RESEARCH ARTICLE

Kun Guo · Robert G. Robertson · Sasan Mahmoodi ·
Yoav Tadmor · Malcolm P. Young

How do monkeys view faces?—a study of eye movements

Abstract Face perception plays a crucial role in primate social communication. We have investigated the pattern of eye movements produced by rhesus monkeys (*Macaca mulatta*) as they viewed images of faces. Eye positions were recorded accurately using implanted eye coils, while neutral upright, inverted and scrambled images of monkey and human faces were presented on a computer screen. The monkeys exhibited a similar eye scan pattern while viewing familiar and unfamiliar monkey face images, or while viewing monkey and human face images. No differences were observed in the distribution of viewing times, number of fixations, time into the trial of first saccade to local facial features, and the temporal and spatial characteristics of viewing patterns across the facial images. However, there was a greater probability of re-fixation of the eye region of unfamiliar faces during the first few seconds of the trial suggesting that the eyes are important for the initial encoding of identity. Indeed, the highest fixation density was found in the eye region of all the face images. The viewing duration and the number of fixations per image decreased when inverted or scrambled faces were presented. The eye region in these modified images remained the primary area of fixation. However, the number of fixations directed to the eyes decreased monotonically from the upright images through the inverted versions to the scrambled face images. Nonetheless, the eyes remain the most salient facial substructure regardless of the arrangement of other features, although the extent of salience which they attain may depend both on the low level properties of the eyes and on the global arrangement of facial features.

K. Guo · S. Mahmoodi · Y. Tadmor · M. P. Young
Department of Psychology, School of Biology,
University of Newcastle upon Tyne,
Newcastle upon Tyne, NE2 4HH, UK

R. G. Robertson (✉)
Department of Psychology, School of Biology,
University of Newcastle upon Tyne,
Newcastle upon Tyne, NE2 4HH, UK
e-mail: r.g.robertson@ncl.ac.uk
Tel.: +44-191-2228642
Fax: +44-191-2225622

Introduction

Visual exploration of the world around us involves a series of saccadic eye movements and fixations (Yarbus 1967). As movements of the eyes re-map the projection of the visual world onto the retina, there is a strong connection between the control of eye movements and the processing of visual inputs. The pattern of eye movements made by a species can, therefore, suggest the goals of vision, and the eye movements themselves have often been studied as indictors of brain mechanisms involved in visual perception, particularly in form and pattern perception (Noton and Stark 1971; Biederman 1987).

In the phylogeny of primates, there is an increasing trend toward larger and more complex social groups in which individuals rely more on visual cues, such as facial signals, than on olfactory cues for communication (Marler 1965). As faces can provide visual information about an individual's gender, age and familiarity, and their expressions provide significant cues to intention and mental state (Bruce and Young 1998; Emery 2000), the ability to recognize these cues and to respond accordingly plays an important role in the social life of higher primates (Andrew 1963; Anderson 1998). Accordingly, the possibility that non-human primates, just like humans, are readily able to perceive differences between individuals of their own species, based on facial cues alone, receives support from numerous behavioural and neurophysiological investigations of face perception in monkeys. At an early age, monkeys respond appropriately to the expressions of other individuals (Mendelson et al. 1982), and they have been shown to be able to recognise the faces of other individuals (Rosenfeld and Van Hoesen 1979). More importantly, they are able to discriminate the

faces of unfamiliar individuals after only a short exposure to sets of their images (Parr et al. 2000). Studies such as these provide compelling evidence that monkeys can recognize the faces of conspecifics based on purely visual cues presented within two-dimensional black-and-white face images.

In humans, the processing of familiar and unfamiliar faces seems to place emphasis on the perception of different facial components. The internal facial features (eyes, nose, mouth) are particularly important in the recognition of familiar faces, whereas the external facial features (hair, face shape) are more salient in the processing of unfamiliar faces (Bruce and Young 1998). In monkeys, neural correlates of face familiarity have been defined by single cell neurophysiology and by recordings of event-related brain potentials (ERPs) (Pineda et al. 1994; Rolls 2000). These observations are consistent with the existence of brain mechanisms that could contribute to the perception of a familiarity distinction with respect to faces.

In the present experiments we have sought to investigate whether the distribution of eye movements in monkeys viewing face images reveals any differences in the salience of faces and their component features according to their familiarity.

Additionally, it remains in dispute whether rhesus monkeys use the same perceptual rules to process monkey and human faces. Several lesion, electrophysiology and brain imaging studies have suggested similarities in the neural mechanisms of face perception between humans and monkeys (for reviews see Bruce and Young 1998; Rolls 2000; Rossion and Gauthier 2002). A characteristic of the perceptual processing of faces in humans is the face inversion effect, which is defined as a larger decrease in recognition performance for faces than for other mono-oriented objects when they are presented upside-down. This effect has also been demonstrated in macaque monkeys (Tomonaga 1994; Parr et al. 1999). However, in the study of Parr et al. (1999), a face inversion effect was observed with the faces of conspecifics, or of capuchin monkeys, or automobiles, but not with human faces, implying differences in the perceptual processing of human faces compared to those of their own species. A similar conclusion has also been drawn from a study of ERPs recorded from macaque monkeys, which revealed waveform differences in latency, amplitude and distribution in response to monkey faces compared to human faces (Pineda and Nava 1993). Therefore, in the present study we have examined monkeys' eye scan patterns to observe whether these reported differences in the perceptual processing of monkey and human faces are manifested in the pattern of visuomotor activity generated during visual inspection of their respective images.

Lastly, it is known that when looking at the faces of conspecifics, the gaze of monkeys is frequently directed to the principal local facial features, especially the eyes, rather than being randomly or evenly distributed across the whole face image (Keating and Keating 1982; Nahm et al. 1997), suggesting that the eye region is more salient,

as it appears to be for humans (Yarbus 1967; McKelvie 1976; Kleinke 1986). However, it remains unclear whether the salience of the eyes is attributable solely to the perception of local properties of the eye, or may derive from the perception of the global relation between local facial features (face configuration). Thus, although face selective responses in the occipitotemporal region of the brain can be activated even when the configuration of inner face components is distorted (Bentin et al. 1996), and may also be activated by isolated face components, particularly the eyes, the timing of these responses is sensitive to the adjustments of the global arrangement of features, which occurs following face inversion (Rossion et al. 2000; Rossion and Gauthier 2002). Therefore, we have also examined whether face inversion has any effect on the salience of faces and their features. However, while face inversion alters the global configuration of facial features, it does so along relatively few axes of symmetry (top-bottom, left-right). More comprehensive disruption of the global arrangement of features is achieved by randomly rearranging the components of a face image. Hence scrambled versions of the same images were presented as a further comparison. In these images major local features, such as the eyes, nose and mouth, were still recognisable, but the relative configuration of features was completely reorganised (see "Materials and methods"). Together these manipulations of the organisation of the face stimuli may indicate whether the salience of the eyes is affected by the context in which they are seen, or whether this can be accounted for by their structure alone.

Investigations of this nature are required to increase our understanding of the relation between the structure of real world stimuli and the organisation of goal-directed eye-movements in non-human primates. This is not only important in its own right, but also for comparison with findings from humans, as the behaviour and neurophysiology of our primate cousins comprises the most significant model for the advancement of research into human brain function.

