous studies

i. Small presentation sizes

In most studies discussed above, the stimuli presented to the participants subtended about 5 to 15 degrees in height (see: Butler et al. (2010); Henderson et al. (2005); Stacey et al. (2005); van Belle, Ramon, et al. (2010), and more). Studies suggest that the macula has a diameter of 5 to 17 degrees (Strasburger et al., 2011), meaning that, for a central fixation point on the face, the entire face would fall within the macula. Thus, this might have been one of the contributing factors to the conflicting results noted in the literature. Since the faces were small enough to be seen completely at high-resolution even under forced fixation, it is possible that eye movements were not necessary for the completion of previous task paradigms, which might explain the lack of definition of clear roles for eye movements in both learning and recognition of faces.

ii. Natural variability of faces

One additional element that remains overlooked in the field of face processing is that familiar faces are naturally seen under a variety of different contexts. Indeed, in real-life situations, someone’s identity remains stable even though local elements are naturally variable, due to internal (facial expressions and movements, aging, etc.) and external factors (viewing angle, lighting conditions, etc.). However, these factors are controlled for in the experimental designs typically used in studies: images are luminance-matched, taken under the same lighting conditions and the same angle, external features are cropped, and so on, and participants are often expected to learn or identify a small number of faces (about ten) for a large number of repeated trials. Then, rather than performing face learning and recognition, participants might rely on details specific to the pictures themselves, and treat the tasks as image-matching problems, as shown below.

One study in particular highlighted the importance of natural variability of faces. Twenty images of two celebrities were shown to two groups of participants: one group was familiar with the celebrities and another group that was not (Jenkins et al., 2011). The participants were asked to group the images by identity. The latter group performed rather poorly, identifying about seven to eight identities on average rather than two, the correct number of identities, as the former group did. This shows that true familiarity with a person's face relies essentially on multiple exposures to that face, to learn how it varies under different viewing conditions.

Results from additional studies have also pointed out the limitations that result from heavily controlling the images used for recognition and learning tasks. For example, one study assessed the effect of using more 'natural' conditions during such tasks (Butler et al., 2010), by employing the widely used Bubbles technique (Gosselin & Schyns, 2001) and comparing their results to studies in which ten newly learned faces had to be identified (Caldara et al., 2005; Schyns et al., 2002). Butler et al. (2010) asked participants to verbally identify 210 images of celebrities when random parts of the faces were visible under Gaussian apertures of different spatial frequencies (i.e. Bubbles). By using different numbers of Bubbles, the researchers assessed which facial areas are necessary for recognition. Although their results were similar to the studies using newly learned faces, they observed differences in specific spatial frequency tuning peaks, which suggested that participants relied more on specific image-based characteristics to recognize the newly learned faces compared to the familiar faces that have been seen numerous times under variable contexts.

The importance of exposure to multiple pictures of a same person in solidifying familiarity is also evident in learning paradigms. In one study, children and adults had to learn new faces by viewing one, three, or six images per identity (Matthews et al., 2018). Both groups were more accurate at face matching when they were exposed to multiple photos of the same person rather than only one image of the person, although adults performed more accurately than children during the test. It seems then that face learning benefits greatly from exposure to multiple unique images of a same identity (under different lighting conditions, from different angles, across time, etc.).

iii. Instructions provided

Finally, in our daily life, no explicit instructions are given to us while we are viewing faces, whereas in studies participants are given specific instructions and they subsequently adjust their gaze strategy according to those, as noted since some of the earliest gaze tracking studies (Yarbus, 1967). As this point is rather difficult to address, in the study below, we presented participants with task instructions.

9. Body of the thesis

a. Rationale for the study

Several studies have examined roles for eye movements in face recognition. However, as outlined above, the absolute requirement of eye movements in recognition paradigms remains poorly understood, as well as the functional characteristics of the eye movements made during these tasks. Thus, the goal of the project was to directly examine the requirement of eye movements in face recognition, and to study them under this task.

Furthermore, in most studies described in the literature, participants viewed relatively small faces which fell completely within the macula. Therefore, we wanted to see if eye movements become more important or requisite for bigger sizes of images, specifically when fixating on a single point does not result in much of the face being viewed with high-resolution central vision.

Due to the concerns regarding image-matching, participants were tested on more than one image per familiar person, to lessen their ability to image-match.

b. <u>Hypotheses and Aims</u>

1) Are eye movements absolutely required for discrimination of face familiarity?

Since previous findings concerning whether or not eye movements are involved in face recognition are mixed, I first aimed to examine their absolute requirement by looking at how the ability to discriminate familiar faces from unfamiliar faces under forced fixation changes across a large range of face sizes. In previous studies, most of the face image fell approximately within central vision, which might have impeded finding clear roles for eye movements, leading us to use a behavioural paradigm including various face sizes.

I hypothesized that familiar faces of a small size (i.e. sizes similar to those used in previous studies) can be discriminated from unfamiliar faces under forced fixation, but that discrimination becomes impossible when faces are large enough that some of the stereotypically viewed internal facial features (e.g. eyes, nose, mouth) fall outside of central vision.

2) Do eye movements provide an advantage for discrimination of face familiarity?

After defining a size range for which face familiarity can be discriminated under forced fixation, I wanted to examine if there was an advantage provided by eye movements, by comparing discrimination of familiar faces from unfamiliar faces under forced fixation versus free viewing conditions. Specially, I wanted to see if eye movements allow discrimination beyond the range of sizes determined in <u>Aim 1</u>. Then, I aimed to compare 'non-functional' eye movements (made while viewing faces of sizes not requiring eye movements for discrimination), to 'functional' ones (made while viewing faces of sizes outside of that range).

I hypothesized that as faces become very large, discrimination rates will decrease under forced fixation but remain high under free viewing, and that, in free viewing conditions, gaze *patterns will change as images get bigger, i.e. as eye movements become functional in discrimination. Furthermore, I hypothesized that, even under free viewing, familiar faces will be discriminated from unfamiliar faces until internal facial features fall out of the field of view (i.e. the oculomotor range during head-fixation), and render the task impossible to complete.*

c. Methods

1) Participants

Each participant gave written informed consent before taking part in this study, which was approved by the Neurosciences and Psychiatry panel of the McGill University Health Centre Research Ethics Board. After study completion, each participant was compensated with $20 CAD per one-hour session attended at the laboratory. Individuals aged 20-70 were recruited through advertisements posted on online social groups. Given previous evidence that performance in face recognition tasks varied based on both the sex of the observer and the face viewed (Lewin & Herlitz, 2002), male and female participants were recruited and presented with face stimuli from both sexes. Additionally, given the evidence that observers from different cultural backgrounds might use different scanning strategies to recognize a face (Caldara et al., 2010; Miellet et al., 2013), participants were recruited across various ethnicities and a broad age range. At the start of the first in-laboratory session, a Snellen chart test was used to ensure that individuals have normal or corrected vision (when wearing glasses or contact lenses) above 20/40 prior to participation. Interested individuals were tested with Ishihara plates, but those with colour vision deficiency (CVD) were nevertheless allowed to sign up in the study as stimuli were in grayscale.

2) End-of-study survey

Following participation in the study, each person filled the 20-Item Prosopagnosia Index questionnaire (Shah et al., 2015) and a background survey. The survey included the following items: year of birth, frequency at which they play video games (‘Never’, ‘A few times a year’, ‘A few times a month’, “Once a week’, or ‘More than once a week’ – this was done to control for possible issues using the joystick in our virtual reality system), sex and gender identities, ethnicity or ethnicities, and geographical area they spent their childhood in (age 0 to 10). Only the first two survey questions were mandatory, and it was explained to participants that all demographic information would only be used for analysis of the data collected, given previous evidence that these characteristics can affect task performance and face-scanning strategies.

3) Screening

After confirming initial eligibility, each participant received a survey listing 500 famous individuals. Participants had to indicate their degree of familiarity with each name listed on a scale from 0 to 10; where 0 corresponds to never having seen the individual before and 10 signifies that they would be able to recognize the individual as easily as they can recognize a family member. To accommodate for the broad age range and background of participants, the dataset included famous individuals from various generations, countries, ethnicities, and occupations (actors, athletes, comedians, fashion models, television personalities, politicians, singers, etc.). A personalized stimulus set was generated from this survey for each participant, containing two face images (Stimuli details are available below) for each of the 150-200 highest ranked individuals (ideally, individuals ranked as six or above). The choice to include two face images per familiar individual was to ensure participants were tested on recognition of the identity of the face seen instead of ability to match a specific memorized image on a pixel level, as faces naturally vary in appearance under different viewing conditions (Mike Burton, 2013). Participants who had not highly ranked a sufficient number of individuals could not to participate in the study.

4) Orientation session

Each participant came to the laboratory for an orientation session, typically lasting an hour and usually scheduled about two to four days preceding the first experiment session. The main purpose of this session was to verify that each participant could identify the individuals that they ranked high on familiarity. The 300-400 images (i.e. 2 images per each of their 150-200 top ranked names) from their personalized dataset were shown in a random order. The vertical height of the faces spanned 30° of visual angle. The stimuli were displayed using a VIVE-Pro Eye Virtual Reality (VR) headset (HTC Corporation) using a custom written VR environment developed in Unity (Unity Technologies). Participants were allowed to freely view each image by moving their eyes and/or head for a duration of four seconds, after which the face disappeared, and they had a short amount of time to verbally respond correctly. To participate in the study, a person had to successfully name or identify both images of 75 individuals out of the set of 150-200 individuals. In this session, identification criteria were flexible: participants could name a movie title or fictional character for an actor, the country of origin for a politician, or someone closely related to the person (parent, spouse, child, etc.). Simply stating an occupation (actor, politician, singer, etc.) was not considered sufficient identification.

5) Experiment sessions

Post-orientation session, the personalized set of 150-200 familiar people was filtered to keep only 75 highest-ranked familiar individuals that were properly identified. In the two weeks following, each participant went through four one-hour long experiment sessions.

6) Systems used

Table 1 and **Figure 1** summarize all the information listed in this section. Participants went through experiments over four separate sessions held on separate days, across a period of two weeks. Two different systems were used to run those experiment sessions. Each week, a person took part in one session using the one system, and another session using the other system, and the order of those sessions was randomized between participants. Each session begun with a calibration step that was repeated as needed to ensure proper recording of eye movements. The left eye was used for eye tracking. A chinrest was used to stabilize the participants' heads.

During two sessions, the visual stimuli were presented on a Samsung LU28E590DS 28-Inch UHD LED-Lit Monitor while eye movements were tracked using an EyeLink 1000 Plus (SR Research Ltd.) system (EL).

The Eyelink eye tracking experiments were written in MATLAB R2020b (MathWorks), using the Psychophysics Toolbox extension (Brainard, 1997; Mario Kleiner, 2007; Pelli, 1997). The EL set-up allows to record timestamps of button-press responses, as well as variables related to eye movements: the horizontal and vertical gaze positions on the monitor, the pupil size and validity (i.e. whether the eye is detected properly), and the eye blinks (partial or total occlusion). This system permitted for eye tracking at a sampling frequency of 2000 Hertz (Hz) with an accuracy of 0.25-0.50° and a trackable range of 60º horizontally and 40º vertically. A 9-point calibration system was used at the start of each session, followed by a validation procedure.

The EL system is more accurate than the VR system and has a higher sampling frequency; but this method of recording is only possible when the participant's head is stabilized using a chinrest. The VR system does not require head stabilization and allows to show virtual images of a wider range of sizes (importantly, allowing us to show much bigger sizes of faces) than the classic eye tracking system. For those reasons, we used both systems.

During the other two experiment sessions, the visual stimuli were presented in a different custom written VR environment while eye movements were tracked using the VR headset's built-in eye tracking, which a previous study has shown to be relatively well-suited for human eye tracking studies (Imaoka et al., 2020). Even though the experimental setup involves minimal body movements, some people might experience feelings of dizziness or nausea in VR. In such a case, participants were given the option to immediately terminate the experiment, to take a break, or to be removed from the study, but this did not happen in any of the sessions.

The virtual reality experiments were written in C# using Visual Studio 2017 (Microsoft) and run using a headset with built-in eye tracking via the SRanipal software development kit (SDK), which allows for eye tracking at a sampling frequency of 90Hz with an accuracy of 0.5-1°. This headset has a horizontal field of view of 110°. Beyond eye tracking, the system allows to record multiple relevant variables: gaze origin and direction, pupil position and size, eye openness, eye validity, as well as head movements and rotations. It also can record the timestamps of button-press responses on the controllers. At the start of each session, a built-in 5-point calibration system was used. This system has lower accuracy than the EL calibration system, and also accuracy fall off as eye movements become larger (i.e. reach the edge of the oculomotor range).

The VR system does not explicitly record the onscreen fixation location, so we had to set up a method to obtain the projection of the gaze vector onto to image. After running the study, we found an issue with this, and realized that only the first fixation was recorded for each trial. This was subsequently fixed, but the experiments were unfortunately run before this realisation, and the eye movements presented here were recorded only using the EL system.

7) Experiment process

During each session (i.e. on a single day), 300 images were presented in a randomized order, half of which were unfamiliar faces, and the rest corresponding to the images from the filtered dataset. Each unfamiliar image was seen only once during the study, whereas each familiar image was seen once per session. For each trial, the familiarity of the face as well as the size and trial type were randomized. The schematic in **Figure 1** provides summary information regarding trial types, face sizes, response collection, and eye tracking sampling rate. The images spanned 0.5 to 30 visual degrees vertically for the EL sessions, and 5 to 150 visual degrees for the VR sessions. The actual values tested were 0.5, 1, 2, 5, 10, 15, 30, 50, 70, 90, 110, 130, and 150°.

How a typical trial was run is demonstrated in **Figure 2**. The trial was self-initiated by the participant via a button press. A keyboard press was used during the EL sessions, and a hand-held controller button press during the VR sessions. Before image presentation, a warning tone played for one second, and a cue announcing the viewing condition followed. Next, a face was presented for four seconds, during which the participant could report recognition by pressing on the same button used for trial initiation without having to verbally identify the face. At each button press, a tone indicated response registration. Participants were instructed to not press the button if the face seen was unfamiliar to them (i.e. selecting 'unfamiliar' was not explicitly enforced with a different button). Immediately after a response, or, in the case where the face was not recognized, after the four seconds elapsed, the image disappeared, and the participant could press to initiate a new trial.

8) Trial types

The images were presented under three different viewing conditions, and the trial type was randomized across trials. Regardless of viewing condition, the images were centered on the point of the face that was equidistant from both eyes and aligned with the nose (see **Figure 2**, bottom-left image), as this was very close to the area most informative for face recognition (Peterson & Eckstein, 2012) while also being easily identified via automated face registration (described below). In the 'Fixation' condition, a fixation point of 0.75° radius needed to be fixated for 1 second before image presentation occurred. The participant then passively viewed the face while maintaining fixation throughout the trial, otherwise, the trial was immediately terminated. **Figure 3** represents the foveated images (i.e. blurred according to visual acuity falloff with eccentricity

(Geisler & Perry, 1998), using code written by Dr. Tong) that approximate how the faces likely appear as seen under forced fixation for the various sizes used in the experiment. As the viewed face gets larger than 70°, increasing portions of it fall outside of the field of view. One of these conditions, 'Head-Free', additionally allowed them to lift their head from the chinrest, a condition we kept out of interest to investigate whether head movements would allow improved performance in the task. This condition was only possible during the VR sessions, whereas the other free viewing condition, 'Free-View', was possible in both systems. At the start of the first session in each system, practice trials were run for the participant to get acquainted with the system used, the trial process, and the various viewing types, using unfamiliar faces that were not shown again.

9) Stimuli details

All images of familiar faces that were presented were unlicensed images obtained through an automated internet search using Selenium Webdriver, via image scraping code adapted from Bosler (2019). The pictures were filtered to select only high-resolution frontal-view images of faces without any distinguishable features, glasses, tattoos, or facial hair. The unfamiliar face images used were obtained from the FEI Face Database (Thomaz, 2005) and from the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office (Phillips et al., 2000; Phillips et al., 1998).

