Perception, 2013, volume 42, pages 1027 – 1042

doi:10.1068/p7552

Size-invariant facial expression categorization and associated gaze allocation within social interaction space

Kun Guo
School of Psychology, University of Lincoln, Lincoln LN6 7TS, UK; e-mail: kguo@lincoln.ac.uk
Received 19 June 2013, in revised form 20 November 2013

Abstract. As faces often appear under very different viewing conditions (eg brightness, viewing angle, or viewing distance), invariant facial information recognition is a key to our social interactions. Although we would clearly benefit from differentiating different facial expressions (eg angry vs happy) at a distance, there is surprisingly little research examining how expression categorization and associated gaze allocation are affected by viewing distance within the range of typical social space. In this study I systematically varied the size of faces displaying six basic facial expressions of emotion with varying intensities to mimic viewing distances ranging from arms length to 5 m, and employed a self-paced expression categorization task to measure participants' categorization performance and associated gaze patterns. Irrespective of the displayed expression and its intensity, the participants showed indistinguishable categorization accuracy and reaction time across the tested face sizes. Reducing face size decreased the number of fixations directed at the faces but increased individual fixation durations, and shifted gaze distribution from scanning all key internal facial features to fixating at mainly the central face region. The results suggest size-invariant facial expression categorization behaviour within social interaction distance which could be linked to a holistic gaze strategy for extracting expressive facial cues.

Keywords: facial expression, face size, fixation, facial region, categorization

1 Introduction

Human faces are probably the most important visual stimuli in our social environment. They provide a wealth of information about an individual's gender, age, familiarity, intention, and emotional state (Bruce and Young 1998). Given that the ability to recognize these visual cues and to respond accordingly plays a crucial role in our social interaction, it is perhaps not too surprising that we are highly skilled at face perception. One of the underlying mechanisms for this efficient cognitive process is the invariance of face representation in our visual system within given limits. For instance, we make little effort to identify familiar faces although they may appear under very different viewing conditions, such as viewing distance, viewing angle, viewing time, facial expression, and brightness (Bruce and Young 1998). Even for unfamiliar synthetic faces, face discrimination performance is size invariant, with faces differing in size up to fourfold (Lee et al 2006).

In addition to facial identity, facial expression of emotion is another important facial cue that we use to guide social judgment and behaviour. Several psychological studies have revealed invariance of facial expression judgment under different viewing conditions. For instance, some common facial expressions—such as happy, sad, fearful, angry, disgusted, contempt, and surprised—have comparable recognition accuracies between frontal and profile view (Kleck and Mendolia 1990; Matsumoto and Hwang 2011; see also Hess et al 2007). Varying resolution of face images has little impact on expression categorization performance. The recognition rate for happiness, fear, anger, or surprise remains consistent until the image is reduced to 10 × 15 pixels, in which almost no useful local facial information is left for visual analysis (Du and Martinez 2011).

However, despite the fact that we would clearly benefit from differentiating different facial expressions (eg angry vs happy) at a distance, there is surprisingly little research examining how expression categorization performance is affected by viewing distance or image size. An early study by Hager and Ekman (1979) has observed that, in general, viewers' expression recognition accuracy declined as the viewing distance increased, but some expressions could still be detected at a long distance (between 30–45 m). This was later confirmed by Smith and Schyns (2009) with a detailed computational and psychophysical study. By manipulating spatial frequency and image size (equivalent of viewing a typical face from 3.3 to 105 m away), they have shown that happy and surprised expressions use several low spatial frequency bands and can be recognized from a distance. Disgusted, fearful, angry, and sad expressions, on the other hand, are proximal expressions which are suited to close-range communication. These studies, however, examined only the impact of long-range viewing distances on affect recognition for faces with peak-intensity or high-intensity emotional expressions. In real-life scenarios our daily social interaction with others often happens at close range [< 12 ft (Hall 1966)], and we are more likely to use nonfacial cues (eg bodily cues, auditory inputs) to detect other people's emotions at a public distance (de Gelder 2009). Furthermore, we see less intense expressions more frequently than intense ones (Gao and Maurer 2009), and facial affects displayed at low intensity would significantly increase our difficulty to interpret subtle expressions (Gao and Maurer 2009, 2010). It is therefore essential to examine how typical social distances will affect our recognition of facial expression with varying intensities.