## Materials and methods

Subjects

Two male adult rhesus monkeys (*Macaca mulatta*, 4.5–6.0 kg) were used in this study. Initially they were trained to fixate a spot on a computer screen for several seconds in a dimming fixation detection task (Guo and Benson 1998). For the purpose of making eye movement recordings, a scleral eye coil and head restraint were then implanted under aseptic conditions. Throughout the period of the recordings, the animal's weight and general health were monitored daily. All procedures complied with the "Principles of laboratory animal care" (NIH publication no. 86–23, revised 1985) and UK Home Office regulations.

Stimuli and apparatus

Visual stimuli were generated using a VSG 2/3W graphics system (Cambridge Research Systems) and displayed on a high frequency

Fig. 1 Examples of face images used in the recording. *Upper left* Upright monkey face, *upper right* upright human face, *bottom left* inverted face, *bottom right* scrambled face. The *white points within the images* indicate the position of each fixation during the image presentation (20 s). Example from one monkey
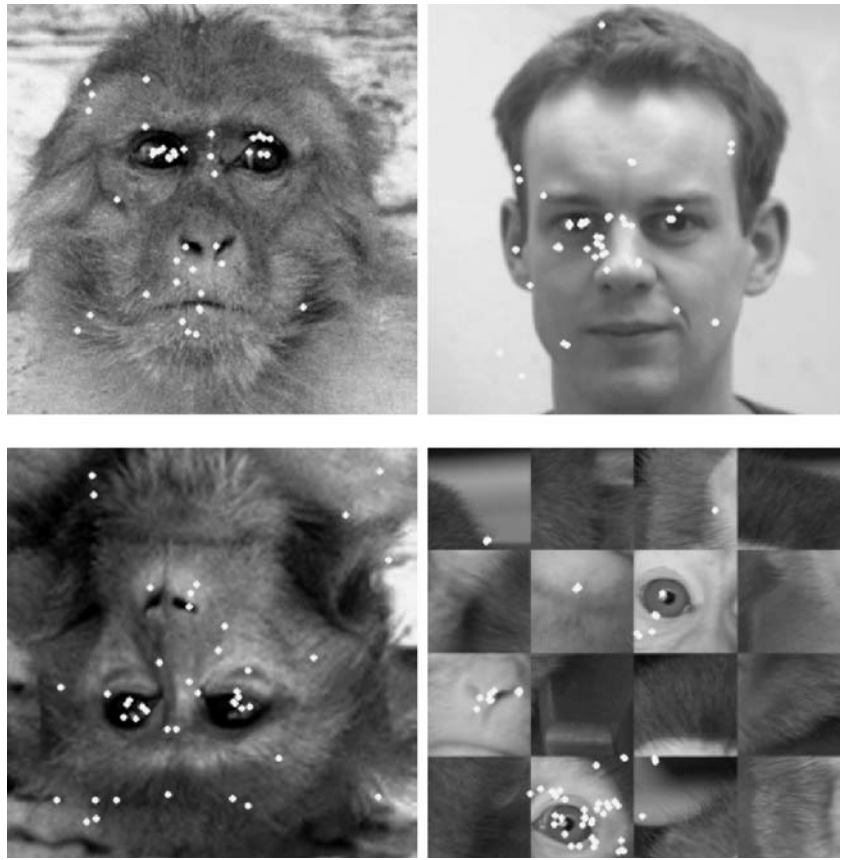
non-interlaced gamma-corrected colour monitor (110 Hz, Sony GDM-F500T9) with the resolution of 1024×768 pixels. At a viewing distance of 57 cm the monitor subtended a visual angle of 40×30°. The mean luminance of the uniform grey background was kept at 6.0 cd/m^2.

Four different classes of digitized neutral face images were used as stimuli (Fig. 1): (1) 20 upright monkey face images, including 10 familiar and 10 unfamiliar face images; the familiar face images were those of other monkeys cohabiting in the same cage; (2) 20 upright human face images; (3) 20 inverted face images (10 human faces and 10 monkey faces all rotated 180° in the plane); and (4) 20 scrambled face images (10 human faces and 10 monkey faces). The scrambled images were generated by dividing each complete face image into a 4×4 matrix and randomly rearranging the parts. By doing so, most of the local facial features (eyes, nose, etc.) were kept intact and recognizable, but the global structure of the face was disrupted. All images were gamma-corrected and displayed once in a random order at the centre of the screen with a resolution of 512×512 pixels (20×20° visual angle).

During the experiments the monkey was seated in a purpose-built primate chair with head restrained, and viewed the display binocularly. To calibrate eye movement signals, a small red fixation point (FP) (0.2° diameter, 7.8 cd/m^2 luminance) was displayed randomly at one of 25 positions (5×5 matrix) across the monitor. The distance between adjacent FP positions was 5°. The monkey was trained to follow the FP and maintain fixation for 1 s. After the calibration procedure, the trial was started with a FP displayed on the centre of the monitor. If the monkey maintained fixation for 500 ms, the FP disappeared and a face image was presented for 20 s. During the presentation, the monkeys passively viewed the images. No reinforcement was given during this procedure, neither were the animals trained on any other task with these stimuli, which could have potentially affected the structure of their behaviour. It was considered that with their lack of training,

and in the absence of instrumental responding, their behaviour should be as natural as possible.

Eye movement recordings and analysis

Horizontal and vertical eye positions were measured using an 18-inch cubic scleral search coil assembly with 6 min arc sensitivity (CNC Engineering). Eye movement signals were amplified by a CNC system and sampled at 500 Hz through the analogue inputs of a CED1401 plus digital interface (Cambridge Electronic Design). The data was then analysed off-line using software developed in Matlab. The software computed horizontal and vertical eye displacement signals as a function of time to determine eye velocity and position. Saccadic eye movements were detected on the basis of their spatiotemporal characteristics. A sample belonged to a saccade if the eye displacement was greater than 0.2° at a velocity of not less than 20 deg/s (Zuber et al. 1965). Samples that did not belong to a saccade were interpreted as a fixation.

We quantified the total amount of time spent looking at each image (face viewing time as percentage of 20-s trial time) and the number of fixations that were detected as a measure of the salience of each image. We also measured the time of the occurrence of the first saccade to each local feature from the start of the trial. The experimental design comprised two levels of image category (familiar faces vs unfamiliar faces; monkey faces vs human faces) and three levels of local features (eyes, nose, mouth) for the first two experiments. The experiment on face structure comprised three levels of face structure (normal, inverted, scrambled). Analysis was carried out after pooling the data from the two animals. Appropriate post hoc testing of differences between levels of face structure was carried out following detection of significant overall F values.

Results

Familiar faces vs unfamiliar faces

Face familiarity did not affect the saliency of the face images (Table 1). No significant difference was observed in the viewing time and number of fixations across the entire set of familiar and unfamiliar faces between the two monkeys (*t*-test, *p*>0.05). Their fixations, however, were not evenly or randomly distributed over the whole face image. The largest proportion of fixations was directed at local facial features, including the eyes, nose and mouth (e.g. Fig. 1). Among these local features, the eyes and surrounding region received a disproportionate share of viewing time (viewing time as percentage of total trial time or as percentage of face viewing time, Table 1) and fixations, considering the relative area they occupy (8.21±0.25%, mean ± SEM, of the face image area). The other parts of the face images were scarcely fixated at all.