All external facial features were removed by cropping the faces around their outline using the machine-learning tool MediaPipe (Lugaresi et al., 2019). The images were rendered greyscale, luminance-matched using the SHINE toolbox (Willenbockel et al., 2010), and put over a grey background (RBG 127, 127, 127). Since the images were obtained from different sources, we wanted them to look qualitatively similar to one another. So, we ran Principal component analysis (PCA) (Pedregosa et al., 2012; van der Walt et al., 2014) on the image textures, followed by dimensionality reduction, to essentially discard the low-level variability across the images. In short, 76 facial features were defined on each face using the AI-based facial segmentation tool Stasm (Milborrow & Nicolls, 2014) and the machine-learning tool MediaPipe (Lugaresi et al., 2019). The features, which contain the face shape information, were aligned across all famous faces in the dataset on the point between the eyes and above the nose (the same point that was used for fixation during the experiment, see **Figure 2**, bottom-left image). The average location of each feature was calculated. Afterwards, each given face was morphed to match the shape population average by translating its features to the corresponding average locations. Dr. Tong wrote code to do this by thin plate spline warping, using the radial basis function: $\varphi(r) = r^2 \times \ln(r)$. PCA was subsequently run on the pixel values of the morphed images, which only contain texture information. Finally, the experimental stimuli were obtained by reconstructing each image texture using only the first 200 principal components (PCs), and then morphing these reduced-texture images back respectively into their original feature shapes using thin plate spline warping. **Figure 4** shows an image reconstructed with different numbers of PCs, to demonstrate the level of dimensionality reduction chosen – which was chosen such that the famous faces still looked highly recognizable, but that the image 'appearance' was more consistent across all photos, both familiar and unfamiliar, of the stimulus set.

d. Analysis of data and results

Analyses were performed using Python 3.9 (Python Software Foundation) in Spyder (Spyder project contributors) and MATLAB R2020b (MathWorks).

1) Demographic information

Thirty-nine people (twenty-five male; fourteen female) signed up to participate in this experiment. The sessions from three male participants were used to finalize the experimental design. Of the thirty-six remaining participants (twenty-two male; fourteen female), only fifteen (six male; nine female) succeeded in identifying 75 individuals twice each during the training session and proceeded to participate in the study. Thus, a higher percentage of female participants (64.29%) were able to fulfill the dataset filtering requirements during the orientation session than male participants (27.27%).

One male participant took part in three experiment sessions (two EL, one VR) only, due to technical difficulties with the VR system during the last session.

Demographic details about all but one of participants who completed the entirety of the study are available in **Table 2**, as one male participant did not fill the survey concerning demographic information. Only one of the participants had a PI-20 score above 65.

2) Trial filtering

'Fixation' trials were discarded if fixation was not maintained properly throughout the trial.

Before running the experiments, a maximum 'False alarm' rate across all trials was set to 10%, with the intention to exclude data coming from participants who exceeded this rate from the analysis. 'False alarm' was defined as the proportion of unfamiliar faces for which 'recognition' button presses were received. None of the participants exceeded the 10% limit, thus, all the recorded sessions were kept for analysis. On average, the rate was 4.97 ± 2.44 %, and **Table 3** presents the average performance rates across trials.

3) Response times

Our main interest in the study was task accuracy, so, although we report response times across the different trial types and face sizes in **Figure 5**, we did not perform any further analysis on these data. Response times appeared slightly shorter under ‘Fixation’ compared to the other two viewing conditions, perhaps because participants were aware that images would disappear if not properly fixated. This might have prompted them to click faster under this condition, whereas, in the other two, there was no penalty for taking more time for task completion. Conversely, the apparently longer time required for free viewing trials (‘Free-View’ and ‘Head-Free’) could be due to a small time cost associated with the motor movements required in those trials.

4) Measures of discrimination accuracy

In the following, a ‘recognition’ corresponds to providing a response via a button click when a face was shown onscreen. These responses can be quantified by measures of discrimination accuracy, which can be computed across experimental variables (e.g. systems used, participant sex, experiment days) to look for differences at these levels. A ‘Hit’ was defined as responding with recognition when a familiar face is presented, a ‘False alarm’ as responding with recognition when an unfamiliar face is presented, a ‘Correct rejection’ as the lack of response to an unfamiliar face, and a ‘Miss’ as the lack of response to a familiar face. All of these values are proportions of detections that range from 0.0 to 1.0. For example, for ‘Hits’, a value of 0.0 corresponds to not responding for any of the familiar faces that were shown, whereas a value of 1.0 corresponds to responding to all the familiar faces.

The design of the experiment is such that the participants only had to respond when a familiar face appeared onscreen, i.e. responses were not ‘symmetrically’ enforced. In other words, the participant only made an ‘active’ choice to push a button to indicate ‘familiar’ but could only make a ‘passive’ choice to not push any button to indicate ‘unfamiliar’. For this reason, we were able to calculate ‘Hit’ and ‘False alarm’ rates accurately, but the comparison to measures of ‘Correct rejections’ and ‘Misses’ may be biased due to the asymmetric nature of the experimental design. Nevertheless, the values of these measures across all trials are provided in **Table 3**.

The goal of the experiment was to characterize how recognition changes as face size increases under different viewing restrictions, to define sizes of face images that do and do not require saccadic eye movements for recognition. For this reason, we were more interested in the overall performance of participants in discriminating familiar faces from unfamiliar faces, i.e. the interaction between ‘Hits’ and ‘False alarms’, rather than those rates themselves. Still, the median values of these rates with their corresponding interquartile ranges are reported in **Figure 6**.

To look at discrimination performance, we opted for the non-parametric measures A’ and B” (bias) used in signal detection theory, due to the assumptions underlying the use of d’. Indeed, d' is a measure of sensitivity that assumes the distributions of the signals to discriminate are normal and have the same standard deviation (Stanislaw & Todorov, 1999). One measure of A’ and B” (calculations were made following Stanislaw and Todorov (1999)) was calculated per participant for each combination of trial type and face size using the corresponding mean values of ‘Hit’ and ‘False alarm’ rates per participant for each condition. These values are reported in **Figure 6**. A’ ranges from 0.5 (unfamiliar and familiar faces cannot be discriminated) to 1.0 (perfect discrimination of unfamiliar and familiar faces), while B” ranges from -1.0 (bias to providing a button response when shown ‘unfamiliar’ faces) to 1.0 (bias to providing a button response when shown ‘familiar’ faces), with a value of 0.0 corresponding to no bias in responding.

Prior to calculation A’ and B”, extreme values of ‘Hit’ and ‘False alarm’ rates were adjusted as suggested in Macmillan and Kaplan (1985), as it is the most commonly used measure. For each combination of participant, trial type, and face size, values of 0 were replaced with $\frac{0.5}{n}$; and values of 1 were replaced with $\frac{n-0.5}{n}$ where n is the number of trials used to calculate those values. In our design, if participants break fixation during the ‘Fixation’ condition, the trial is thrown away. Hence, for some combinations of participant, trial type, and face size, it was impossible to calculate one of the ‘Hit’ or ‘False alarm’ rates, and thus, values of A’ and B”. Additionally, for some combinations of trial types and face size, there was only one valid trial (i.e. fixation was maintained throughout the trial); such values were excluded from the analysis.

5) Psychometric functions

To test our hypotheses, we decided to make use of psychometric functions to look at the variation of A’ values when the face size is increased for each viewing condition, instead of comparing those values across the conditions for each face size separately. The curves were fitted using the following equation for a sigmoid with values ranging from 1.0 to 0.5:

$$y = 1 - \frac{1}{1 + e^{-k\times(x-x_0)}} \times \frac{1}{2}$$

where x represents the face size, x_0 is the ‘sensory threshold’ parameter (the inflection point of the curve), and k is the ‘scale’ parameter (the steepness of the curve). We chose this equation as values of A’ span 0.5 to 1.0 and because we assumed that discrimination worsens when face sizes are increased, regardless of trial type.

One psychometric function was fit per participant for each viewing condition, using the values of A' for each face size above 1 degree in visual angle, and R^2 values of the fits are reported in **Table 4**. Performance was at chance level for faces of 0.5 degrees in height, perhaps due to monitor resolution, or because faces become harder to recognize at such sizes. Those trials were excluded from the analysis, as they would have led to issues fitting the functions, and our main interest was participant performance for the larger face sizes. Estimates of the 'sensory threshold' and the 'scale' were obtained for each function. Then, per viewing condition, we calculated the median values of those parameters to obtain psychometric functions over all the participants.

Aim 1: Are eye movements absolutely required for discrimination of face familiarity?

To determine if eye movements are absolutely required for recognizing familiar faces, we first looked at the median psychometric curve for A' under the 'Fixation' trial type, which is shown in **Figure 7**. We had hypothesized that faces presented in larger sizes will not be recognized under such a condition. However, the 'sensory threshold' was on average 105.694° ± 9.04° SEM across participants, with a median value of 114.40° [Q1 = 112.27°, Q3 = 123.39°]. This parameter reflects the inflection point of the curve, i.e. the face size for which performance is just in between chance and perfect performance. This means that participants were still above chance level for faces of large sizes. In fact, referring to **Figure 3**, the totality of the face can be seen in the field of view for faces only up to 70° in height. Thus, under forced fixation, participants were able to recognize faces of sizes large enough that the internal facial features were well outside of the macula.

Aim 2: Do eye movements provide an advantage for discrimination of face familiarity?

To follow-up on this result, we wanted to see if saccadic eye movements provide an advantage for recognition of familiar faces – in other words, if they extend the size range whereby familiar and unfamiliar faces could be readily distinguished from each other. To this end, we compared the psychometric functions between the 'Fixation' and 'Free-View' trial types, as shown in **Figure 8**.

For the 'Free-View' fitted curves, the mean 'sensory threshold' across participants was 119.45° ± 4.18° SEM, with a median of 118.20° [Q1 = 112.80°, Q3 = 122.58°]. We compared this parameter to the one obtained under 'Fixation' using the Mann-Whitney U test and found no significant difference across the two trial types (statistic = 69.0, p-value = 0.44). Hence, eye movements alone did not expand the range of sizes for which familiar faces can be discriminated from unfamiliar faces.

As for the scale, the mean was 0.19 ± 0.073 SEM for 'Free-View' with a median of 0.10 [Q1 = 0.047, Q3 = 0.18], and 0.15 ± 0.047 for 'Fixation', with a median of 0.075 [Q1 = 0.046, Q3 = 0.20]. Again, the parameter values were compared by Mann-Whitney U test and were not statistically different from each other (statistic = 63.0, p-value = 0.31). This means that eye movements did not decrease the steepness of the psychometric curve for discrimination of face familiarity, i.e. that participants reached chance level at the same rate for both viewing conditions.

<u>Additional analyses beyond defined aims:</u>

In <u>Aim 2</u>, we allowed participants to make free eye movements while their head was stabilized with a chin rest. This limited them to make eye movements only within the oculomotor range, a range under which we found that eye movements did not confer a discrimination advantage compared to viewing the faces under forced fixation. Following our observations, we wanted to see if removing the chin rest restriction (i.e. permitting participants to accompany their eye movements with head movements) allowed participants to discriminate between familiar and unfamiliar faces over a wider range of face sizes.

In **Figure 9**, we compare the parameters of the psychometric curve for the 'Head-Free' condition against those of the other viewing conditions. The mean 'threshold' was 141.77° ± 7.55° SEM with a median of 138.41° [Q1 = 131.31°, Q3 = 152.73°], and the mean 'scale' was 0.081 ± 0.034 SEM with a median of 0.029 [Q1= 0.020, Q3 = 0.087]. The Kruskal-Wallis test was used to compare the parameter values across the trial types. The 'scales' were not found to be significantly different across the trial types (statistic = 4.94, p-value = 0.084), meaning that the participants reached chance level at the same rate of decrease in A' for all the viewing conditions. However, the 'sensory thresholds' were found to be different across the conditions (statistic = 9.18, p-value = 0.010). Pairwise post-hoc Dunn's tests revealed that the 'threshold' from the 'Head-Free' viewing condition was statistically different from the one under the 'Fixation' (p-value = 0.0098) and 'Free-View' trial types (p-value = 0.029).

We thought the difference found in the 'sensory thresholds' might be a result of participants getting better at the task and/or using a new strategy for completion of 'Head-Free' trials, as they took part in multiple experiment sessions. To investigate this, we fit psychometric curves first on the pooled the data from the first week (first sessions in EL and VR) and then on the data from the second week (second sessions in EL and VR). These curves are shown in **Figure 10**. Using the Wilcoxon signed-rank test, we compared the 'sensory threshold' parameter across the two days, for each viewing condition separately. We found that the 'sensory threshold' parameter was significantly different between the first week and the second week of the experiments for both 'Free-View' (statistic = 11.0, p-value = 0.027) and 'Head-Free' trials (statistic = 5.0, p-value = 0.039). Indeed, the threshold increased by a mean of 14.27° ± 4.84 SEM for 'Free-View' and 29.97° ± 11.40 SEM for 'Head-Free'. Finally, we ran the Kruskal-Wallis test separately on first week data versus second week data to compare the 'sensory threshold' parameter values across the trial types, to see if the advantage found in 'Head-Free' trials was only observed the second week, but this was not the case (Week 1: statistic = 4.92, p-value = 0.086; Week 2: statistic = 3.94, p-value = 0.14). Together, these analyses suggest that although some learning permitted for an improvement for both 'Free-View' and 'Head-Free' trials during the second week of experiment sessions, the better performance observed during 'Head-Free' trials was not a result of task learning or training, but rather of allowing participants to make head movements.

6) False alarm rates

We found that 'False alarm' rates seemed rather different in 'Fixation' compared to the other viewing conditions, mainly across the range of face sizes from 5 to 50° in height, which can be seen in **Figure 6**. We first wondered if this difference could be due to the experimental set-up used, as the difference was mainly evident over a range of face sizes mostly shared across both experimental set-ups, or if it could be due to the sex of the participants, given previous studies that this variable can affect task performance. The discrimination measures were plotted separately for EL and VR sessions in **Figure 11**, and separately for male and female participants in **Figure 12**.

We thought that the difference in 'False alarm' rate between 'Fixation' and the other viewing conditions was mainly attributable to the male participants, and we set out to compare their data to that of female participants. In **Figure 13**, we compare these rates between male and female participants across the viewing conditions. We pooled the values across participant sex for faces ranging 5 to 50° in height and compared the values between sexes for each trial type using Mann-Whitney U tests, followed by a False discovery rate (FDR) correction of p-values across the trial types. We found that, for 'Fixation' trials, male participants (n = 6) had statistically higher 'False alarm' rates than female participants (n = 9) (FDR-corrected p-value = 0.026), but that the 'False alarm' rates of two groups were not different under the other trial types.

7) Eye movements

Aim 2: Do eye movements provide an advantage for discrimination of face familiarity?

We then analysed the sequences of eye movements made when viewing familiar faces, to see if any differences arise across face sizes. Only free viewing eye movement recordings were of interest in this section. Instances during which the eyes partially or totally closed were filtered out of the eye movement analysis. For the eye movement sequences recorded during the EL system, the different fixations clusters and saccades were defined for each trial. First, velocity was calculated between each pair of consecutive recorded time points. Then, the different fixations were identified using a Density-Based Spatial Clustering of Applications with Noise (DBSCAN) data clustering algorithm (Pedregosa et al., 2012) on the velocity and the vertical and horizontal fixation positions, with the outliers corresponding to the saccades. To generalize the analysis of the free viewing gaze trajectories across different faces (whose facial features were located in relatively different positions) and to allow easy comparison of gaze trajectories, all of the recorded trajectories were morphed onto the average face shape using thin plate spline warping. Examples of eye movement sequences morphed onto the average face with the corresponding identified saccades and fixations are shown in **Figure 14**.