Additionally, at a close social distance, the faces falling in our visual field are large enough to elicit saccadic eye movements. A few behavioural and computational studies have suggested that basic facial expressions have minimal overlap in transmitted facial information and different facial features can provide diagnostic information in recognizing different expressions (eg the eyes and mouth transmit crucial cues for detecting angry and happy expressions, respectively) (Calvo and Nummenmaa 2008; Smith et al 2005). This hypothesis has been confirmed by recent eye-tracking studies. While performing a self-paced expression categorization task, participants tended to look at local facial regions that are most characteristic for each facial expression (Eisenbarth and Alpers 2011; Guo 2012; Jack et al 2009; Sullivan et al 2007). Compared with the same local region in different facial expressions, participants looked more often at the mouth region in happy faces, at the nose region in disgusted and sad faces, and at the eyes in angry, fearful, and surprised faces. However, participants rarely labelled an expression after fixating at only a single characteristic facial region. Instead, they analyzed facial information sampled from the diagnostic region (eg mouth in happy faces) in conjunction with those from other key internal facial features (eg eyes and nose regions in happy faces) when labelling (especially low-intensity or medium-intensity) facial expressions (Guo 2012). It has been proposed that a 'holistic' viewing strategy (scanning all key internal facial features to integrate all featural information into a single representation of the whole face) could be adopted to reliably categorize these common facial affects (Guo 2012). It is unclear, however, whether this holistic but also expression-sensitive gaze pattern would change with the viewing distance (or face image size) and expression intensity.

In this study a morphing technique was applied to varying intensities of six basic facial expressions of emotion (happy, sad, fearful, angry, disgusted, and surprised). Each expressive face image was then systematically varied in size to mimic different viewing distances. With a self-paced expression categorization task, this study aimed to examine to what extent participants' expression categorization performance and associated gaze patterns were affected by various image sizes (or social distances). As the average adult human face [~18 cm in height (Fang et al 2011)] subtends a visual angle of 13.4−22 deg at phases of close personal distances (~76−45 cm), 8.4−13.4 deg at far personal distances (~122−76 cm), and 4.8−8.4 deg at close social distances [~213−122 cm (Hall 1966)], this study tested four different face sizes which were roughly equivalent to viewing a typical face at 62 cm, 124 cm, 248 cm, and 496 cm, respectively. Also, 5 m was chosen as the longest viewing distance as this distance is already outside of typical social interaction space (Hall 1966), and the face presented at this distance should fall within fovea (~2 deg in size) and could be recognized with only 1 or 2 fixations. If facial expression shares similar size-invariant representation as facial identity within given limits, we would expect a consistent expression categorization performance across the tested face sizes, regardless of the displayed expression and its intensity. Furthermore, as size reduction would reduce the need to make frequent saccades and result in decreased discriminability due to the loss in fine detail in the scene (Loftus and Harley 2005), we would expect a shift of gaze allocation towards the central face region and prolonged fixation duration.

2 Materials and methods

To control potential gender, age, or culture difference in expression categorization performance and associated gaze pattern (Jack et al 2009; Vassallo et al 2009), only young female Caucasian participants were recruited. In total, twenty-six female undergraduate students, age ranging from 18 to 22 years old, volunteered to participate in this study. All participants had normal or corrected-to-normal visual acuity for each eye. The Ethical Committee in School of Psychology, University of Lincoln approved this study. Written informed consent was obtained from each participant, and all procedures complied with the British Psychological Society Code of Ethics and Conduct and with the World Medical Association Helsinki Declaration as revised in October 2008.

Digitized grey-scale face images in full frontal view were presented through a ViSaGe graphics system (Cambridge Research Systems, UK) and displayed on a noninterlaced gamma-corrected colour monitor (30 cd m^{-2} background luminance, 100 Hz frame rate, Mitsubishi Diamond Pro 2070SB) with the resolution of 1024 × 768 pixels. At a viewing distance of 57 cm, the monitor subtended a visual angle of 40 × 30 deg.