No statistical difference was found between the familiar and unfamiliar faces for the proportion of face viewing time (Fig. 2A) and the proportion of fixations within the face images (Fig. 2B) directed at local features (*t*-test, *p*>0.1). The eyes were fixated for 28.93±3.46% of the face viewing time, with 38.36±3.77% of the total number of fixations in the familiar faces, and for 30.58±2.55% of the face viewing time with 44.40±3.48% of the total fixations in the unfamiliar faces. The nose or mouth region, in contrast, attracted less than 7.5% of face viewing time with less than 10.5% of fixations (Table 1) (viewing time: local features, $F_{(2,97)}=77.37$, $p<0.01$; image category, $F_{(1,97)}=0.79$, $p>0.1$; fixations: local features, $F_{(2,97)}=95.68$, $p<0.01$; image category, $F_{(1,97)}=0.87$, $p>0.1$).

At the start of each trial the monkey's first saccade was directed to the eyes with a high probability, although often they were not the closest feature to the initial fixation point. To quantitatively compare the sequence of saccade destinations across different images, we measured the time into the trial of the occurrence of the first saccade to the eyes, nose and mouth regions for each tested image. No statistical difference was observed between the familiar and unfamiliar faces (*t*-test, *p*>0.2, Table 1). The first saccade to the eyes in both familiar and unfamiliar faces clearly occurred at the shortest time into the trial (444±114 and 341±48 ms, respectively), while the times into the trial of the first saccades to the nose and mouth regions were much longer (Fig. 2C; significant main effects for local features: $F_{(2,97)}=13.62$, $p<0.01$; image category: $F_{(1,97)}=2.54$, $p>0.1$).

Analysis of the fixation duration and size of the saccades for each image was also carried out. While viewing the images, the monkeys made frequent short saccades with a mean value of 7.84±0.33° and 7.85±0.29° for the familiar and unfamiliar faces, respectively (Fig. 3A), and the durations for individual fixations were distributed around 250 ms with mean values of 308±14 and 268±8 ms, respectively (Fig. 3B). No significant
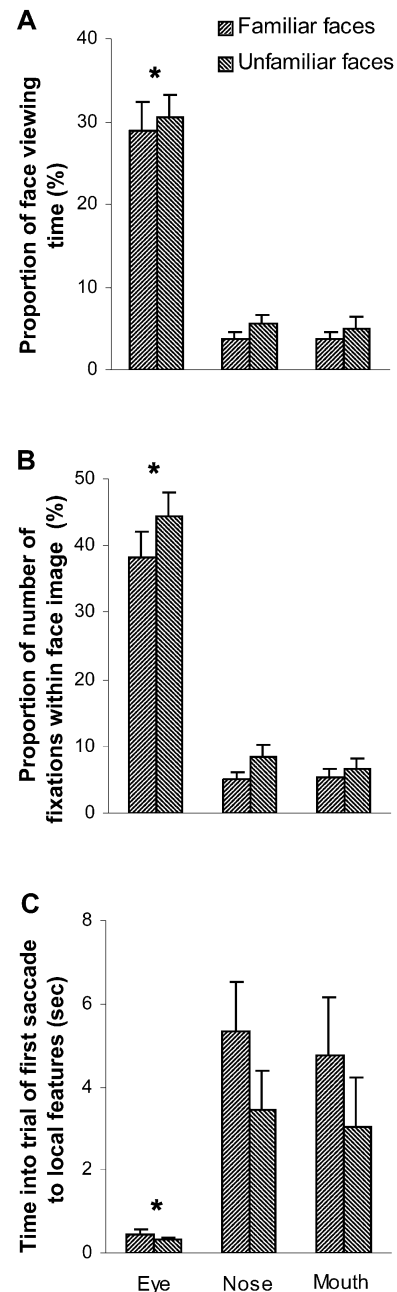

Fig. 2 Proportion of cumulative face viewing time (**A**) and number of fixations (**B**) within eye, nose and mouth regions, and time into the trial of occurrence of first saccade to these local facial features (**C**) for familiar and unfamiliar monkey face images. *Error bars* indicate standard error of the mean. *Asterisks* indicate the values that were significantly different from corresponding values of other local features (Tukey's least significant difference procedure, *p*<0.01)

difference was observed for the distribution of saccade sizes and fixation durations between the two face categories (Kolmogorov-Smirnov test, *p*>0.1).

It appears that during the entire 20-s presentation, the monkeys generated similar eye scan patterns to examine both familiar and unfamiliar faces. However, in order to determine whether there were any differences in the

Table 1 Cumulative viewing times and number of fixations for face images and local facial features, and time into the trial of first saccade to local facial features. Values presented in the table are mean ± SEM

|	Familiar faces	Unfamiliar faces	t-test (p)	Monkey faces	Human faces	t-test (p)	Upright faces	Inverted faces	Scrambled faces	ANOVA (p)
Viewing time as percentage of total trial time (%)										
Whole face	60.70±5.32	66.29±3.76	0.40	65.37±3.35	69.40±3.76	0.43	70.11±3.22	52.19±3.64	56.20±3.34	0.0008
Eye	18.41±3.42	20.50±2.37	0.62	19.68±2.02	18.48±2.21	0.73	20.52±1.87	13.78±1.84	9.37±1.38	5.53E-5
Nose	2.12±0.49	3.51±0.72	0.11	3.09±0.46	2.67±0.54	0.63	–	–	–	–
Mouth	3.32±0.89	2.13±0.49	0.25	2.84±0.51	3.15±0.65	0.70	–	–	–	–
Viewing time as percentage of face viewing time (%)										
Eye	28.93±3.46	30.58±2.55	0.62	29.40±2.10	25.97±2.23	0.26	28.92±1.94	26.64±2.60	17.63±2.51	0.002
Nose	3.73±0.87	5.54±1.14	0.21	4.93±0.72	4.05±0.83	0.42	–	–	–	–
Mouth	3.68±0.83	4.97±1.36	0.42	4.12±0.67	4.61±1	0.68	–	–	–	–
Number of fixations for whole face image										
	35.20±2.02	39.90±3.03	0.16	35.30±2.15	36.60±2.08	0.74	38.89±2.30	25.53±1.99	27.81±1.82	1.69E-5
Number of fixations within local facial features as percentage of number of fixations within whole face image (%)										
Eye	38.36±3.77	44.40±3.48	0.25	39.94±2.35	35.21±2.10	0.07	39.25±2.68	37.41±3.57	22.72±2.99	0.0004
Nose	5.13±0.99	8.4±1.91	0.14	7.85±1.22	5.4±0.94	0.12	–	–	–	–
Mouth	5.47±1.06	6.51±1.57	0.58	6.6±1.06	6.32±1.09	0.85	–	–	–	–
Time into the trial of occurrence of first saccade to local facial feature (ms)										
Eye	444±114	341±48	0.41	386±59	374±65	0.90	305±30	592±111	1655±327	3.67E-5
Nose	5325±1195	3456±917	0.23	4074±721	5560±998	0.22	–	–	–	–
Mouth	4739±1424	3024±1213	0.37	3796±873	4271±909	0.71	–	–	–	–
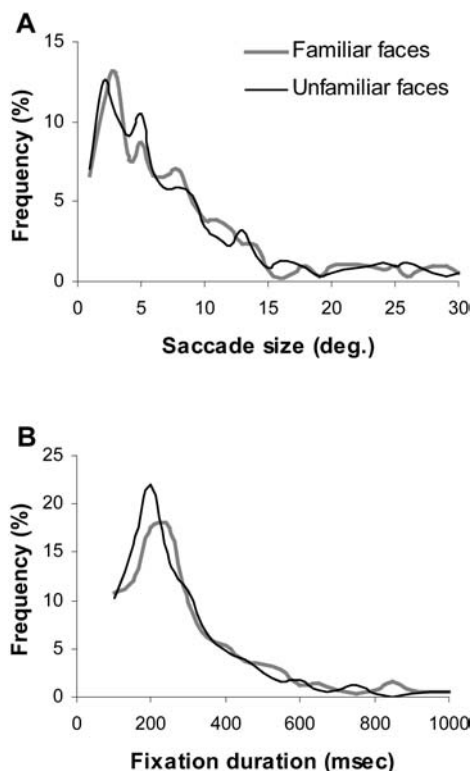
Fig. 3 Frequency distributions of individual saccade size (**A**) and fixation duration (**B**) measured while viewing familiar and unfamiliar monkey face images