Then, we wanted to look for differences in the time spent observing different areas of the image across face sizes. For this analysis, we used the full eye movement sequences to produce heat maps, instead of the identified clusters. First, eye movement sequences that were shorter than 500ms were discarded, and the rest of the sequences were trimmed to only keep the first 500ms. We chose this value because and there is evidence that the first two fixations are the most instrumental for face recognition (Hsiao & Cottrell, 2008), and, although fixation durations vary, they were found to be on average 261.78ms in a study involving face recognition (Chuk et al., 2014), and around 250ms long for other tasks involving scanning images, such as scene perception and visual search (Rayner, 2009). After that, regions of interest (ROIs) were defined by partitioning each face into a Voronoi diagram using the locations of the identified facial features (the same features that were used to prepare the experiment stimuli). This type of diagram is generated from pairs of coordinates (here, of the facial features) by using each pair to define a cell of the diagram, which will contain the points of the image that are closer to this coordinate pair than to the rest of the pairs. The eyes were grouped as a single ROI. For 'Free-View' trials, for each face size, the amount of time (in ms) spent in each area was calculated. Since the ROIs have different surface areas, the amount of time spent in each ROI was divided by its corresponding surface area, and this density of time is plotted as a heat map in **Figure 15**. The absolute amount of time spent in each area was plotted in **Figure 16**. To ensure that our results are not due to a biased definition of ROIs, this same analysis was repeated by binning the face area into 400 squares of equal sizes and plotting the amount of time spent in each of the squares, as shown in the heat map in **Figure 17**.

We used those heat maps to quantify the differences in the amount of time participants looked at the binned areas across the face sizes. First, we excluded the heat maps produced when observing faces smaller than 5° from this analysis, because of the lack of a clear stereotypic gaze pattern ('T-shape') and because of the limitations of the tracker in terms of recording accuracy (see **Table 1**). Then, to be able to make this comparison, we normalized the time spent in each binned area to the size of the face viewed (in metric units). Then, for each of the 400 binned areas, separately, a one-way repeated measures Analysis of variance (ANOVA) was run on this normalized amount of time, with face size (5-30° in height) as the independent variable, and participants as the subject variable. This analysis was followed by FDR correction, and the corrected p-values are shown in the left heat map of **Figure 18**. The middle heat map in the same figure displays the twenty binned areas where a significant difference was observed across sizes.

Given that a difference in dwell time was observed across face sizes for some binned areas, we decided to see if there was a relationship between face size and time spent looking at a binned area. For the twenty binned areas identified in the above analysis, we calculated the Spearman rank-order correlation coefficient of time against face size. In the right heat map of **Figure 18**, we plot the correlations that were significantly different from 1000 shuffles. Interestingly, we find a center-surround difference in the correlations: with negative correlations at the center and positive ones at the surround, suggesting that for smaller face sizes, participants look at the center more than the surround, whereas for bigger sizes, they view the surround more than the center.

10. Discussion

a. Study themes

Although various studies looked into the role of eye movements in face recognition, the extent to which scan paths made when looking at a face are functional for its recognition remains debated. There is evidence that humans can recognize faces with only very few eye movements allowed or even without eye movements. The project presented in this document was thus elaborated to see: first, under forced fixation, over which range of image sizes discrimination between familiar and unfamiliar faces is possible; second, if allowing eye movements under a head-fixed condition would extend the range of sizes whereby familiarity could be discriminated; third, if head-free conditions could further enhance task performance. Our results are in agreement with a subset of previous studies, in that we find that, for the task we designed, eye movements might not necessarily play a role in recognition of familiar faces, but that participants nevertheless make stereotypic 'T-shape' eye movements when viewing faces.

We designed an experiment that consisted in presenting participants with a mix of familiar and unfamiliar faces of different sizes under different restrictions of viewing, only some of which allowed eye movements. As a measure of performance, we focused on discriminability of familiar faces from unfamiliar faces rather than recognition of familiar faces. To be able to quantify how performance varies across two main conditions of interest (forced fixation onto a point equidistant from both eyes and aligned with the nose, versus free viewing of the face) as the presentation size is increased, participants had to indicate familiarity to the faces seen. An additional viewing condition allowed participants to make head movements along with free eye movements.

Two main hypotheses were driving this experimental design. First, that familiar and unfamiliar faces can be discriminated under forced fixation as long as the internal facial features that are stereotypically viewed are located within central vision, after which the task becomes impossible under this viewing condition. Second, that beyond the face size for which this happens (and up to a certain 'threshold size' above which faces become impossible to recognize at all), eye movements become necessary to discriminate the familiarity of the faces, and their structure becomes more informative when they start playing an essential role for the task at hand.

Experiment sessions were recorded with fifteen participants, and measures of A' (an interaction of 'Hits' and 'False alarms') were calculated across the conditions of interest to evaluate task performance. Psychometric functions were constructed to examine how task performance changed as face sizes were increased, for each viewing condition separately.

Task performance stayed stable up to large face sizes (around 115°), even under forced fixation. This implies that eye movements were, in fact, not necessary to successfully discriminate familiar faces from unfamiliar faces, even for sizes for which the internal facial features fall out of central vision under fixation. In addition, this range extends beyond that seen in day-to-day life. Indeed, faces are about 18 cm from forehead to chin, and, if a person was nose-to-nose with another person, their faces would be separated by approximately 8 cm; meaning that the biggest visual angle that we (occasionally) see faces at subtends 97°.

In addition to this, allowing eye movements during head fixation did not permit participants to discriminate familiar faces from unfamiliar faces over a larger range of sizes, and did not change the rate of decrease in performance as faces get bigger. This means that eye movements do not provide an advantage for discrimination of face familiarity. Thus, although our eye movements center the most salient face information into central view (as can be seen by the 'T-shape' heat maps), it seems that this is not necessary for the task at hand, given that familiar faces could be discriminated from unfamiliar faces even under forced fixation. This means that high-resolution input for the salient facial features is not required to discriminate face familiarity, as they are located in the peripheral vision for bigger sizes of faces under forced fixation, and as increasing the presentation size gets rid of the high spatial frequency content in the images.

Then, we wanted to analyse the sequences of eye movements made when viewing familiar faces, to see if any differences arise across face sizes. The analysis showed that for faces of 5 to 30° in height, participants looked at the areas for different durations of time. Specifically, for smaller sizes, participants looked at the center longer than for bigger sizes, whereas for bigger sizes, they looked at the immediate surround of the central area more than for the smaller sizes. We speculate that this could be because, as the face gets bigger, some of the features are further away from the fovea, which requires participants to look around the center rather than at it directly, to gather more information and perform the task. However, no difference was observed in success rates between 'Fixation' and 'Free-View' trials, perhaps because participants were able to gather enough information to make a decision about recognition without requiring eye movements.

Beyond answering our hypotheses, we wanted to determine whether combining eye movements with head movements provided an additional perceptual advantage for the task. Surprisingly, this extended the range of sizes for which discrimination of face familiarity was possible. However, for all three viewing conditions, participants reached a chance level of A' at the same rate. Thus, the saccadic eye movements made during face viewing themselves were not essential for this discrimination task. In addition, the central spot of fixation (used in our experiment and shown in previous study to be the most useful location for face recognition) does not carry information itself, but rather is a convenient location to access all of the facial features.

Even though participants performed better on 'Free-View' and 'Head-Free' trials during the second week of the study, when looking at the data from the first or second week alone, we did not observe an effect of improved performance in 'Head-Free' compared to the other trial types. This means that the ability of the participants to train on the task and/or to learn a new strategy from session to session were not responsible for the effect seen in 'Head-Free' trials. Indeed, allowing head movements permitted participants to perform better at the task.

Finally, male participants made significantly more 'False alarms' under forced fixation compared to female participants, when viewing faces of 5 to 50° in height. Somewhat consistent with this finding, compared to women, a smaller percentage of men passed the initial session required to participate in the study. As suggested by the Spearman correlation heat maps observed across face sizes, eye movements could have permitted participants to make more accurate judgements concerning unfamiliar faces, perhaps by allowing them to sample more information around the face and reduce their uncertainty prior to making a decision. However, we do not have enough data to explore this option, as our study included only six male participants and nine female participants, and the effect could be specific to the individual participants recruited. A bigger sample size is probably needed to pull out a real effect of participant sex on task performance.

b. Limitations

As discussed in the Introduction, participants possibly rely on certain characteristic internal facial features to make a familiarity judgement in face recognition tasks, which introduces questions regarding the nature of such tasks: whether they truly are recognition experiments or are simply image-matching experiments. For this reason, our task involved the presentation of two images per familiar identity. Still, it is impossible to state with certainty that participants were employing one or the other method, since each image of a familiar individual was seen five times across the study. So, we believe that the conclusions made here relate more to discriminating familiar faces from unfamiliar faces, rather than recognition of faces.

In our chosen experimental design, participants clicked on a button when they saw a face that was familiar to them. In other words, each person had to ask themself: ‘Is this face familiar to me?’; and collect visual information regarding the face seen. Within four seconds, if they had collected enough information to exceed their own internal absolute threshold for ‘familiarity’, they clicked on the button. In any other situation, when they did not collect enough information (either because the face is unfamiliar of because the participants were unsure), they did not provide a response. Thus, to a certain extent, the choices that the participants had to make were asymmetric, unlike other more controlled paradigms usually used in face recognition studies, such as the Alternative forced choice (AFC). In an AFC paradigm, there are two physical buttons within participants’ reach, one for familiar faces and one for unfamiliar faces, and an answer must be submitted for each trial. In such a situation, the participant has to answer the following question ‘Is the face I see more familiar than unfamiliar or is it more unfamiliar than familiar?’. Opting for an AFC task gets rid of individual bias and variability in decision threshold, because participants are always performing a relative comparison between the amount of information collected in favour of ‘familiarity’ versus ‘unfamiliarity’. The difference in task design, as illustrated in **Figure 19**, creates difficulty in comparing our results to the literature. Indeed, our measures of ‘Hit’ and ‘False alarm’ rates were most likely affected by the experimental design and the way the task was approached by the participants (i.e. the internal question they ask themselves). As the ‘Correct rejection’ and ‘Miss’ rates may not be indicative of participant performance, these data were not analysed. In addition, for the trials involving the bigger face sizes, due to the lack of enforcement of a choice (i.e. the task was not an AFC), only a very small number of trials where familiar faces were presented were actually completed, meaning that a response was received. **Table 5** notes the detailed breakdown of the average number of trials with received responses and the average number of completed trials across participants, per experimental condition. The reason why only a small number of trials were completed for some experimental conditions might be that the threshold set internally for familiarity was very high, or that it was never reached (see **Figure 19**). So, it is hard to make any conclusions with certainty regarding participant performance for those trials, as the calculations of A' were certainly affected by this sample size.

If fixation was not maintained properly throughout a trial, then the trial would end and would be marked as incomplete and would be discarded from the analysis. We opted for this experimental design instead of making participants repeat the trials for which they broke fixation because we did not want specific images to be seen more times than others. However, two issues arise because of the way broken fixations were dealt with. First, there was a smaller number of recorded trials under forced fixation compared to the other two conditions, for which trials could not be marked as incomplete (see **Table 5**). The calculated values of A' for this viewing condition were affected by the small sample size, and our statistical comparisons across viewing conditions involved unequal sample sizes. Second, since participants were aware that when fixation was not maintained, faces disappeared, participants might have had the incentive to finish the trials as quickly as possible. Then, it is unclear if response times related more to the paradigm chosen and task training or were actually informative of participant performance, which is why we did not compare them between the forced fixation and the free viewing trials.

Finally, our design involved a third condition, 'Head-Free', for face sizes bigger than 5 degrees in height, for which participants were allowed to make both eye and head movements freely. There is a potential issue with our comparison of performance under this viewing condition to the other two conditions: for sizes for which the face is fully visible within the field of view (i.e. it can be seen without the need for head movements), this condition and 'Free-View', the head-fixed condition, are essentially the same. Still, we believe the comparisons we made are valid, as they mainly pertained to larger face sizes for which facial features fall out of the field of view.

c. Future directions

It is clear, after discussing the limitations of the experiment, that the validity of the findings can be confirmed both by redesigning the study and running it differently, to deal with certain technical difficulties; as well as running the study with more participants and/or sessions and less testing conditions, to gain statistical power. Alternatively, the data presented here could be upsampled, and the A' calculations and consequent statistical analyses could be run again to ensure cleaner results and to interpret the data more clearly.

Additionally, it might be worth exploring if eye movements truly lower the chance of falsely recognizing unfamiliar faces. More experiments are needed to reach real conclusions, but, for a subset of participants, this seemed to be the case. Potentially, eye movements provide this advantage by allowing to sample more information about the face observed to make a judgement on familiarity. Another possibility is that under conditions of limited information, which is the case for forced fixation where most of the face is outside of central vision, participants cannot accurately template-match the face viewed, and conflate unfamiliar faces with stored templates of faces, which, by definition, can only be familiar.

Although we did not find a face size absolutely requiring eye movements for discrimination of face familiarity, it remains possible that eye movements are essential for other tasks, such as learning. Indeed, when Henderson et al. (2005) examined this, they found eye movements helped, but forced fixation led to above chance performance. However, participants learned identities by viewing one 10° image per identity, which fell completely within the macula. The paradigm we use could be employed to examine if eye movements become more important for learning as faces are large enough that a significant portion of them falls outside of the macula for each fixation, and when multiple images are presented per learned identity. The goal would be to contrast situations for which eye movements are absolutely required for face learning and recognition from those for which they are required to learning but not recognition, to further investigate the role of eye movements. The hypothesis that we would test is that even for natural sizes of faces, i.e. sizes seen in daily life, learning becomes impossible without eye movements, and that the viewing patterns made under free viewing change as eye movements become necessary for learning.

Finally, free viewing familiar faces with the purpose to discriminate them from unfamiliar faces produced ‘T-shape’ heat maps (see **Figure 17**, 5-30° faces), even though the eye movements did not enhance performance. Thus, we would like to investigate the role of the continual ‘task-unrelated’ eye movements made viewing familiar faces, and we hypothesize that they are ‘learning-style’ eye movements, important to constantly update an internal template of this face and maintain stable familiarity to it despite changes in its appearance. This is driven by the idea that under ‘natural’ conditions, even the faces most familiar to us constantly change from one state to another, rapidly (due to applying skin cosmetics, shaving facial hair, etc.) or slowly (changes due to aging, etc.). Our hypothesis is also rooted in previous studies: recognition was shown to benefit from multiple different exposures to a same face (Matthews et al., 2018), and the nature of eye movements changes with more exposures. Generally, external features are more important to recognize not-so familiar faces, but internal features are more looked at for familiar faces (Ellis et al., 1979; Young et al., 1985), and, as a faces becomes more familiar, the eyes are focused on more than the rest of the face (Heisz & Shore, 2008). To test the hypothesis that the eye movements made when viewing very familiar faces are to maintain stable familiarity to those faces, we set up a paradigm in which learned faces are dynamically altered, but the pilot has not been run yet.

11. Conclusion

The data collected in the experiment show that eye movements are not necessary for discriminating familiar faces from unfamiliar faces. Indeed, participants were able to make such judgements for faces bigger than sizes naturally seen (up to 115° in height) even under fixation, probably due to their ability to simultaneously attend to features in their peripheral vision.

Importantly, although we wanted to see if allowing eye movements during face viewing would be instrumental to bind facial features and permit the discrimination of familiarity under bigger presentation sizes, we found that eye movements did not lead to an advantage compared to when face familiarity had to be discriminated under forced fixation. However, as the presentation size was increased, participants looked at the areas surrounding the center rather than the center itself, probably to gather more information from the face and reach a decision regarding familiarity. Thus, it seems that eye movements are not necessary to assess the familiarity of faces ranging from extremely small sizes (1°), to face sizes beyond those seen naturally (115°), even though we tend to look at the most salient features when viewing faces, resulting in 'T-shape' gaze patterns.

Interestingly, this result, along with the fact that combining eye movements with head movements extended the range of presentation sizes over which faces could be discriminated, mean that we do not need high-resolution input from these facial features to perform face recognition tasks. In addition, although previous studies have found the central spot of fixation to be important for recognition of faces (perhaps due to the small stimuli sizes used in those studies), it is not itself important, but rather allows the facial features to all be accessible in peripheral view.

Finally, participants were able to perform better during the second week of the study, probably due to learning a new strategy or to task training.