Western Caucasian face images, consisting of two female and two male models, were selected from the Karolinska Directed Emotional Faces CD ROM (Lundqvist et al 1998). Each of these models posed one neutral and six high-intensity facial expressions (happy, sad, fearful, angry, disgusted, and surprised). Although they may have real-world limitations, and categorization performance for some expressions could be subject to culture influence, these well-controlled face images were chosen for their comparability and universality in transmitting facial expression signals, at least for our observer group (Western Caucasian adults). The faces were processed in Adobe Photoshop to remove external facial features (eg hair) and to ensure a homogenous grey background, face size, and brightness. For each of the six expressions of each model, Morpheus Photo Morpher was used to create 3 levels of intensity in two physically equal steps (20%, 60%, and 100%) by morphing the emotional face with the neutral face. Each morphed face was then processed in Adobe Photoshop to create four presentation sizes: 430 × 568 pixels (16.5 × 21.8 deg, size 1), 215 × 284 pixels (8.3 × 10.9 deg, size $\frac{1}{4}$), 108 × 143 pixels (4.1 × 5.5 deg, size $\frac{1}{16}$), and 54 × 72 pixels (2.0 × 2.7 deg, size $\frac{1}{64}$). The largest and smallest face presentation sizes were roughly equivalent to viewing a face from 62 cm (arms length) and 500 cm away, respectively. As a result, 288 expressive face images were generated for the testing session (4 sizes × 6 expressions × 3 intensities × 4 models—see figure 1 for examples). These images were gamma corrected and displayed once in a random order during the testing. To mimic natural vision, the face images were presented in the peripheral visual field. To further control different perceptual sensitivities in face perception between left and right visual fields [eg left visual field advantage (Guo et al 2009)], the faces were presented in the left visual field.

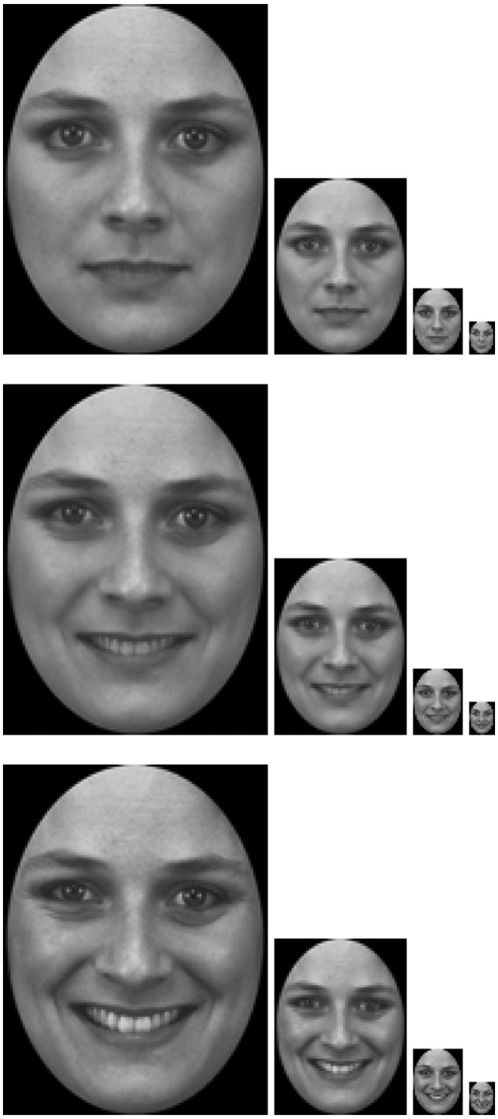

Figure 1. Examples of happy expression at varying intensities (from top to bottom: 20%, 60%, and 100%) and varying face sizes (from left to right: size 1, size $\frac{1}{4}$, size $\frac{1}{16}$, and size $\frac{1}{64}$).

All of our participants were aware of universal facial expressions. Before the testing, they were shown a PowerPoint presentation containing one male and one female model posing happiness, sadness, fear, anger, disgust, and surprise (sampled from Pictures of Facial Affect), and were asked to label each facial expression as carefully as possible without time constraint. All of them could recognize these facial expressions or agree with the classification proposed by Ekman and Friesen (1976).