Fig. 4 The probability of the destination of the first ten saccades measured while viewing familiar and unfamiliar monkey face images

distribution of saccades during the time course of viewing the images, we compared the timing and destination of the first ten saccades in each sequence (this number was chosen as it represented the maximum number of saccades for some images). The probability of saccade distribution to different local facial features is plotted in Fig. 4. The eyes had a much higher probability as the first saccade destination (90%) compared to other local elements. For the next four saccades, the eye region in the unfamiliar faces received more fixations (65±0.00%) than in the familiar faces (35±7.08%), while the forehead region in the familiar faces (28.75±4.33%) had slightly more fixations than that of the unfamiliar faces (13.75±5.46%) (Kolmogorov-Smirnov test, $p<0.05$). At later stages, no statistical difference was observed (Kolmogorov-Smirnov test, $p>0.1$). The other local features in the familiar and unfamiliar faces had the same chance to be the next saccade destination (Kolmogorov-Smirnov test, $p>0.1$).

No significant difference was observed in the duration of the first ten fixations directed to each of these features ($F_{(6,386)}=1.65$, $p>0.1$), nor between the familiar and unfamiliar faces ($F_{(1,386)}=0.71$, $p>0.1$), nor between the saccade sequences ($F_{(9,390)}$, $p>0.05$).
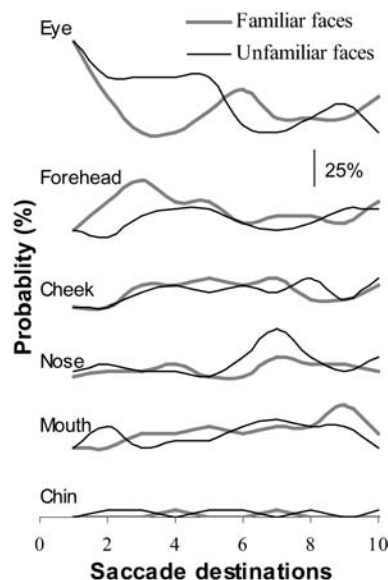
Monkey faces vs human faces

The upright monkey and human face images appeared equally salient to the monkeys. No difference was observed in the viewing time and number of fixations across the image categories (*t*-test, $p>0.4$; Table 1). In both monkey and human faces, gaze directed at the eyes comprised the largest component of face viewing time (Fig. 5A, local features, $F_{(2,204)}=115.6$, $p<0.01$); the largest number of fixations (Fig. 5B, local features, $F_{(2,204)}=33.23$, $p<0.01$), and the shortest time into the trial of the first saccade (Fig. 5C, local features, $F_{(2,204)}=29.51$, $p<0.01$).

The proportion of the total trial time or face viewing time (Fig. 5A, image category, $F_{(1,204)}=0.27$, $p>0.1$) or number of fixations directed to each feature (image category, $F_{(1,204)}=1.02$, $p>0.1$) did not differ between monkey and human face images; neither did the time into the trial of the first saccade to each feature (Fig. 5C, image category, $F_{(1,204)}=1.51$, $p>0.1$). Fixations of the eye region in the monkey images contributed 29.4±2.1% of the face viewing time and 39.94±2.35% of fixations, while fixations of the eye region in the human images contributed 25.97±2.23% of the face viewing time and 35.21±2.1% of fixations. In contrast, the nose and mouth attracted much less viewing time (<7.5%) and fewer fixations (<9.5%, Fig. 5A, B). In the saccade sequences, the times into the trial of the first saccade to the eyes in the monkey and human face images were 386±59 and 374±65 ms, respectively, while the times into the trial to the nose and mouth regions were much longer (>3700 ms, Fig. 5C).

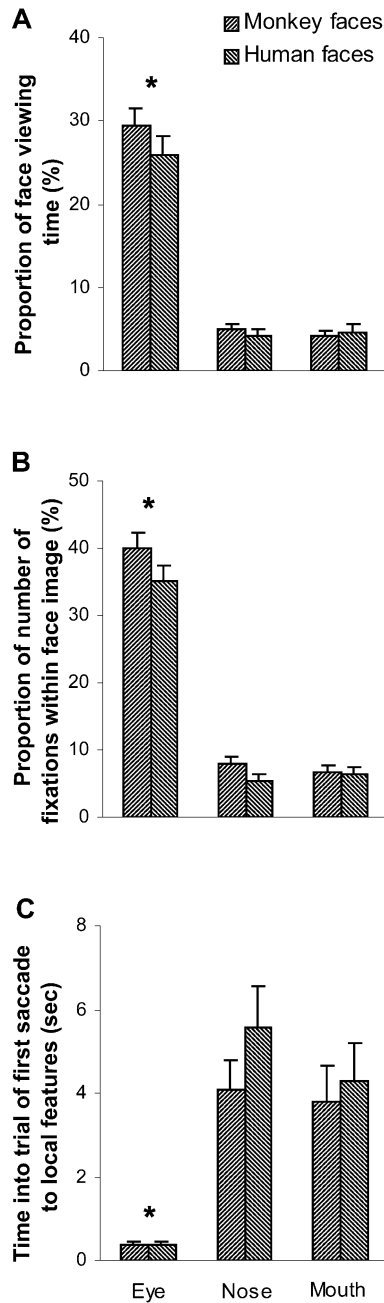Other studies make the claim that the characteristics of the visual scene can influence the frequency and size of

Fig. 6 Frequency distributions of individual saccade size **(A)** and fixation duration **(B)** measured while viewing upright monkey and human face images

Fig. 5 The proportion of cumulative face viewing time **(A)** and number of fixations **(B)** within eye, nose and mouth regions, and time into the trial of occurrence of first saccade to these local facial features **(C)** for upright monkey and human face images. *Error bars* indicate standard error of the mean. *Asterisks* indicate the values that were significantly different from the corresponding values of other local features (Tukey's least significant difference procedure, *p*<0.01)

distributions of fixation durations and the size of each saccadic eye movement under these conditions. The distributions of saccade size and fixation duration for monkey and human face images were very similar (Kolmogorov-Smirnov test, saccade size: ks=0.2, *p*>0.1; fixation duration: ks=0.16, *p*>0.1). The peaks of the frequency distribution of saccade size (Fig. 6A) and fixation duration (Fig. 6B) were 3° and 200 ms. The mean saccade sizes were 7.7±0.21° and 7.08±0.18° for the monkey and human face images, and the mean fixation durations were 294.16±8.68 and 302.23±6.7 ms.