12. Figures

	Classic eye-tracker	**Virtual reality**
System used	EyeLink 1000 Plus	VIVE Pro Eye + SRanipal SDK
Sampling frequency	2000Hz	90Hz
Accuracy	0.15°	0.5-1° *
Trackable field of view	32° (horizontal); 25° (vertical)	110° (horizontal)
Calibration system	9-point	5-point

Table 1. Specifications of the two systems used for the experiments. * Within central 20° of the field of view.

	Average	**Minimum**	**Maximum**	**Percentage**
‘False alarm’ rate (%)	4.97	0.38	9.32	
Age	27.4	22	44	
PI-20 score *	40.8	23	69	
Video game-playing score **	1.3	0	4	
Corrective glasses/ lenses				66.66
CVD ***				0.0

Table 2. Participant demographics. The table lists the average, minimum and maximum values of the participants’ ‘False alarm’ rate, age, PI-20 scores, video game-playing scores, as well as the percentage of participants who wear corrective glasses or contact lenses, and the percentage with CVD. One of the participants did not fill the survey regarding demographic information and was excluded from this table. * Generally, scores of 65-74, 75-84, and 85-100 respectively indicate mild, moderate, and severe developmental prosopagnosia (Shah et al., 2015). ** Self-reported frequency at which participants play video games: ‘Never’ (0), ‘A few times a year’ (1), ‘A few times a month’ (2), ‘Once a week’ (3), or ‘More than once a week’ (4). *** Measured with an Ishihara colour plates test.

Recognition * / Familiarity	Yes	No
Familiar	0.79	0.21 **
Unfamiliar	0.050	0.95 **

Table 3. Average rates of familiar and unfamiliar face recognition across all trials. Values are reported as a proportion over all trials. * Here, 'recognition' is defined as clicking on the response button while the face is shown on screen. ** Due to the chosen experimental paradigm, the values for 'Misses' (No – Familiar) and 'Correct rejections' (No – Unfamiliar) may not be indicative of participant performance.

Sessions / Trial type	All	Week 1	Week 2
‘Fixation’	0.81 ± 0.066	0.85 ± 0.048	0.88 ± 0.015
‘Free-View’	0.93 ± 0.016	0.84 ± 0.031	0.92 ± 0.013
‘Head-Free’	0.79 ± 0.032	0.58 ± 0.056	0.65 ± 0.048

Table 4. Mean R^2 fits for the psychometric curves. For each participant, three psychometric curves were fit for each trial type: one using all experiment sessions, another using sessions from the first week of the study, and a last one using sessions from the second week of the study. R^2 fits were calculated for each curve, and the average values across participants are reported with their corresponding SEM.

Face size	Trial type / Trials	'Fixation'	'Free-View'	'Head-Free'
0.5°	**Received responses**	9.3 ± 1.4	11.2 ± 1.6	
	Completed trials	25.9 ± 3.8	40.3 ± 1.7	
1°	**Received responses**	18.7 ± 1.3	22.0 ± 1.2	
	Completed trials	31.3 ± 2.4	38.5 ± 1.4	
2°	**Received responses**	16.7 ± 0.8	17.1 ± 1.1	
	Completed trials	28.8 ± 2.2	40.5 ± 1.2	
5°	**Received responses**	26.5 ± 2.5	30.8 ± 1.6	9.7 ± 0.8
	Completed trials	44.2 ± 4.6	71.5 ± 2.0	18.7 ± 1.1
10°	**Received responses**	22.5 ± 1.5	31.0 ± 1.1	9.5 ± 1.0
	Completed trials	36.3 ± 3.4	60.7 ± 1.3	19.6 ± 1.4
15°	**Received responses**	22.4 ± 1.8	31.3 ± 1.4	9.7 ± 0.7
	Completed trials	35.5 ± 3.9	64.3 ± 2.0	19.4 ± 1.3

Face size	Trial type / Trials	'Fixation'	'Free-View'	'Head-Free'
30°	**Received responses**	20.8 ± 1.9	35.3 ± 1.7	10.6 ± 0.6
	Completed trials	30.5 ± 3.4	64.3 ± 2.3	20.1 ± 0.9
50°	**Received responses**	6.3 ± 1.0	9.7 ± 0.9	10.3 ± 1.0
	Completed trials	10.8 ± 1.8	20.2 ± 1.2	19.1 ± 1.5
70°	**Received responses**	4.9 ± 0.7	9.5 ± 0.6	8.8 ± 0.9
	Completed trials	9.2 ± 1.3	20.4 ± 1.0	18.1 ± 1.2
90°	**Received responses**	3.0 ± 0.7	6.3 ± 1.0	8.0 ± 1.2
	Completed trials	8.1 ± 1.1	16.6 ± 1.6	18.8 ± 1.5
110°	**Received responses**	1.1 ± 0.4	2.3 ± 0.6	4.4 ± 0.7
	Completed trials	10.3 ± 1.6	18.6 ± 1.4	19.1 ± 1.4
130°	**Received responses**	0.1 ± 0.1	0.7 ± 0.3	2.5 ± 0.8
	Completed trials	8.9 ± 1.6	19.4 ± 1.3	20.1 ± 1.5

Face size	Trial type / Trials	'Fixation'	'Free-View'	'Head-Free'
150°	**Received responses**	0.1 ± 0.1	0.4 ± 0.2	0.5 ± 0.2
	Completed trials	10.2 ± 1.3	19.1 ± 1.2	18.2 ± 1.1

Table 5. Average number of trials across experimental conditions. For each experimental condition (face size and trial type), we calculated the average number of trials with received responses, and the average number of completed trials across participants, with their corresponding SEM. 'Fixation' trials were discarded if participants looked away from the fixation point during the trial. 'Head-Free' trials were only presented for half of the sessions (only in VR sessions).

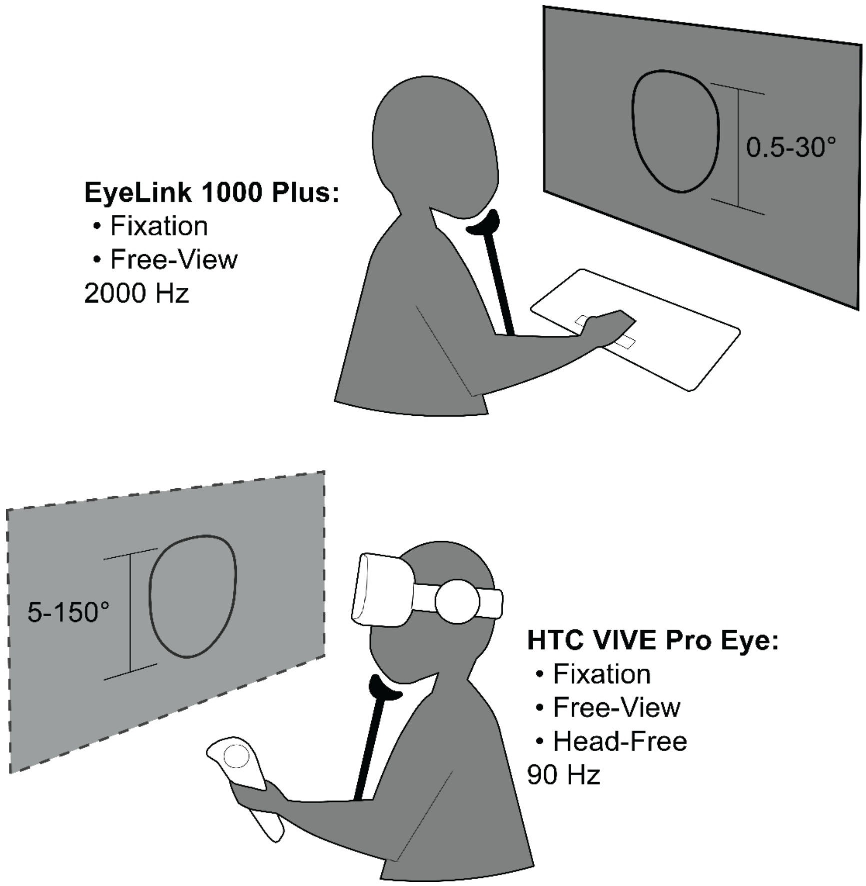

Figure 1. Summary schematic of the two set-ups used in the experiments. A chinrest was used to stabilize participants' heads for both set-ups. During the EL sessions, a SR Research EyeLink 1000 Plus was used for eye tracking at a sampling frequency of 2000Hz. Images were presented on a 28-Inch UHD LED-Lit Monitor under 'Fixation' and 'Free-View' trial types, and the faces subtended 0.5, 1, 2, 5, 10, 15, and 30° in height. Participants used the spacebar of a keyboard to provide responses. During the VR sessions, an HTC VIVE Pro Eye virtual reality headset with built-in eye tracking was used to record eye movements. Images were presented within this headset under 'Fixation', 'Free-View', and 'Head-Free' trial types, and the face sizes tested were 5, 10, 15, 30, 50, 70, 90, 110, 130, and 150° in height. Participants provided responses by pressing on the button of a hand-held controller.

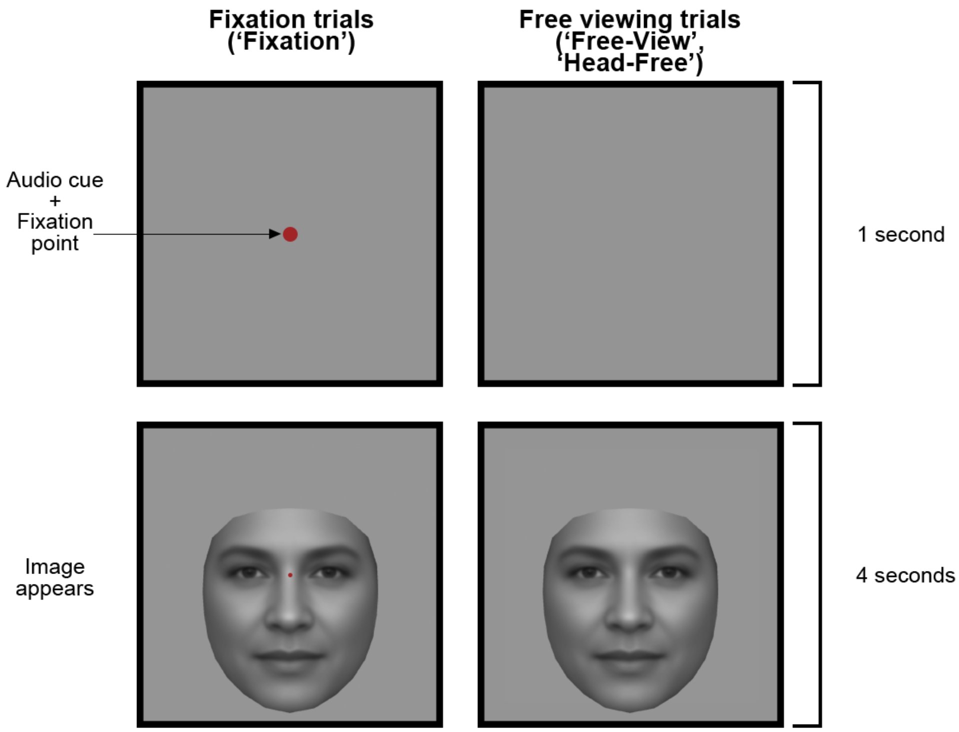

Figure 2. Schematic contrasting forced fixation versus free viewing trials. At the start of a trial, an audio cue stating the viewing condition of the trial ('Fixation', 'Free-View', and 'Head-Free') played. At the start of fixation trials, a red point appeared, and after having been fixated for a second, it shrunk, and an image of a face appeared. During free viewing trials, the image appeared one second after the audio cue. In the EL sessions, participants were asked to sit with their head 60 cm away from the monitor. In the VR sessions, the faces were presented at a distance of 11 Unity metric units away from their head. In both systems, the faces stayed onscreen for a duration four seconds, or until a response was received from the participant. The faces were shown at various sizes in height but were all centered on a point equidistant from both eyes and aligned with the nose, which coincided with the location of the fixation point. This location was chosen due to being one of the more informative for recognition (Peterson & Eckstein, 2012). The face shown in this figure corresponds to the average shape and texture from our dataset of images.

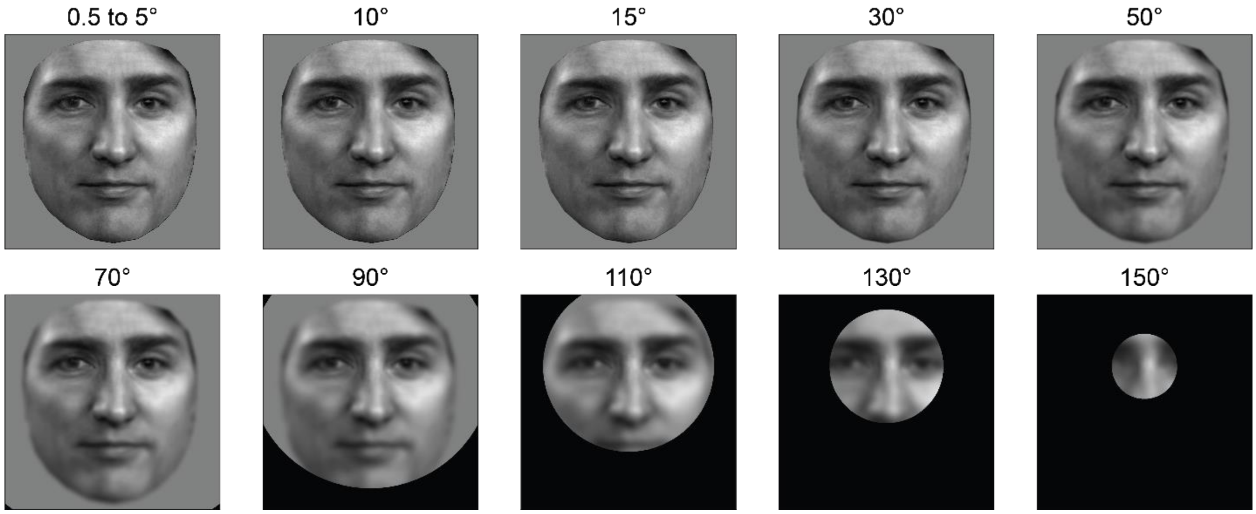

Figure 3. Foveated images demonstrating faces of various sizes as seen under forced fixation. A face of an individual that is familiar to the reader, Justin Trudeau, is used to display the foveation effect. The images are foveated, i.e. blurred according to visual acuity falloff with eccentricity (Geisler & Perry, 1998), and shown as seen when fixating faces of the various sizes presented in experiment on the point above the nose that is equidistant ton both eyes. The circular black masks indicate the portions of the face that fall outside of the field of view during fixation. Dr. Tong wrote the code used to obtain these images.

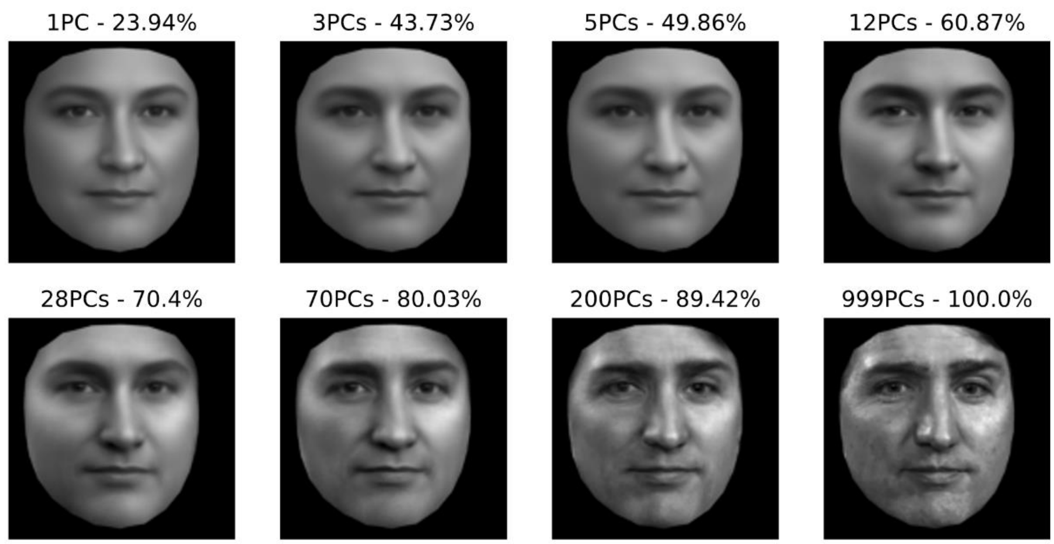

Figure 4. Reconstruction of an example face using different numbers of texture principal components following projection onto PCA space. A face of an individual familiar to the reader, Justin Trudeau, was chosen as an example to display the effect of using different numbers of texture PCs. The corresponding percentages of variance explained by the different numbers of PCs are included in the figure. The images used as stimuli in the experiment were constructed using 200 PCs for texture (or 89.42% of the explained variance).