Upright faces vs inverted faces and scrambled faces

The upright faces attracted longer viewing times ($F_{(2,105)}$= 7.64, *p*<0.01) and more fixations ($F_{(2,105)}$= 12.22, *p*<0.01) than the inverted or scrambled faces. The two monkeys spent 70.11±3.22% of the 20-s image presentation time viewing the upright faces, making 38.89±2.3 fixations across the images. The proportion of time spent viewing the image decreased to 52.19±3.64% and 56.20±3.34%, and the number of fixations declined to 25.53±1.99 and 27.81±1.82 for the inverted and scrambled face images, respectively. No significant difference was observed between the inverted and scrambled face images (Tukey's honestly significant difference, *p*>0.05).

human saccadic eye movements (Kowler et al. 1992; Andrews and Coppola 1999). In order to determine whether the temporal and spatial characteristics of the saccades of monkeys, viewing the faces of other conspecific monkeys, differ from the characteristics of saccades when viewing the faces of humans, we calculated the
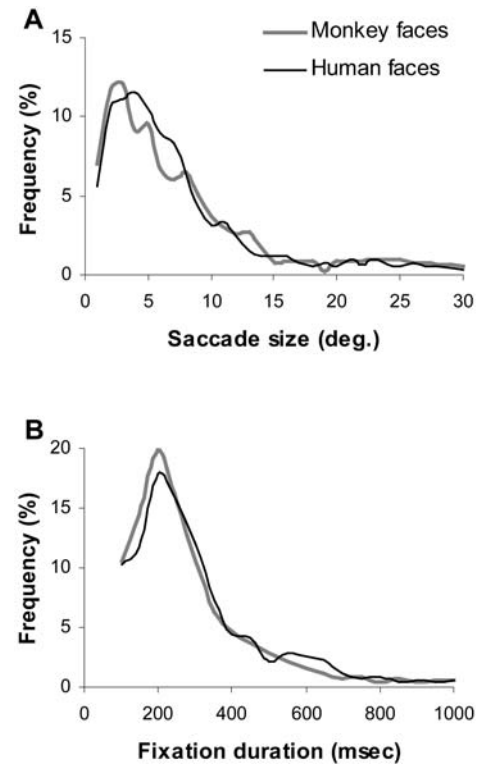
Fig. 7 Cumulative viewing time (**A**) and number of fixations (**B**) for the eye region, and the time into the trial of the occurrence of the first saccade to the eye region (*Time into trial to eyes*) (**C**) within upright, inverted and scrambled face images (mean ± SEM); the values in each image category were statistically different from all other image categories (Tukey's least significant difference procedure, *p*<0.05). **D, E** Cumulative viewing time (**A**) and number of fixations (**B**) for the eye region as percentages of the face viewing time (mean ± SEM); values for scrambled images significantly different from upright and inverted faces (Tukey's least significant procedure, *p*<0.05)

As in the upright images, the eyes and surrounding region in the inverted and scrambled images received the highest proportion of fixations. The cumulative viewing time (Fig. 7A) and the number of fixations (Fig. 7B) was decreased in these two conditions (viewing time: upright 4.10±0.37 s, inverted 3.21±0.4 s, scrambled 1.87±0.28 s, $F_{(2,105)}$=10.78, *p*<0.01; number of fixations: upright 15.69±1.66, inverted 11.83±1.61, scrambled 6.31±0.96, $F_{(2,105)}$=13.46, *p*<0.01). However, the reduced probability of eye region viewing for the inverted faces appeared to be due to the reduced accumulated viewing duration for these face images; this difference is not apparent when expressed as a percentage of face viewing time; the proportion of eye region viewing remains significantly reduced for the scrambled condition only [Fig. 7D, upright 28.92±1.94%, inverted 26.64±2.6%, scrambled 17.63±2.51% ($F_{(2,105)}$=6.35, *p*<0.01; scrambled significantly different from other two groups, Tukey's least significant difference, *p*<0.01)] (Table 1). Accordingly, the proportion of the number of fixations of the eye region for the reorganised images was only reduced in the scrambled condition (Fig. 7E, upright 39.25± 2.68%, inverted 37.41±3.57%, scrambled 22.72±2.99% ($F_{(2,105)}$=8.54, *p*<0.01; scrambled significantly different from the other two groups, Tukey's least significant difference, *p*<0.01)] (Table 1).

The time into the trial of the first saccade to the eye region in the reorganised images was also delayed (Fig. 7C). The mean time into the trial increased from
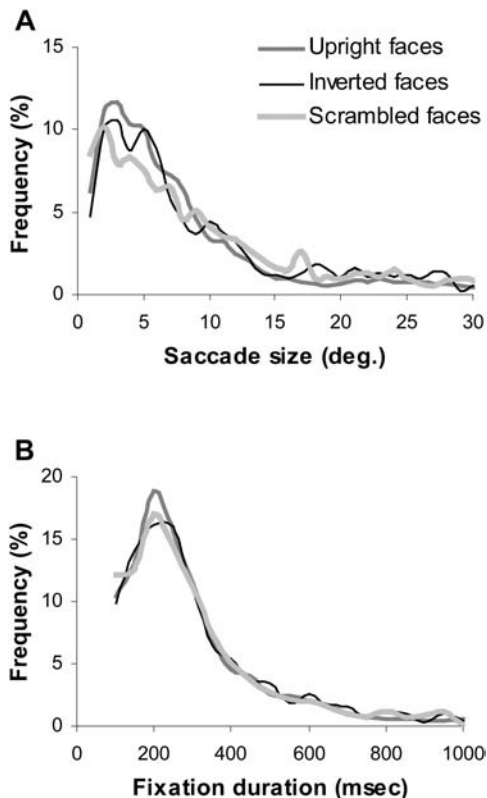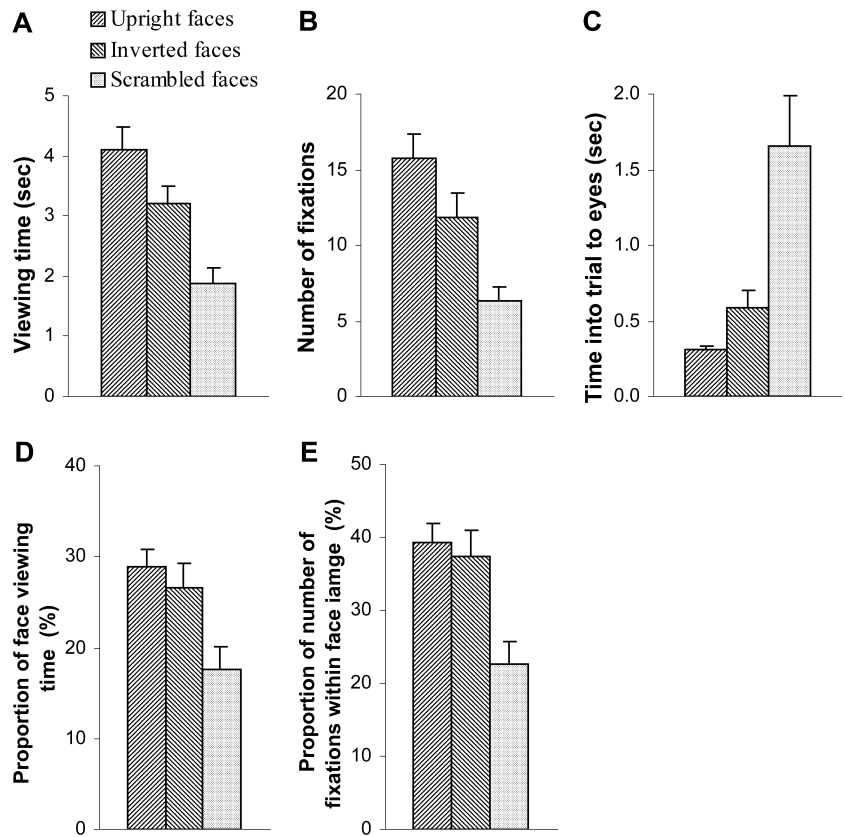
Fig. 8 Frequency distributions of individual saccade size (**A**) and fixation duration (**B**) measured while viewing upright, inverted and scrambled face images