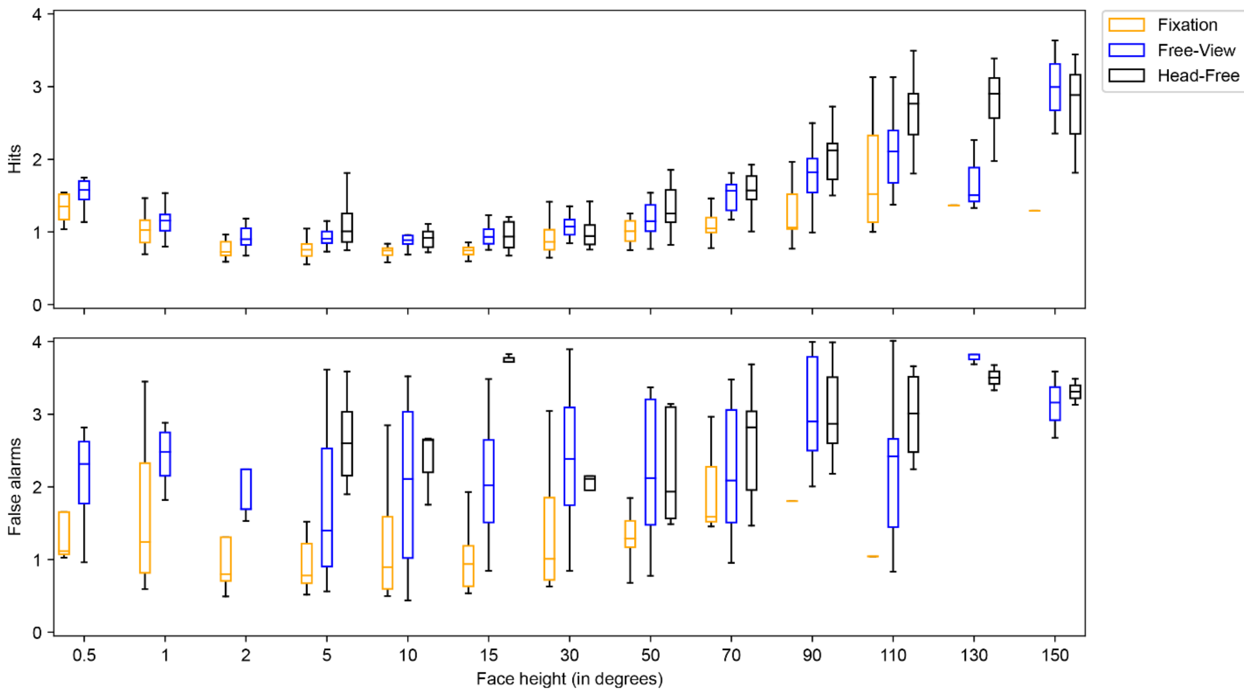

Figure 5. Median response times of 'Hits' and 'False alarms' across different stimulus sizes and trial types. Mean response times (ranging from 0 to 4 seconds) were calculated for 'Hit' (i.e. recognizing a familiar face) and 'False alarm' responses (i.e. recognizing an unfamiliar face) per participant for each combination of face size and trial type. Median response times with their corresponding interquartile ranges are shown for 'Hits' (top graph) and 'False alarms' (bottom graph). The lack of data for 'Head-Free' trials for faces ranging 0.5 to 2° in height is not due to discarded trials, but rather due to these combinations of stimulus sizes and viewing condition not being tested in the VR system.

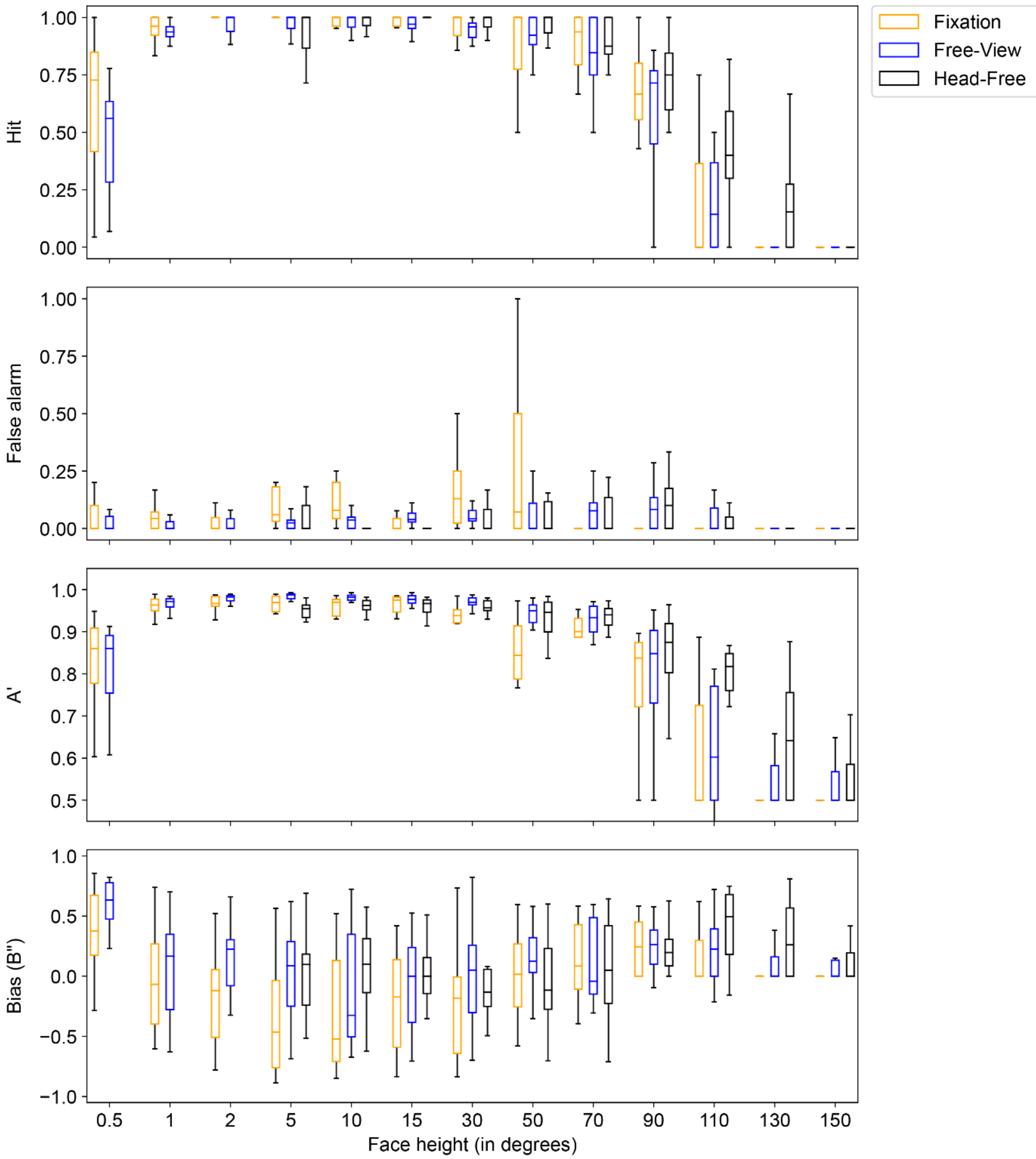

Figure 6. Median measures for discrimination of familiar faces from unfamiliar faces across different stimulus sizes and trial types. Mean values of 'Hit' and 'False alarm' rates (ranging from 0 to 1), A' (0.5 to 1.0), and B" (-1.0 to 1.0) were calculated per participant for each combination of face size and trial type. Box plots represent median values and their interquartile ranges, with whiskers extending the ranges 1.5 times. Faces 0.5 to 2° were not shown for 'Head-Free' trials, hence the lack of data.

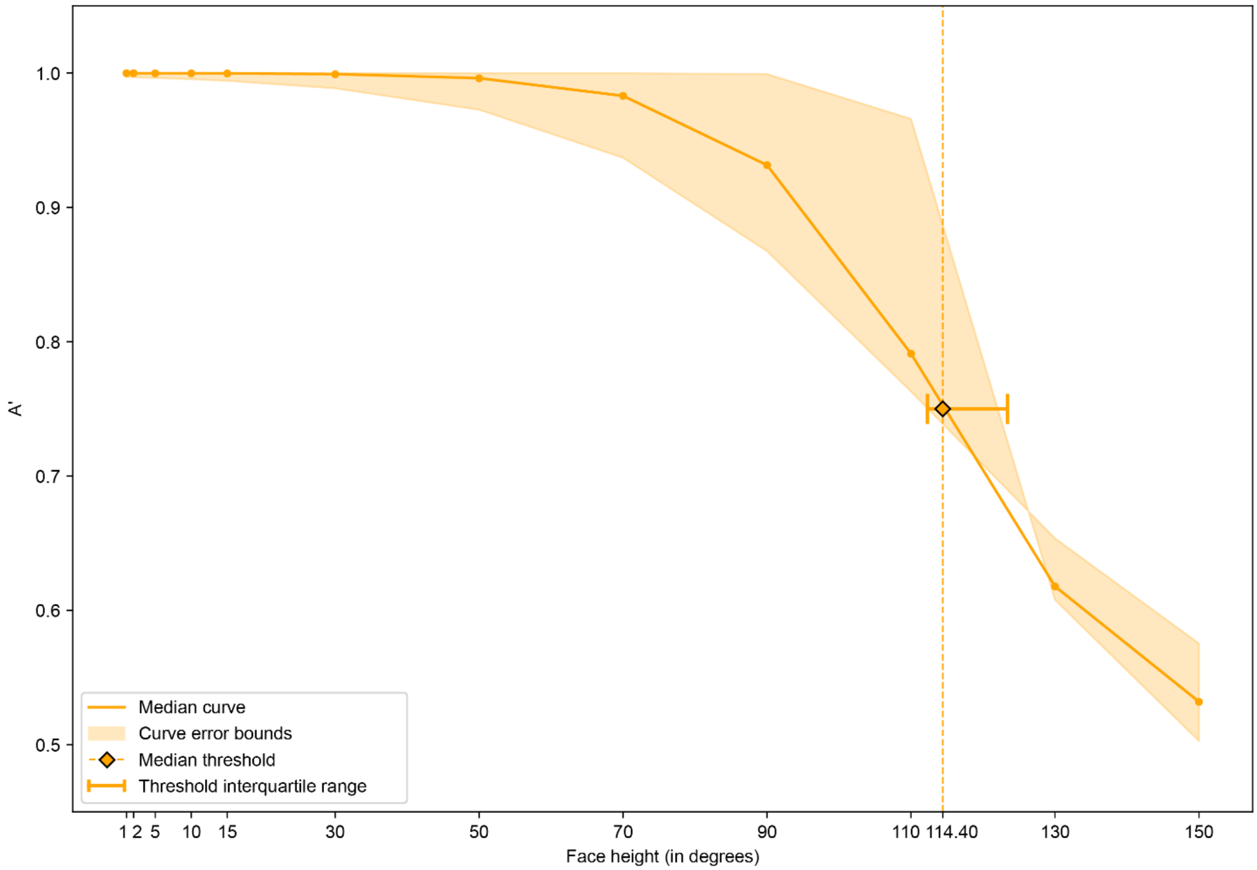

Figure 7. Median A' psychometric curve under the forced fixation condition. For each participant, a value of A' was calculated for each face size. These values were used to fit a sigmoid curve for each participant and obtain estimates of the 'sensory threshold' and the 'scale' parameters. The median values of those parameters were used to plot this figure, with error bounds defined from the interquartile ranges of the two fitted parameters. The median 'sensory threshold' value is indicated on the graph with its interquartile range. The mean 'sensory threshold' across participants was 105.69° ± 9.04° SEM, and the mean 'scale' was 0.15 ± 0.047 SEM.

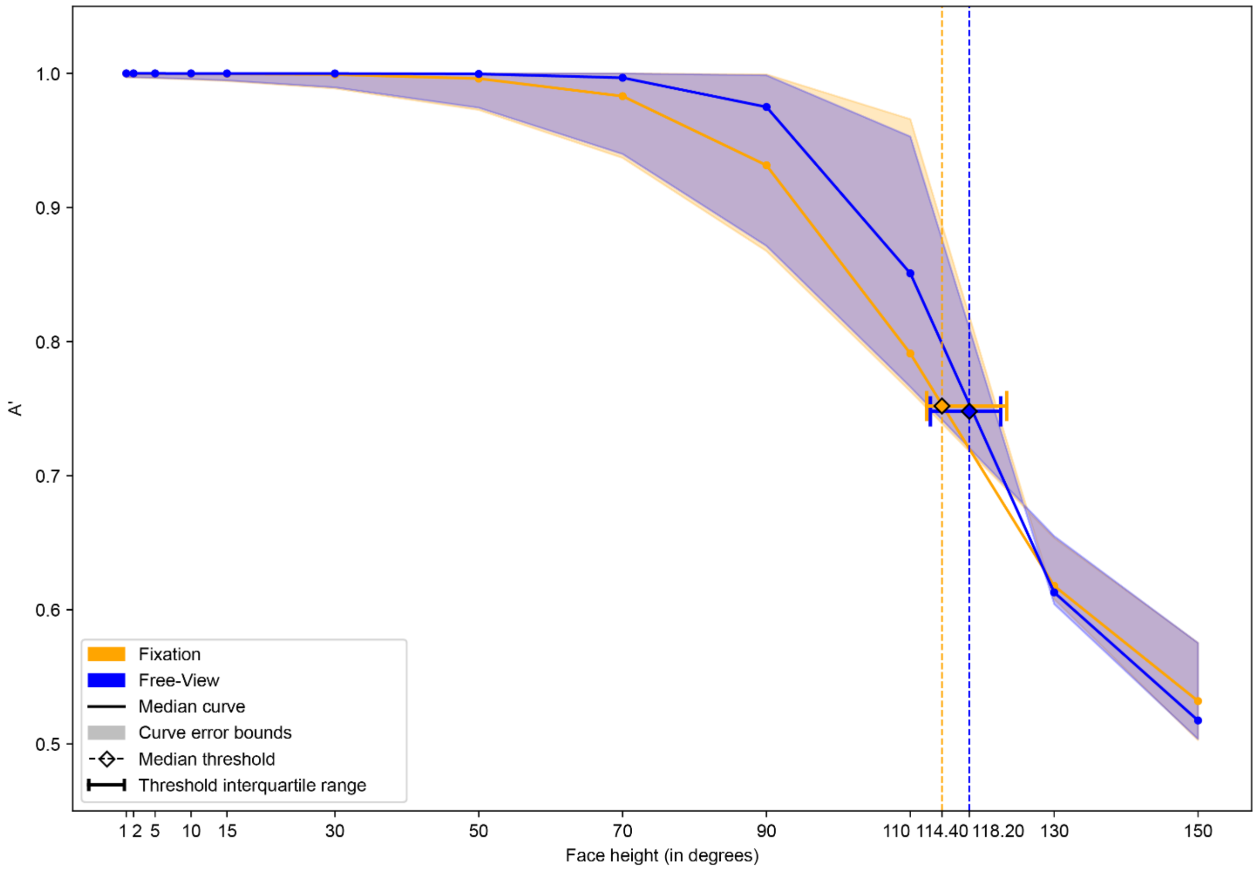

Figure 8. Comparison of median A' psychometric curves under 'Fixation' versus 'Free-View' conditions. For each participant, values of A' were calculated per experimental condition, used to fit two sigmoid curves for each participant, one for 'Fixation' and one for 'Free-View' and obtain the corresponding estimates of the 'sensory threshold' and the 'scale' parameters. The median values of those parameters were used to plot this figure, with error bounds defined from the interquartile ranges of the two fitted parameters. The median 'sensory threshold' values are indicated on the graph with their interquartile ranges. Across the 'Free-View' fitted curves, the mean 'threshold' across participants was 119.45° ± 4.18 SEM°, and the mean 'scale' was 0.19 ± 0.072 SEM. The values of the parameters were compared using the Mann-Whitney U test and were not found to be significantly different from each other ('sensory threshold': statistic = 69.0, p-value = 0.44; 'scale': statistic = 63.0, p-value = 0.31).

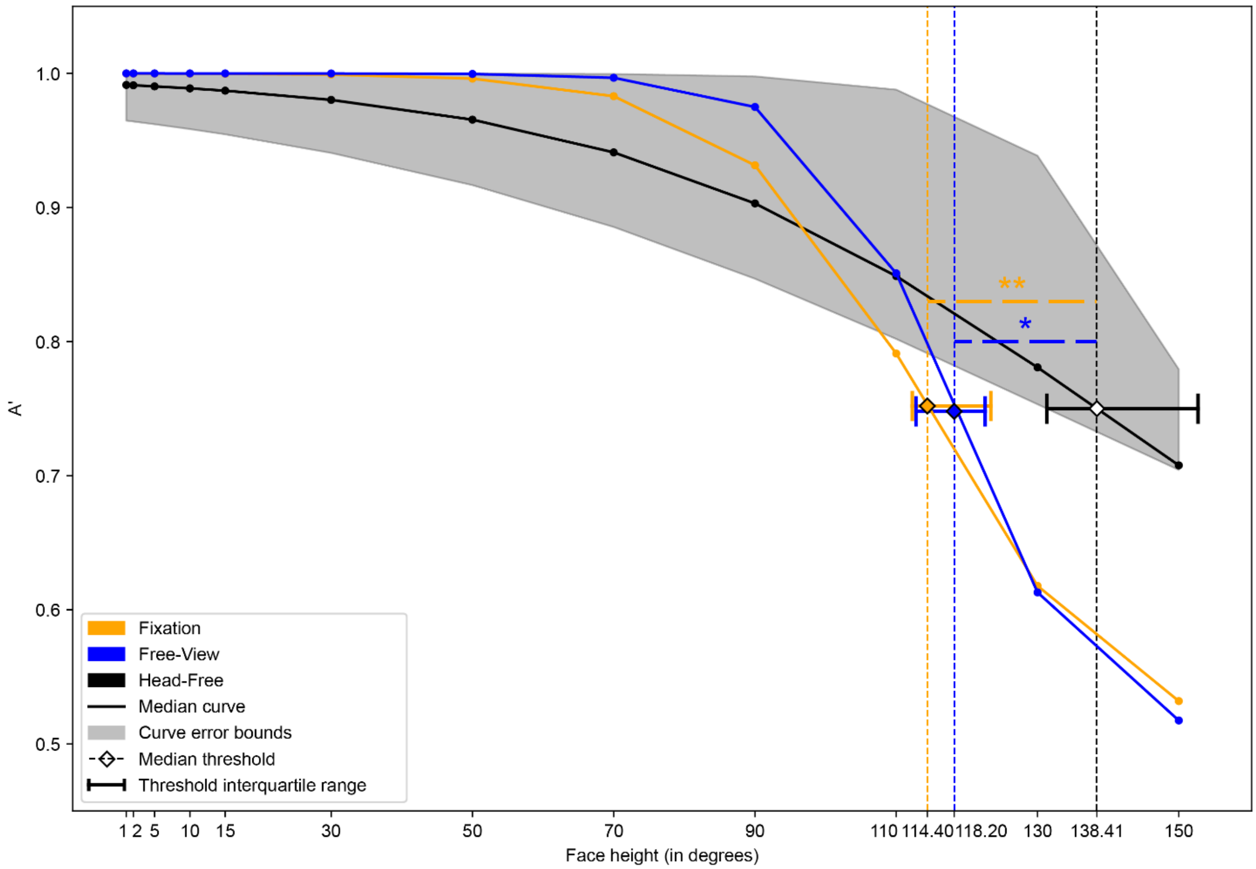

Figure 9. Comparison of the median A' psychometric curve under the 'Head-Free' condition against the curves under other viewing conditions. A' values were calculated per experimental condition. One sigmoid curve was fitted per trial type and estimates of the 'threshold' and the 'scale' parameters were obtained. For 'Head-Free', the mean 'threshold' was 141.77° ± 7.55° SEM, and the mean 'scale' was 0.081 ± 0.034 SEM. The 'Head-Free' median values were used to plot the curve, with error bounds defined from the parameters' interquartile ranges. The 'threshold' median values and their interquartile ranges are indicated on the graph for all trial types. Via the Kruskal-Wallis test, the 'scales' were not significantly different across the trial types (statistic = 4.94, p-value=0.084), but the 'sensory thresholds' were (statistic = 9.18, p-value = 0.010). Pairwise post-hoc Dunn's tests revealed the 'Head-Free' 'threshold' to be statistically different from 'Fixation' (p-value = 0.0098) and 'Free-View' (p-value = 0.029), as shown on the figure.

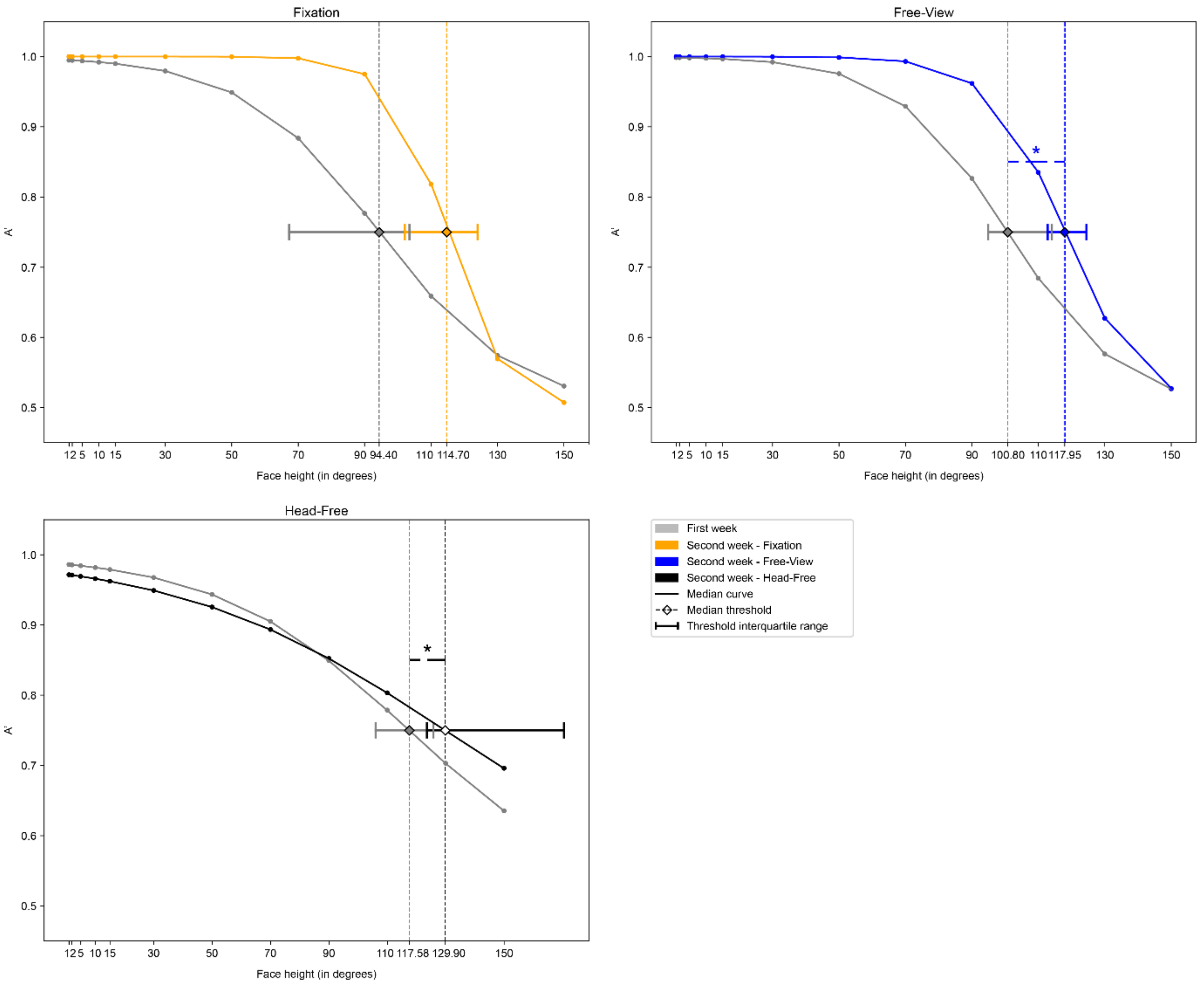

Figure 10. Comparison of the 'sensory threshold' parameter between the two weeks of the study, for each viewing condition. Data were separated according to the week of study they were recorded in, and A' values were calculated for experimental condition. One sigmoid curve was fitted per trial type per week and estimates of the 'threshold' and the 'scale' parameters were obtained. The 'threshold' median values and interquartile ranges are indicated on the graph for all trial types. Using the Wilcoxon signed-rank test, the 'sensory threshold' parameter was found to be significantly different between the first week and the second week of the experiments for both 'Free-View' (statistic = 11.0, p-value = 0.027) and 'Head-Free' trials (statistic = 5.0, p-value = 0.039). The threshold increased by a mean of 14.27° ± 4.84 SEM for the 'Free-View' viewing condition and 29.97° ± 11.40 SEM for the 'Head-Free' condition.

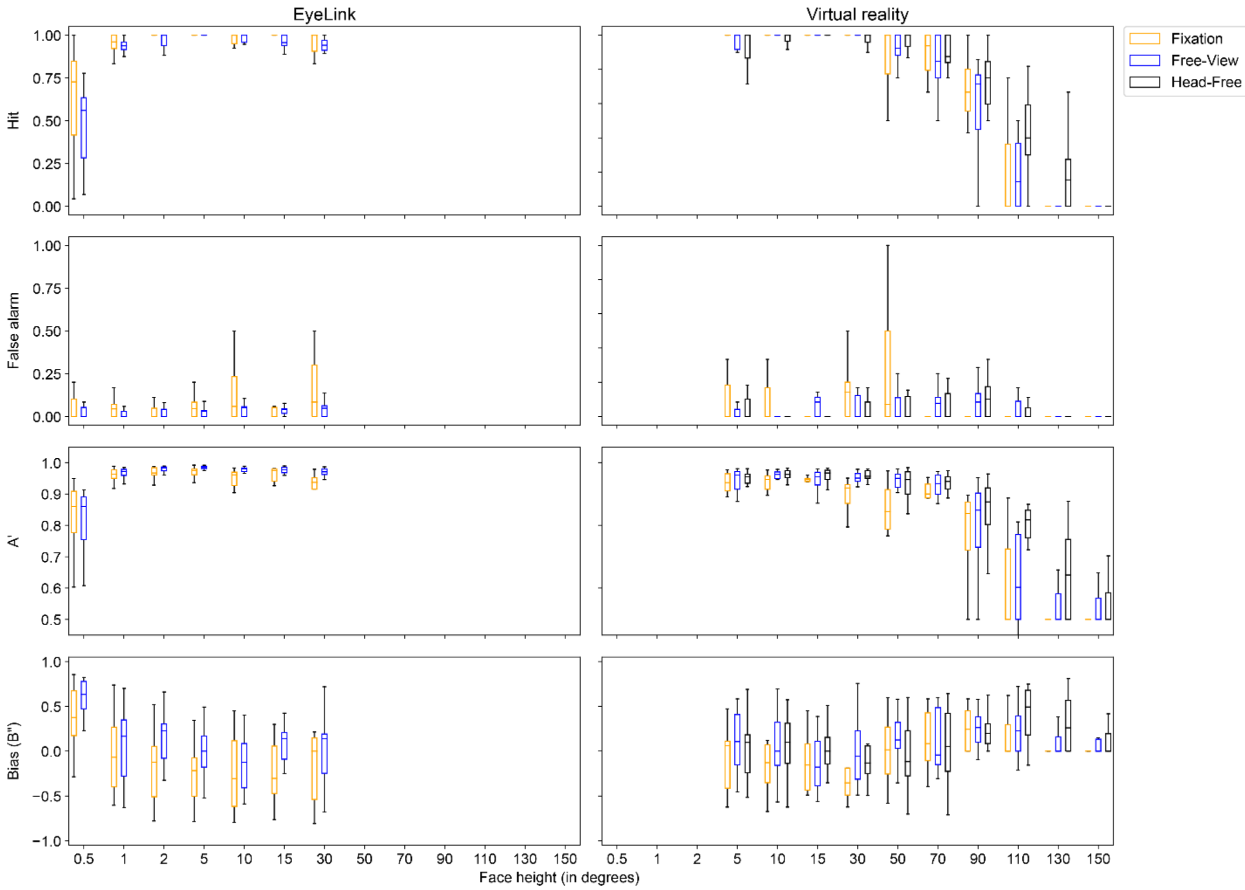

Figure 11. Comparison of median measures for discrimination of familiar faces from unfamiliar faces across the experimental set-ups. Mean values of 'Hit' and 'False alarm' rates (ranging from 0 to 1), A' (0.5 to 1.0), and B" (-1.0 to 1.0) were calculated per participant for each combination of face size and trial type. Median values with their corresponding interquartile ranges were then plotted as box plots, separately for sessions run in EL (left graph) and VR (right graph). Due to differences in the systems' specifications, different ranges of presentation sizes were shown in each system. The lack of data points outside of these ranges for each system is due to not having tested those combinations of stimulus sizes and experimental system, not due to discarded trials.