305±30 ms in the upright images to 592±111 ms in the inverted versions and 1655±327 ms in the scrambled versions ($F_{(2,105)}$=11.27, $p<0.01$).

There were no significant differences in the size of saccadic eye movements and the duration of fixations between the three image categories (Kolmogorov-Smirnov test, $p>0.1$). The peaks of the frequency distribution of saccade size (Fig. 8A) and fixation duration (Fig. 8B) were 3° and 200 ms, respectively. The mean saccade sizes for the upright, inverted and scrambled faces were 7.95±0.2°, 9.0±0.29° and 9.63±0.3°, and the mean fixation durations were 295.68±7.28, 305.83±8.7 and 305.2±8.54 ms. The temporal and spatial characteristics of saccadic eye movements, therefore, remained the same for all three types of face image.

Discussion

Eyes are salient features in face images

When humans observe a complex scene, instead of distributing their attention evenly across the entire scene, their fixations are restricted to information rich details. The scene, therefore, can be divided into informative features and redundant regions (Mackworth and Morandi 1967; Yarbus 1967; Krieger et al. 2000). Generally, informative features include angles and line ends when dealing with simple patterns, and the intricate, unpredictable details or curved boundaries in more complex patterns (Attneave 1954; Mackworth and Morandi 1967; Yarbus 1967; Antes 1974). In pictures of faces, the eyes and surrounding region are the informative features that attract most of the observer's attention (McKelvie 1976). Early eye movement recordings showed that humans fixate mainly on the eyes while viewing a face picture (Yarbus 1967). This preference for attending to the eyes appears to develop very early; even 2-month-old infants display a preference for looking at the eyes compared to other regions of the face (Maurer 1985).

Our findings confirm that the visual system of monkeys also appears to be tuned to these features in the faces of other monkeys and in the faces of humans. The largest proportion of face viewing time (>25%) and the number of fixations (>35%) were devoted to viewing the eyes regardless of the face familiarity or of the species. This is consistent with evidence for the critical role of eyes in face perception and social cognition (Emery 2000). Similar findings have also been reported for rhesus monkeys (Nahm et al. 1997) and baboons (Kyes and Candland 1987), both of which demonstrated an exaggerated interest in the eye region of the faces of conspecifics.

The eyes were usually the first fixation target following the appearance of the upright face stimuli. This is consistent with the observation from human studies that the informative regions of an image appear to be identified at the earliest viewing stage (Mackworth and Morandi 1967; Antes 1974) and knowledge about the setting of the main objects in an image is present during the first fixation (Biederman et al. 1982).

This preferential interest in the eyes during face perception may be linked to monkey ecology. Naturally living in large social groups, monkeys have sophisticated social systems that are reliant on visual communication, and, as with humans, facial signals are an important means of communication (Marler 1965; Emery 2000). Behavioural studies have shown that the eyes, eyelids and eyebrows can provide at least 13 different facial expressions for catarrhine primates (van Hoof 1967), suggesting that the eye region plays a pivotal role in facial expression of non-human primates and provides individuals with a means of evaluating one another's current intentions (Jolly 1972; Redican 1975).

Alternatively, as perception of face identity in humans is sensitive to the manipulation of the metric properties of a face (Cooper and Wojan 1996), gauging the position of the eyes relative to other facial features may provide crucial information for face perception. Indeed, rhesus monkeys have considerable difficulty in recognising the faces of other monkeys when the eyes are masked from the stimulus (Parr et al. 2000). While this may indicate the presence of important cues within the structure of the eyes themselves, the use of the eyes as critical facial landmarks for gauging facial dimensions should not be overlooked, as these parameters are explicitly encoded in the brain and may be used for face discrimination (Young and Yamane 1992; Sugase et al. 1999).

Comparison of familiar face and unfamiliar face images

We observed that the monkeys used similar oculomotor strategies to examine both familiar and unfamiliar monkey faces. No statistical difference was observed in terms of viewing time, number of fixations and the time into the trial of the first saccade to the component features between the two categories. Among the prominent facial features, the eyes and surrounding region attracted the most fixations, indicating that the eyes are the most salient facial structures. Indeed, facial feature-masking suggests that the eyes play a critical role in individual face recognition (Parr et al. 2000), as they do in humans (McKelvie 1976).

Analysis of the earliest saccade destinations revealed a difference between the processing of familiar and unfamiliar faces. After the first saccade to the eye region, there was a much greater probability that the eyes would be the targets for the next few saccades in the unfamiliar faces, whereas the forehead region was the most probable target in the familiar faces. Apparently, the cognitive processing of unfamiliar faces stimulates more frequent inspection of the eye region initially. The fixation density at any region of a scene appears to correspond to the informativeness of the region (Mackworth and Morandi 1967; Antes 1974), and hence refixation of the eyes following the appearance of the unfamiliar faces may be an important mechanism to promote incorporation of this

new information into memory. Young and Yamane (1992) demonstrated that information regarding the width of the face and between the eyes and the hairline of individuals in a set of faces of Japanese males was encoded by a population of face selective neurones in the inferior temporal cortex of monkeys. This observation suggests that there may be a connection between the encoding of spatial dimensions used in face perception and the generation of a saliency map to guide eye movements during face viewing.