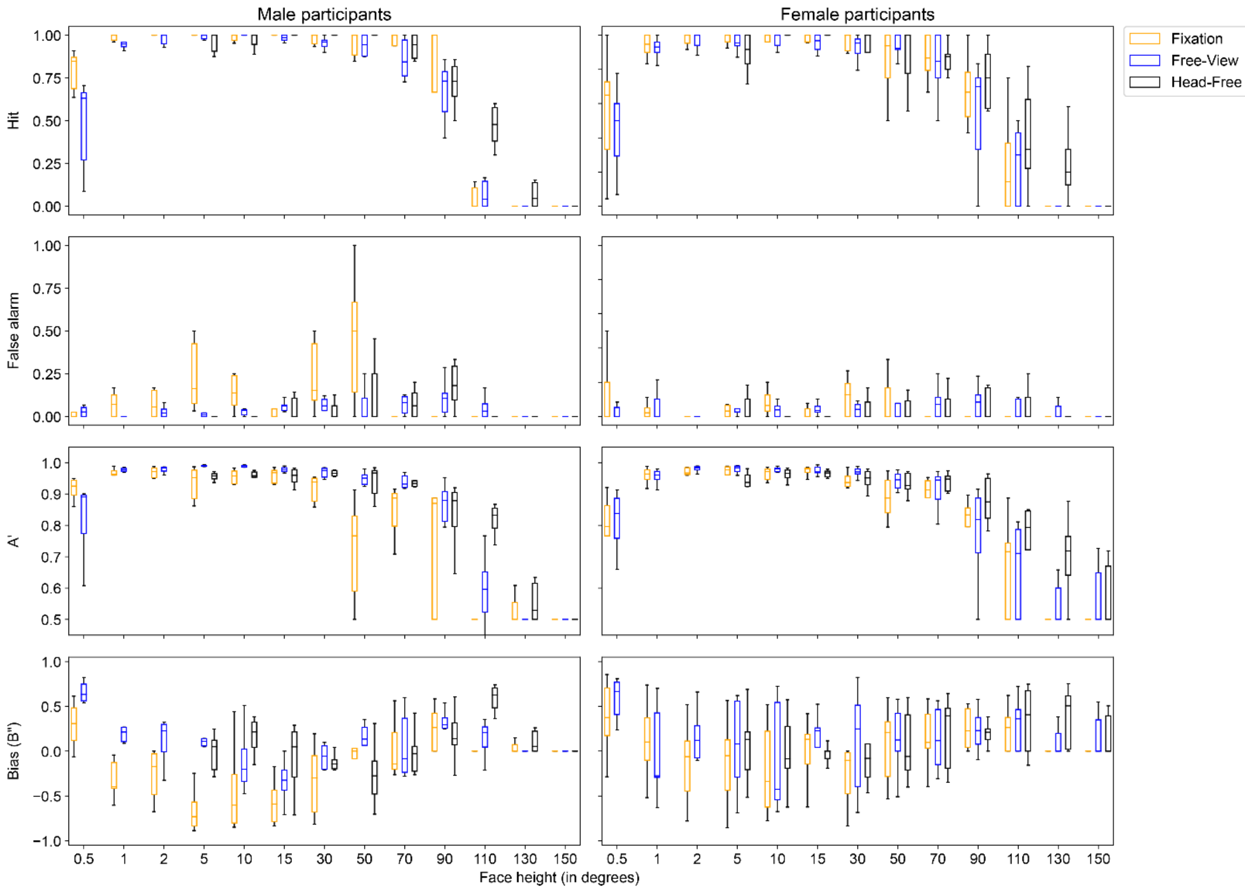

Figure 12. Comparison of median measures for discrimination of familiar faces from unfamiliar faces across participant sexes. Mean values of 'Hit' and 'False alarm' rates (ranging from 0 to 1) and of A' (ranging from 0.5 to 1.0) and B" (ranging from -1.0 to 1.0) were calculated for each participant for each combination of face size and trial type. Median values were then plotted separately for the six male (left graph) and nine female (right graph) participants, with corresponding interquartile ranges as box plots. The lack of data for 'Head-Free' trials for faces ranging 0.5 to 2° in height is not due to discarded trials, but rather due to these combinations of stimulus sizes and viewing condition not being tested in the VR system.

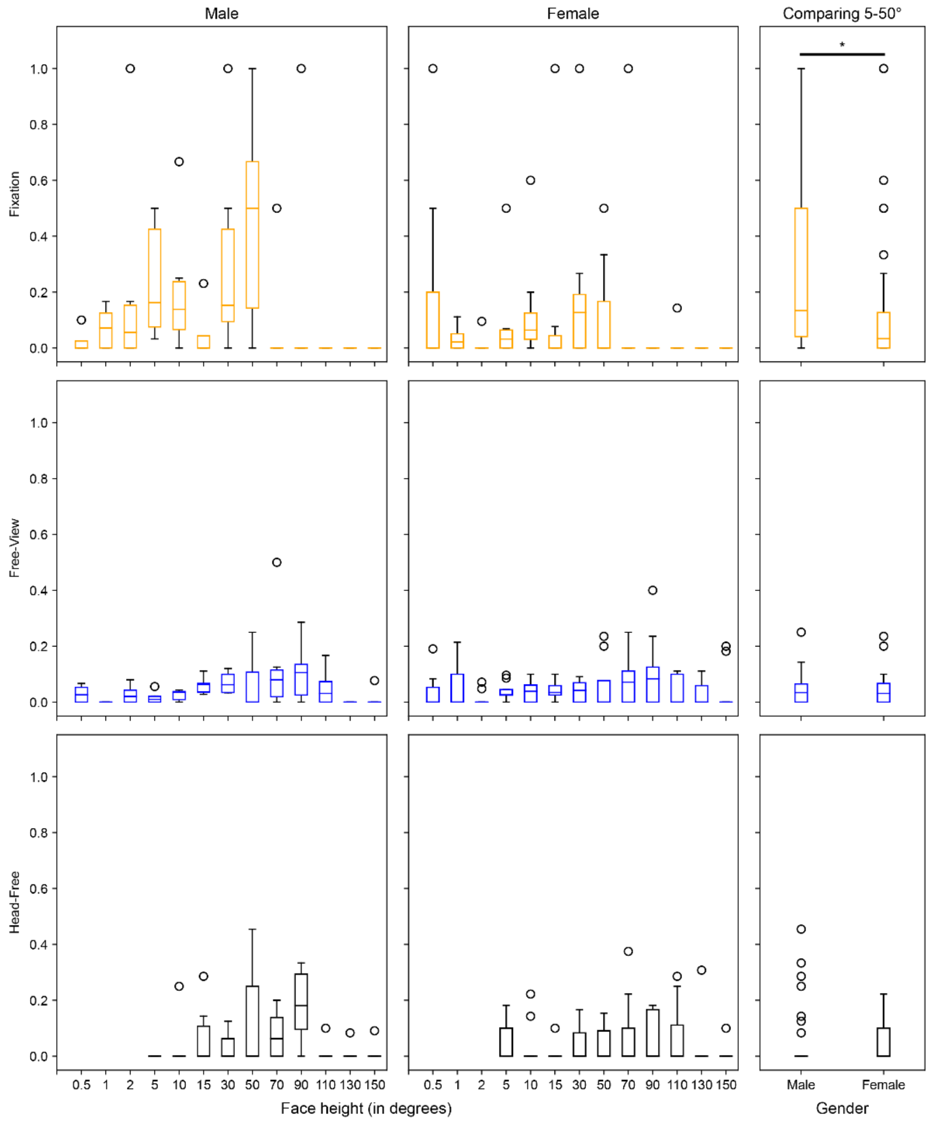

Figure 13. Comparison of 'False alarm' rates between male and female participants across conditions. Median values are represented as horizontal lines, interquartile ranges (Q1 to Q3) as boxes, and fliers as circles. Whiskers extend the interquartile range 1.5 times. 'False alarm' values were calculated per participant for each combination of face size and trial type. Median values were plotted separately for male and female participants in the first two columns. The third column shows the median values across sex for pooled 5-50° faces. Mann-Whitney U tests were used to compare values across sex for each trial type. For 'Fixation', male (n = 6) and female (n = 9) participants had statistically different 'False alarm' rates (FDR-corrected p-value = 0.026).

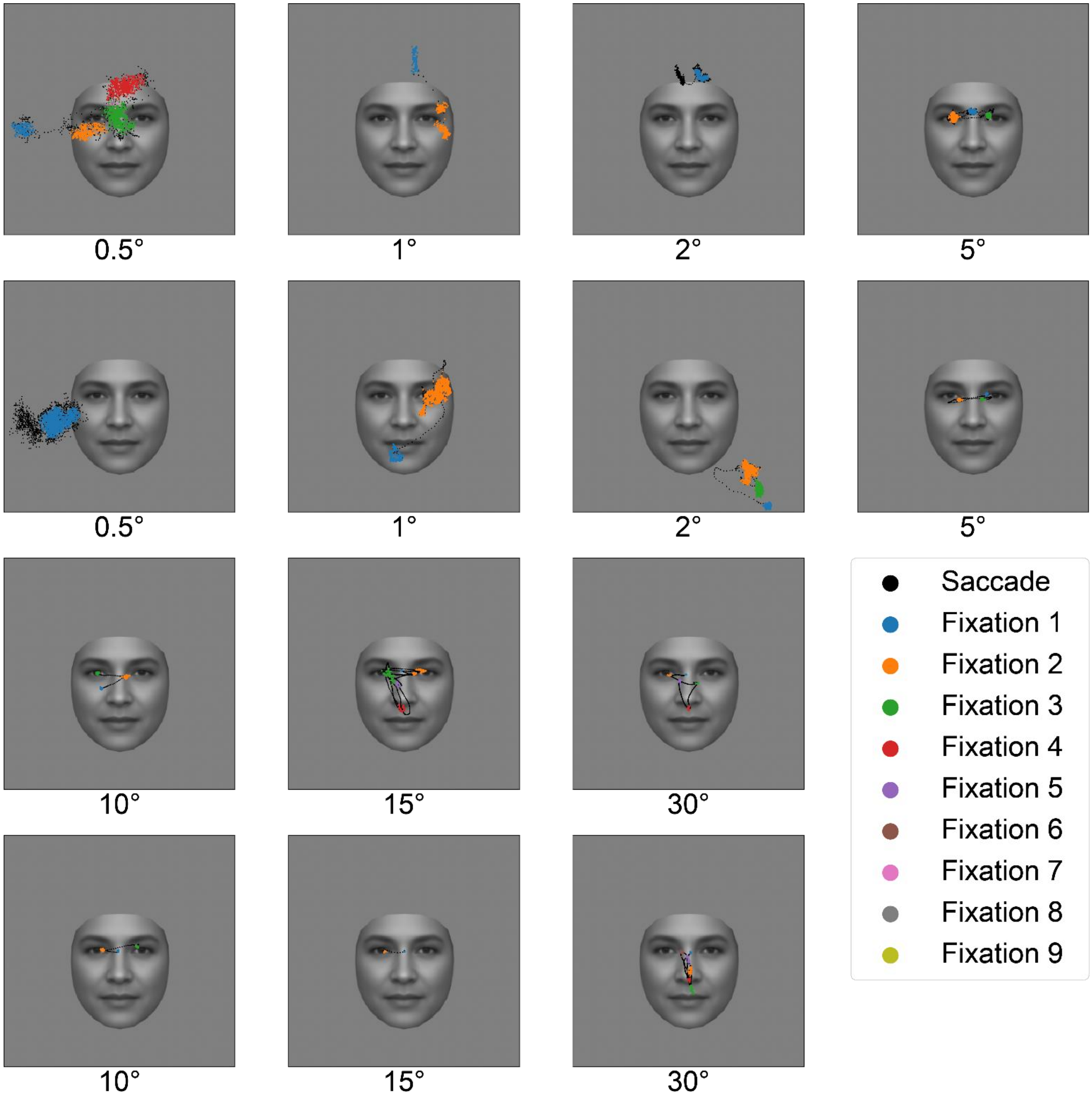

Figure 14. Example free viewing eye movement sequences. The saccades (marked in black) and the different fixation clusters (in colours, ordered chronologically) were identified using DBSCAN. The scan paths made on the face viewed were morphed into scan paths made onto the average face shape before plotting.

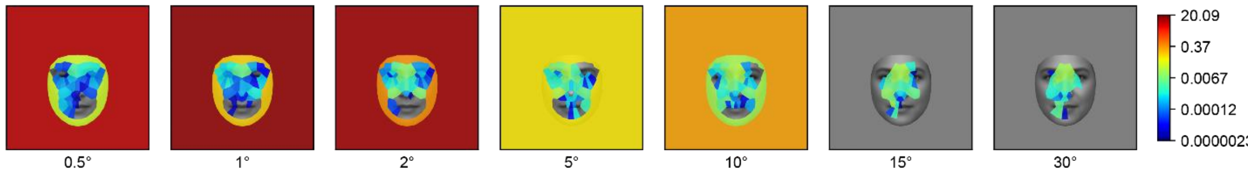

Figure 15. Heat maps of the density of time spent looking at 49 regions of interest of different sizes during 'Free-View' trials across face sizes. Of these regions, 47 correspond to Voronoi cells obtained from facial features, one corresponds to the general face outline, and the last one corresponds to off-face locations. The colours were plotted on a log scale and represent the density of time spent in each region of interest (i.e. ms/units of surface area, as the time spent in each region of interest was divided by the surface area of the ROI). Areas were left uncoloured if no fixations were made on them.

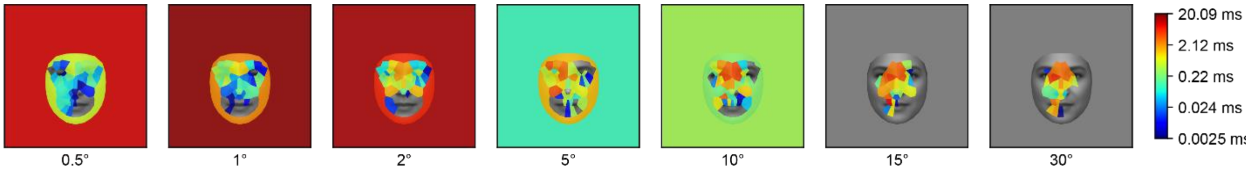

Figure 16. Heat maps of the amount of time spent looking at 49 regions of interest of different sizes during 'Free-View' trials across face sizes. Of these regions, 47 correspond to Voronoi cells obtained from facial features, one corresponds to the general face outline, and the last one corresponds to off-face locations. Each region has a different surface area. The colours were plotted on a log scale and represent the amount of time spent in each region of interest. Areas were left uncoloured if no fixations were made on them.

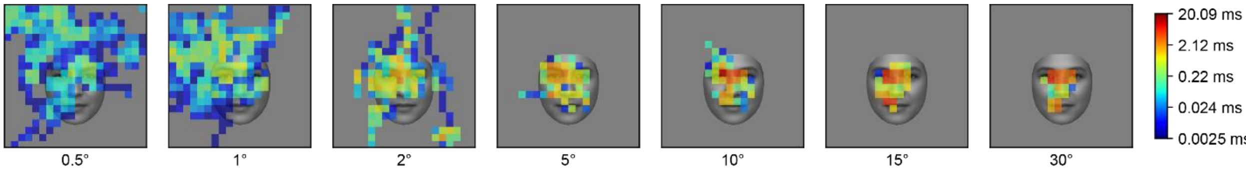

Figure 17. Heat maps of the amount of time spent looking at 400 square areas of equal sizes during 'Free-View' trials across face sizes. The heat maps were produced by calculating the amount of time spent in each area and normalizing the values. The colours were plotted on a log scale. Areas were left uncoloured if no fixations were made on them.

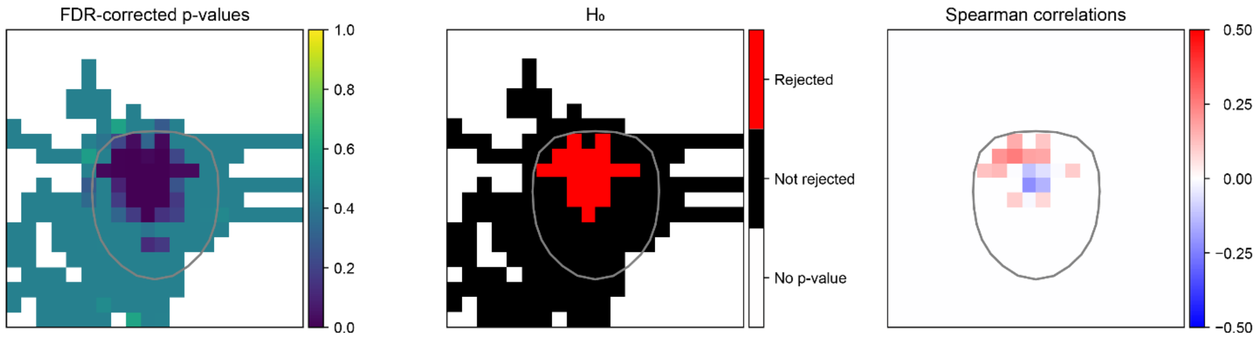

Figure 18. Heat maps showing the ANOVA and post-hoc Spearman analysis comparing the amount of time spent in the binned areas across faces of 5-30° in height. The image was binned into 400 areas arranged in a 20 by 20 grid. For each binned area, a one-way repeated measures ANOVA with amount of time (in ms) spent looking at the area in the first 500ms of each trial as a dependent variable and face size as an independent variable was run across the participants. The obtained p-values were FDR corrected and were plotted to the left. The middle heat map displays the binned areas where a significant difference was observed across the face sizes after FDR correction, i.e. the binned areas for which H_0 is rejected. The ANOVA could not be run on binned areas that were not fixated, and these missing p-values are indicated by white pixels in these two heat maps. For the areas for which H_0 was rejected, we calculated post-hoc Spearman rank-order correlation coefficients on the time spent in the area against face sizes, as shown in the heat map to the right.

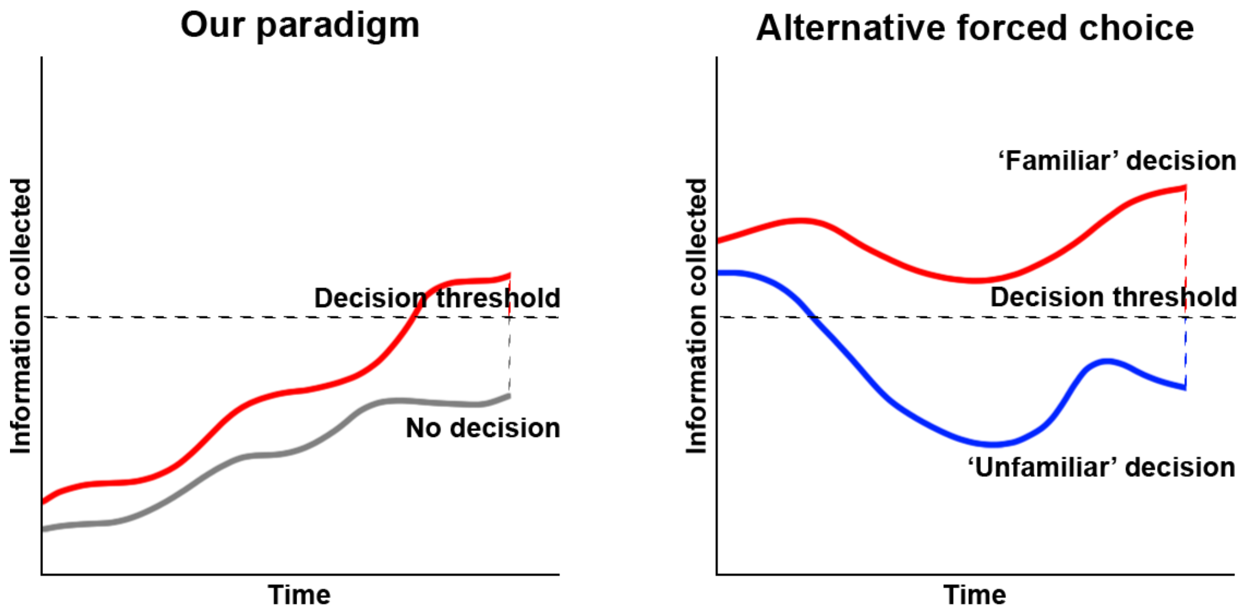

Figure 19. Our task as compared to AFC tasks. In our task (left), participants click on a button if they have accumulated enough evidence to reach a decision threshold that they set internally. In an AFC task (right), participants compare the amount of information collected a threshold mid-way between 'Familiar' and 'Unfamiliar' and click on the corresponding button. They essentially choose the category they are closer to, whereas in our task, they only choose the 'Familiar' category if they have reach (or pass) their internal threshold.

13. Copyright

The unfamiliar faces presented in all the experiments were obtained from two online databases, for which we obtained permission to use for research purposes only, but not distribute or reproduce:

- the FEI Face Database (Thomaz, 2005)
- the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office (Phillips et al., 2000; Phillips et al., 1998).

14. References

Arizpe, J. M., Noles, D. L., Tsao, J. W., & Chan, A. W. (2019). Eye Movement Dynamics Differ between Encoding and Recognition of Faces. *Vision (Basel)*, *3*(1). https://doi.org/10.3390/vision3010009

Barton, J. J., Radcliffe, N., Cherkasova, M. V., Edelman, J., & Intriligator, J. M. (2006). Information processing during face recognition: the effects of familiarity, inversion, and morphing on scanning fixations. *Perception*, *35*(8), 1089-1105. https://doi.org/10.1068/p5547

Bosler, F. (2019). Image Scraping with Python: A code-along guide to learn how to download images from Google with Python! Retrieved 31/08/2021, from https://towardsdatascience.com/image-scraping-with-python-a96feda8af2d

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis*, *10*(4), 433-436. https://www.ncbi.nlm.nih.gov/pubmed/9176952

Brecht, M., & Freiwald, W. A. (2012). The many facets of facial interactions in mammals. *Curr Opin Neurobiol*, *22*(2), 259-266. https://doi.org/10.1016/j.conb.2011.12.003

Buchan, J. N., Pare, M., & Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamic face processing. *Soc Neurosci*, *2*(1), 1-13. https://doi.org/10.1080/17470910601043644

Burton, A. M., Schweinberger, S. R., Jenkins, R., & Kaufmann, J. M. (2015). Arguments Against a Configural Processing Account of Familiar Face Recognition. *Perspect Psychol Sci*, *10*(4), 482-496. https://doi.org/10.1177/1745691615583129

Butler, S., Blais, C., Gosselin, F., Bub, D., & Fiset, D. (2010). Recognizing famous people. *Atten Percept Psychophys*, *72*(6), 1444-1449. https://doi.org/10.3758/APP.72.6.1444

Caldara, R., Schyns, P., Mayer, E., Smith, M. L., Gosselin, F., & Rossion, B. (2005). Does prosopagnosia take the eyes out of face representations? Evidence for a defect in representing diagnostic facial information following brain damage. *J Cogn Neurosci*, *17*(10), 1652-1666. https://doi.org/10.1162/089892905774597254

Caldara, R., Zhou, X., & Miellet, S. (2010). Putting culture under the 'spotlight' reveals universal information use for face recognition. *PLoS One*, *5*(3), e9708. https://doi.org/10.1371/journal.pone.0009708

Cattaneo, Z., Renzi, C., Bona, S., Merabet, L. B., Carbon, C. C., & Vecchi, T. (2014). Hemispheric asymmetry in discriminating faces differing for featural or configural (second-order relations) aspects. *Psychon Bull Rev*, *21*(2), 363-369. https://doi.org/10.3758/s13423-013-0484-2

Chuk, T., Chan, A. B., & Hsiao, J. H. (2014). Understanding eye movements in face recognition using hidden Markov models. *J Vis*, *14*(11). https://doi.org/10.1167/14.11.8

Chuk, T., Chan, A. B., & Hsiao, J. H. (2017). Is having similar eye movement patterns during face learning and recognition beneficial for recognition performance? Evidence from hidden Markov modeling. *Vision Res*, *141*, 204-216. https://doi.org/10.1016/j.visres.2017.03.010

Chuk, T., Crookes, K., Hayward, W. G., Chan, A. B., & Hsiao, J. H. (2017). Hidden Markov model analysis reveals the advantage of analytic eye movement patterns in face recognition across cultures. *Cognition*, *169*, 102-117. https://doi.org/10.1016/j.cognition.2017.08.003

Corrow, S. L., Dalrymple, K. A., & Barton, J. J. (2016). Prosopagnosia: current perspectives. *Eye Brain*, *8*, 165-175. https://doi.org/10.2147/EB.S92838

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: face detection in just 100 ms. *J Vis*, *10*(4), 16 11-17. https://doi.org/10.1167/10.4.16

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. *Perception*, *8*(4), 431-439. https://doi.org/10.1068/p080431

Geisler, W., & Perry, J. (1998). *Real-time foveated multiresolution system for low-bandwidth video communication* (Vol. 3299). SPIE. https://doi.org/10.1117/12.320120

Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Res*, *41*(17), 2261-2271. https://doi.org/10.1016/s0042-6989(01)00097-9

Harrison, M. T., & Strother, L. (2020). Does right hemisphere superiority sufficiently explain the left visual field advantage in face recognition? *Atten Percept Psychophys*, *82*(3), 1205-1220. https://doi.org/10.3758/s13414-019-01896-0

Heisz, J. J., & Shore, D. I. (2008). More efficient scanning for familiar faces. *J Vis*, *8*(1), 9 1-10. https://doi.org/10.1167/8.1.9

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Mem Cognit*, *33*(1), 98-106. https://doi.org/10.3758/bf03195300

Hillger, L. A., & Koenig, O. (1991). Separable mechanisms in face processing: evidence from hemispheric specialization. *J Cogn Neurosci*, *3*(1), 42-58. https://doi.org/10.1162/jocn.1991.3.1.42

Hsiao, J. H.-W., & Cottrell, G. (2008). Two Fixations Suffice in Face Recognition. *Psychological Science*, *19*(10), 998-1006. https://doi.org/10.1111/j.1467-9280.2008.02191.x

Hsiao, J. H. (2010). Eye movements in face recognition. In M. Oravec (Ed.), *Face Recognition* (pp. 255-272). Intechopen. https://www.intechopen.com/books/3735

Imaoka, Y., Flury, A., & de Bruin, E. D. (2020). Assessing Saccadic Eye Movements With Head-Mounted Display Virtual Reality Technology. *Front Psychiatry*, *11*, 572938. https://doi.org/10.3389/fpsyt.2020.572938

Jenkins, R., White, D., Van Montfort, X., & Mike Burton, A. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313-323. https://doi.org/10.1016/j.cognition.2011.08.001

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: a review. *Memory*, *17*(5), 577-596. https://doi.org/10.1080/09658210902976969

Kanan, C., Bseiso, D. N., Ray, N. A., Hsiao, J. H., & Cottrell, G. W. (2015). Humans have idiosyncratic and task-specific scanpaths for judging faces. *Vision Res*, *108*, 67-76. https://doi.org/10.1016/j.visres.2015.01.013

Kelly, D. J., Liu, S., Rodger, H., Miellet, S., Ge, L., & Caldara, R. (2011). Developing cultural differences in face processing. *Dev Sci*, *14*(5), 1176-1184. https://doi.org/10.1111/j.1467-7687.2011.01067.x

Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition--women's faces make the difference. *Brain Cogn*, *50*(1), 121-128. https://doi.org/10.1016/s0278-2626(02)00016-7

Liu, J., Li, J., Feng, L., Li, L., Tian, J., & Lee, K. (2014). Seeing Jesus in toast: neural and behavioral correlates of face pareidolia. *Cortex*, *53*, 60-77. https://doi.org/10.1016/j.cortex.2014.01.013

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M. G., & Lee, J. (2019). Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172*. https://arxiv.org/abs/1906.08172

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. *Psychol Bull*, *98*(1), 185-199. https://www.ncbi.nlm.nih.gov/pubmed/4034817

Mario Kleiner, D. B., Denis Pelli. (2007). *"What's new in Psychtoolbox-3?"*. Pion Ltd.

Matthews, C. M., Davis, E. E., & Mondloch, C. J. (2018). Getting to know you: The development of mechanisms underlying face learning. *J Exp Child Psychol*, *167*, 295-313. https://doi.org/10.1016/j.jecp.2017.10.012

Miellet, S., Vizioli, L., He, L., Zhou, X., & Caldara, R. (2013). Mapping Face Recognition Information Use across Cultures. *Front Psychol*, *4*, 34. https://doi.org/10.3389/fpsyg.2013.00034

Mike Burton, A. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Q J Exp Psychol (Hove)*, *66*(8), 1467-1485. https://doi.org/10.1080/17470218.2013.800125

Milborrow, S., & Nicolls, F. (2014). Active shape models with SIFT descriptors and MARS. Proceedings of the 9th International Conference on Computer Vision Theory and Applications - Volume 1: VISAPP, (VISIGRAPP 2014),

Or, C. C., Peterson, M. F., & Eckstein, M. P. (2015). Initial eye movements during face identification are optimal and similar across cultures. *J Vis*, *15*(13), 12. https://doi.org/10.1167/15.13.12

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2012). Scikitlearn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825-2830. https://doi.org/10.48550/ARXIV.1201.0490

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis*, *10*(4), 437-442. https://www.ncbi.nlm.nih.gov/pubmed/9176953

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proc Natl Acad Sci U S A*, *109*(48), E3314-3323. https://doi.org/10.1073/pnas.1214269109

Peterson, M. F., & Eckstein, M. P. (2013). Individual differences in eye movements during face identification reflect observer-specific optimal points of fixation. *Psychol Sci*, *24*(7), 1216-1225. https://doi.org/10.1177/0956797612471684

Phillips, P. J., Hyeonjoon, M., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(10), 1090-1104. https://doi.org/10.1109/34.879790

Phillips, P. J., Wechsler, H., Huang, J., & Rauss, P. J. (1998). The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, *16*(5), 295-306. https://doi.org/10.1016/s0262-8856(97)00070-x

Ramon, M., & Rossion, B. (2012). Hemisphere-dependent holistic processing of familiar faces. *Brain Cogn*, *78*(1), 7-13. https://doi.org/10.1016/j.bandc.2011.10.009

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Q J Exp Psychol (Hove)*, *62*(8), 1457-1506. https://doi.org/10.1080/17470210902816461

Sadr, J., Jarudi, I., & Sinha, P. (2003). The role of eyebrows in face recognition. *Perception*, *32*(3), 285-293. https://doi.org/10.1068/p5027

Schurgin, M. W., Nelson, J., Iida, S., Ohira, H., Chiao, J. Y., & Franconeri, S. L. (2014). Eye movements during emotion recognition in faces. *J Vis*, *14*(13), 14. https://doi.org/10.1167/14.13.14

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychol Sci*, *13*(5), 402-409. https://doi.org/10.1111/1467-9280.00472

Shah, P., Gaule, A., Sowden, S., Bird, G., & Cook, R. (2015). The 20-item prosopagnosia index (PI20): a self-report instrument for identifying developmental prosopagnosia. *R Soc Open Sci*, *2*(6), 140343. https://doi.org/10.1098/rsos.140343

Stacchi, L., Ramon, M., Lao, J., & Caldara, R. (2019). Neural Representations of Faces Are Tuned to Eye Movements. *J Neurosci*, *39*(21), 4113-4123. https://doi.org/10.1523/JNEUROSCI.2968-18.2019

Stacey, P. C., Walker, S., & Underwood, J. D. (2005). Face processing and familiarity: evidence from eye-movement data. *Br J Psychol*, *96*(Pt 4), 407-422. https://doi.org/10.1348/000712605X47422

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behav Res Methods Instrum Comput*, *31*(1), 137-149. https://doi.org/10.3758/bf03207704

Strasburger, H., Rentschler, I., & Juttner, M. (2011). Peripheral vision and pattern recognition: a review. *J Vis*, *11*(5), 13. https://doi.org/10.1167/11.5.13

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Mem Cognit*, *25*(5), 583-592. https://doi.org/10.3758/bf03211301

Tanaka, J. W., & Simonyi, D. (2016). The "parts and wholes" of face recognition: A review of the literature. *Q J Exp Psychol (Hove)*, *69*(10), 1876-1889. https://doi.org/10.1080/17470218.2016.1146780

Thomaz, C. E. (2005). *FEI Face Database.* FEI Face Database. https://fei.edu.br/~cet/facedatabase.html

Van Belle, G., De Graef, P., Verfaillie, K., Busigny, T., & Rossion, B. (2010). Whole not hole: expert face recognition requires holistic perception. *Neuropsychologia*, *48*(9), 2620-2629. https://doi.org/10.1016/j.neuropsychologia.2010.04.034

van Belle, G., Ramon, M., Lefevre, P., & Rossion, B. (2010). Fixation patterns during recognition of personally familiar and unfamiliar faces. *Front Psychol*, *1*, 20. https://doi.org/10.3389/fpsyg.2010.00020

van der Walt, S., Schonberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., Yu, T., & scikit-image, c. (2014). scikit-image: image processing in Python. *PeerJ*, *2*, e453. https://doi.org/10.7717/peerj.453

Visconti di Oleggio Castello, M., & Gobbini, M. I. (2015). Familiar Face Detection in 180 ms. *PLoS One*, *10*(8), e0136548. https://doi.org/10.1371/journal.pone.0136548

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behav Res Methods*, *42*(3), 671-684. https://doi.org/10.3758/BRM.42.3.671

Yarbus, A. L. (1967). Eye Movements During Perception of Complex Objects. In (pp. 171-211). Springer US. https://doi.org/10.1007/978-1-4899-5379-7_8

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception*, *14*(6), 737-746. https://doi.org/10.1068/p140737