In their influential model of face recognition, Bruce and Young (1986) suggested that face recognition starts with a perceptual structural encoding stage, where local facial features and their spatial relationships are analysed. This is followed by a recognition stage, where structural representations are compared with stored face representations. When a positive match is achieved, person identity nodes in semantic memory are activated, resulting in face identification. Recently, evidence for possible neurophysiological correlates of these concepts has been obtained from human ERP recordings (Bentin and Deouell 2000; Eimer 2000). A short latency negative potential (N170) is not affected by face familiarity, and may correspond to the structural encoding stage; whereas longer latency negative (N400) and positive potentials (P600) are enhanced when a face becomes familiar, suggesting that the activity responsible for these signals is associated with the recognition stage. An enhancement of the amplitude of the N400 has also been recorded from monkeys as they viewed familiar monkey faces (Pineda et al. 1994). In the present experiment, following the stimulus onset, the first saccade was usually directed to the eyes regardless of the face familiarity. However, the next few saccade destinations were distributed to different parts of the face images according to the familiarity of the faces. This dynamic of saccade destinations is likely to be correlated with the above neurophysiological measurements.

The perirhinal cortex appears to have a critical role in signalling the familiarity of objects (Brown and Bashir 2002), the neurons in this region being tuned to the novelty of specific objects. Neurons in this region should provide a strong signal indicating that a given face is unfamiliar, and this information may increase the salience of the eyes during the early stage of viewing, and assist the encoding of identity into memory.

Comparison of monkey face and human face images

The spatial configuration of monkey faces is similar to that of human faces (Carmel and Bentin 2002). They also share some morphological similarities in the evolution of certain facial expressions. The relaxed open-mouth face and the bared-teeth display in macaques have been proposed to be homologous with laughter and smiling in humans (Preuschoft 1992). Therefore, it is not unexpected that the visual system of monkeys is tuned to the same features in the faces of monkeys and humans. Indeed, the monkeys employed the same oculomotor strategy to inspect monkey and human faces, distributing fixations to the whole face image and major local facial features in a similar manner for both species. Comparison of the data for the human and monkey faces did not reveal any differences in the measures that we chose to quantify. These results suggest that the monkey visual system employs a common visuomotor mechanism to parse images of monkey and human faces. This conclusion is reinforced by the observation that the effects of inversion and scrambling on the salience of the images and the evoked pattern of eye movements were equivalent for both the monkey and the human faces (see below).

Comparison of upright, inverted and scrambled face images

Recognition of face identity is considerably more difficult for human subjects when the face is inverted (Valentine 1988; Rossion and Gauthier 2002). Evidence suggests that it is the ability to perceive the spatial relationships between facial features that is disrupted by the inversion (Searcy and Bartlett 1996; Leder and Bruce 1998). Rearranging the local or global arrangement of a face image (i.e. the scrambled faces in our experiment) can make identity even harder to recognize. Apparently, the brain mechanisms for face recognition work optimally with an upright face. Once the image is inverted or scrambled, the brain is deprived of the information regarding the global structure of the face, which informs efficient identity judgments. There have been several attempts to determine whether face recognition in non-human primates is also subject to a face inversion effect, but the results have been inconsistent. While several studies claim to have found a face inversion effect in various species of macaque, including long-tailed macaques (Dittrich 1990), pig-tailed macaques (Swartz 1983) and Japanese macaques (Tomonaga 1994), others have found the effect to be absent in either long-tailed macaques (Bruce 1982) or rhesus monkeys (Rosenfeld and Van Hoesen 1979). These different outcomes have been attributed to inconsistencies in the methodologies employed by each study, this being exacerbated by the absence of non-face control stimuli throughout. Attempting to address these issues, Parr et al. (1999) provided evidence in favour of an inversion effect in rhesus monkeys, but found that the effect was not specific to faces.

Interestingly, ERP studies in humans have demonstrated that the peak latency of the N170 component, which might represent the activation of brain processes related to face perception, is significantly delayed following the presentation of inverted faces or isolated facial features, such as the eyes (e.g. Bentin et al. 1996; Taylor et al. 2001; Rossion and Gauthier 2002), and similar results have been obtained with macaques (Pineda and Nava 1993).

In the present experiments the reduced viewing times and reduced number of fixations for the inverted and scrambled face images suggest that these images are less salient than the upright images. The observation that the scrambled images appear to be no less salient than the inverted faces may be accounted for by assuming that the scrambled images still retain meaningful components of the original images.

The upright versions of the images were more salient than the inverted or scrambled versions. The actual proportion of viewing time that the gaze was directed to the eye region of the inverted faces was equivalent to that found in the upright images (*viewing time as a percentage of face viewing time*, Table 1); this was also reflected in the proportion of fixations on the eyes as a percentage of the fixations within the face image (Table 1). Thus, the eyes in the images remained the most salient features despite their new spatial relationship with other face features. However, the eyes in the scrambled face images were much less salient than the eyes in either the upright or the inverted images, suggesting that the saliency of the eyes is affected by the context in which they are embedded. This interpretation is affirmed by the observation that the time into the trial of the first saccade to the eye region is also sensitive to the context in which the eyes appear. Thus, it is not clear from the present results whether the time into the trial of the occurrence of the first saccade to the eyes (Table 1) is a consequence of the reduced salience of the eyes themselves or of the two classes of images. The viewing time as a percentage of total trial time (Table 1) shows that the saliencies of the inverted and scrambled faces are approximately equivalent, suggesting that the increased latency for looking at the eyes in the scrambled images is attributable to the reduced coherency of the face in these images.

What causes the loss of salience of the face following inversion? One account of the effects of face inversion claims that the effect on behaviour arises from the strength of the signal originating from neurons in the temporal cortex that are tuned to different views of faces (Perrett et al. 1998). According to this idea fewer neurons are tuned to the inverted face images as the monkeys are probably not used to viewing faces in this orientation. Furthermore, it has recently been demonstrated that decisions concerning the selection of visual targets for eye movements are encoded by neurons in the posterior parietal cortex (Platt and Glimcher 1999), and prior knowledge concerning the likelihood of the occurrence of a stimulus may play a significant role in this process (Carpenter and Williams 1995). According to this explanation, selection of the face as a target for fixation would depend on the prior probability of observing a face in a given orientation. This would be consistent with both the observed increase in the time into the trial for the first saccade directed at the eyes and the reduced salience of the inverted faces and scrambled faces. Yet, the fact that the eyes remain the most salient targets for fixation, even for the scrambled face images, suggests either that their intrinsic structure (e.g. local contrast or local edges) may

contribute to their salience, or that the visual system retains prior knowledge of the occurrence of "eyes" within the context of a face from past experience (Bentin et al. 2002).

These results imply that eye movements in non-human primates, while viewing faces, are subject to control by more than a single level of perceptual processing. The fact that the eyes attain such a high degree of salience may ultimately derive from their biological significance, particularly their role in social communication. However, the mechanisms that control the targeting of the eye region may occur at a relatively low level of visual processing, i.e. prior to the level of object classification. At the same time, the probability that the eye region will become the target of a saccade is evidently affected by higher levels of perceptual processing. Thus, the eyes are more likely to be the target of a saccade if they occur within their normal context. Additionally, if a face has been classified as being unfamiliar to the viewer, then the probability of the eye region becoming the target of a saccade is again increased.

References

Anderson JR (1998) Social stimuli and social rewards in primate learning and cognition. Behav Process 42:159–175

Andrew RJ (1963) Evolution of facial expressions. Science 142:1034–1041

Andrews TJ, Coppola DM (1999) Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. Vision Res 39:2947–2953

Antes JR (1974) The time course of picture viewing. J Exp Psychol 103:62–70

Attneave F (1954) Some informational aspects of visual perception. Psychol Rev 61:183–193

Bentin S, Deouell LY (2000) Structural encoding and identification in face processing: ERP evidence for separate mechanisms. Cogn Neurophysiol 17:35–54

Bentin S, Allison T, Puce A, Perez A, McCarthy G (1996) Electrophysiological studies of face perception in human. J Cogn Neurosci 8:551–565

Bentin S, Sagiv N, Mecklinger A, Friederici A, von Cramon YD (2002) Priming visual face-processing mechanisms: electrophysiological evidence. Psychol Sci 13:190–193

Biederman I (1987) Recognition-by-components: a theory of human image understanding. Psychol Rev 94:115–147

Biederman I, Mezzanotte RJ, Rabinowitz JC (1982) Scene perception: detecting and judging objects undergoing violation. Cognit Psychol 14:143–177

Brown MW, Bashir ZI (2002) Evidence concerning how neurons of the perirhinal cortex may effect familiarity discrimination. Philos Trans R Soc Lond B Biol Sci 357:1083–1095

Bruce C (1982) Face recognition by monkeys: absence of an inversion effect. Neuropsychologia 20:515–521

Bruce V, Young A (1986) Understanding face recognition. Br J Psychol 77:305–327

Bruce V, Young A (1998) In the eye of the beholder. Oxford University, New York

Carmel D, Bentin S (2002) Domain specificity versus expertise: factors influencing distinct processing of faces. Cognition 83:1–29

Carpenter RH, Williams ML (1995) Neural computation of log likelihood in control of saccadic eye movements. Nature 377:59–62

Cooper EE, Wojan TJ (1996) Differences in the coding of spatial relations in faces and objects. Invest Ophthalmol Vis Sci 37:177

Dittrich W (1990) Representation of faces in longtailed macaques (*Macaca fascicularis*). Ethology 85:265–278

Eimer M (2000) Event-related brain potentials distinguish processing stages involved in face perception and recognition. Clin Neurophysiol 111:694–705

Emery NJ (2000) The eyes have it: the neuroethology, function and evolution of social gaze. Neurosci Biobehav Rev 24:581–604

Guo K, Benson PJ (1998) Involuntary eye movements in response to first- and second-order motion. Neuroreport 9:3543–3548

Henderson JM, Hollingworth A (1999) High-level scene perception. Annu Rev Psychol 50:243–271

Jolly A (1972) The evolution of primate behavior. Macmillan, New York

Keating CF, Keating EG (1982) Visual scan patterns of rhesus monkeys viewing faces. Perception 11:211–219

Kleinke CL (1986) Gaze and eye contact—a research review. Psychol Bull 100:78–100

Kowler E, Pizlo Z, Zhu G, Erkelens CJ, Steinman RM, Collewijn H (1992) Coordination of head and eyes during the performance of natural (and unnatural) visual tasks. In: Berthoz A, Vidal PP, Graf W (eds) The head-neck sensory motor system. Oxford University, Oxford, pp 419–426

Kreiger G, Rentschler I, Hauske G, Schill K, Zetsche C (2000) Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics. Spat Vis 13:201–214

Kyes RC, Candland DK (1987) Baboon (*Papio hamadryas*) visual preferences for regions of the face. J Comp Psychol 101:345–348

Leder H, Bruce V (1998) Local and relational aspects of face distinctiveness. Q J Exp Psychol A 51:449–473

Mackworth NH, Morandi AJ (1967) The gaze selects informative details within pictures. Percept Psychophys 2:547–552

Marler P (1965) Communication in monkeys and apes. In: DeVore I (ed) Primate behavior. Holt, Rinehart & Winston, New York, pp 544–584

Maurer D (1985) Infant's perception of facedness. In: Field T, Fox N (eds) Social perception in infants. Ablex

McKelvie SJ (1976) The role of eyes and mouth in the memory of a face. Am J Psychol 89:311–323

Mendelson MJ, Haith MM, Goldman-Rakic PS (1982) Face scanning and responsiveness to social cues in infant rhesus monkeys. Dev Psychol 18:222–228

Nahm FKD, Perret A, Amaral DG, Albright TD (1997) How do monkeys look at faces? J Cogn Neurosci 9:611–623

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

Parr LA, Winslow JT, Hopkins WD (1999) Is the inversion effect in rhesus monkeys face-specific? Anim Cogn 2:123–129

Parr LA, Winslow JT, Hopkins WD (2000) Recognizing facial cues: individual discrimination by chimpanzees (*Pan troglo-dytes*) and rhesus monkeys (*Macaca mulatta*). J Comp Psychol 114:1–14

Perrett DI, Oram MW, Ashbridge E (1998) Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. Cognition 67:111–145

Pineda JA, Nava C (1993) Event-related potentials in macaque monkey during passive and attentional processing of faces in a priming paradigm. Behav Brain Res 53:177–187

Pineda JA, Sebestyen G, Nava C (1994) Face recognition as a function of social attention in nonhuman primates—an ERP study. Cognit Brain Res 2:1–12

Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. Nature 400:233–238

Preuschoft S (1992) "laughter" and "smile" in barbary macaques (*Macaca sylvanus*). Ethology 91:220–236

Redican WK (1975) Facial expression in nonhuman primates. In: Rosenblum LA (ed) Primate behaviour: developments in field and laboratory research. Academic, New York, pp 104–194

Rolls ET (2000) Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. Neuron 27:205–218

Rosenfeld SA, Van Hoesen GW (1979) Face recognition in the rhesus monkey. Neuropsychologia 17:503–509

Rossion B, Gauthier I (2002) How does the brain process upright and inverted faces? Behav Cognit Neurosci Rev 1:63–75

Rossion B, Gauthier I, Tarr MJ, Despland P, Bruyer R, Linotte S, Crommelinck M (2000) The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. Neuroreport 11:69–74

Searcy JH, Bartlett JC (1996) Inversion and processing of component and spatial-relational information in faces. J Exp Psychol Hum Percept Perform 22:904–915

Sugase Y, Yamane S, Ueno S, Kawano K (1999) Global and fine information coded by single neurons in the temporal visual cortex. Nature 400:869–873

Swartz KB (1983) Species discrimination in infant pigtail monkeys with pictorial stimuli. Dev Psychobiol 16:219–231

Taylor MJ, Itier RJ, Allison T, Edmonds GE (2001) Direction of gaze effects on early face processing: eyes-only versus full faces. Cognit Brain Res 10:333–340

Tomonaga M (1994) How laboratory-raised Japanese monkeys (*Macaca fuscata*) perceive rotated photographs of monkeys: evidence of an inversion effect in perception. Primates 35:155–165

Valentine T (1988) Upside-down faces: a review of the effects of inversion upon face recognition. Br J Psychol 79:471–491

van Hoof JARAM (1967) The facial displays of the catarrhine monkeys and apes. In: Morris D (ed) Primate ethology. Weidenfeld & Nicholson, London, pp 7–68

Yarbus A (1967) Eye movements and vision. Plenum, New York

Young MP, Yamane S (1992) Sparse population coding of faces in inferior temporal cortex. Science 256:1327–1331

Zuber BL, Stark L, Cook G (1965) Micro-saccades and the velocity-amplitude relationship for saccadic eye movements. Science 150:1459–1460