A self-paced task was used to mimic natural viewing condition. During the self-paced experiments the participants sat in a chair with their head restrained by a chin-rest, and viewed the display binocularly. To calibrate eye movement signals, a small red fixation point (FP, 0.3 deg diameter, 15 cd m^{-2} luminance) was displayed randomly at one of 9 positions (3×3 matrix) across the monitor. The distance between adjacent FP positions was 10 deg. The participant was instructed to follow the FP and maintain fixation for 1 s. After the calibration procedure, the participant pressed the response box to initiate a trial. The trial was started with an FP displayed at the centre of the monitor. If the participant maintained fixation for 1 s, the FP disappeared and a face image was presented 10 deg left to the FP location (centre-to-centre distance). During the self-paced, free-viewing presentation, the participant was instructed to "categorize this facial expression as accurately and as quickly as possible", and to respond by pressing a button on the response box (for collecting reaction time data) with the dominant hand followed by a verbal report of the perceived facial expression (6-alternative forced choice: happiness, sadness, fear, anger, disgust, and surprise). No reinforcement was given during this procedure.

Horizontal and vertical eye positions from the self-reported dominant eye (determined through the Hole-in-Card test or the Dolman method if necessary) were measured using a Video Eyetracker Toolbox with 250 Hz sampling frequency and up to 0.25 deg accuracy (Cambridge Research Systems). The software developed in Matlab computed horizontal and vertical eye displacement signals as a function of time to determine eye velocity and position. Fixation locations were then extracted from the raw eye-tracking data using velocity (less than 0.2 deg eye displacement at a velocity of less than 20 deg s^{-1}) and duration (greater than 50 ms) criteria (Guo et al 2006).

While determining fixation allocation within key internal facial features (ie eyes, nose, and mouth), a consistent criterion was adopted to define boundaries between local facial features for different faces (for details see Guo et al 2010) to ensure equal size of individual internal feature across faces of different expressions and intensities from the same model. Specifically, the 'eye' region included the eyes, eyelids, and eyebrows; the 'nose' or 'mouth' region consisted of the main body of the nose (glabella, nasion, tip-defining points, alar-sidewall, and supra-alar crease) or mouth and immediate surrounding area (up to 0.5 deg). The division line between the mouth and nose regions was the midline between the upper lip and the bottom of the nose. Each fixation was then characterized by its location among feature regions and its time of onset relative to the start of the trial, and the number of fixations directed at each feature was normalized to the total number of fixations sampled in that trial.

3 Results

3.1 *Analysis of behavioural responses in expression categorization*

To examine to what extent image size would affect participants' overall task performance in categorizing facial expressions, two repeated-measures analyses of variance (ANOVAs) were conducted with image size as the independent variable, and percentage of correct expression identification and reaction time as the dependent variables. For each ANOVA, Greenhouse–Geisser corrections were applied where sphericity was violated. The analysis demonstrated that reducing face size had no impact on expression categorization accuracy ($F_{3,100} = 0.06$, $p = 0.98$, $\eta_p^2 = 0.02$; figure 2a) and reaction time ($F_{3,100} = 0.25$, $p = 0.86$, $\eta_p^2 = 0.007$; figure 2b).

Previous studies have demonstrated different perceptual sensitivities in recognizing facial expressions of different intensities, and in recognizing different facial expressions. Specifically, increasing expression intensity would improve categorization accuracy and shorten reaction time (eg Guo 2012), and people often have the most accurate and fastest identification performance for happiness, but are least accurate in recognizing fearful (or anxious) expressions (Guo 2012; Kirouac and Doré 1985; Palermo and Coltheart 2004; Rutishauser et al 2011). To examine whether image size would affect participants' behavioural responses in categorizing individual facial expressions with different intensities, 4 (image size) × 6 (expression type) × 3 (expression intensity) ANOVAs were conducted with categorization accuracy and reaction time as the dependent variables.

The analysis did not reveal a significant main effect of image size (accuracy: $F_{3,75} = 0.32$, $p = 0.82$, $\eta_p^2 = 0.001$; figure 2c; reaction time: $F_{2.22,55.38} = 1.90$, $p = 0.16$, $\eta_p^2 = 0.07$; figure 2d) or any significant interaction between image size and expression type or between image size and expression intensity (all ps > 0.21). It seems that, across the tested range of face

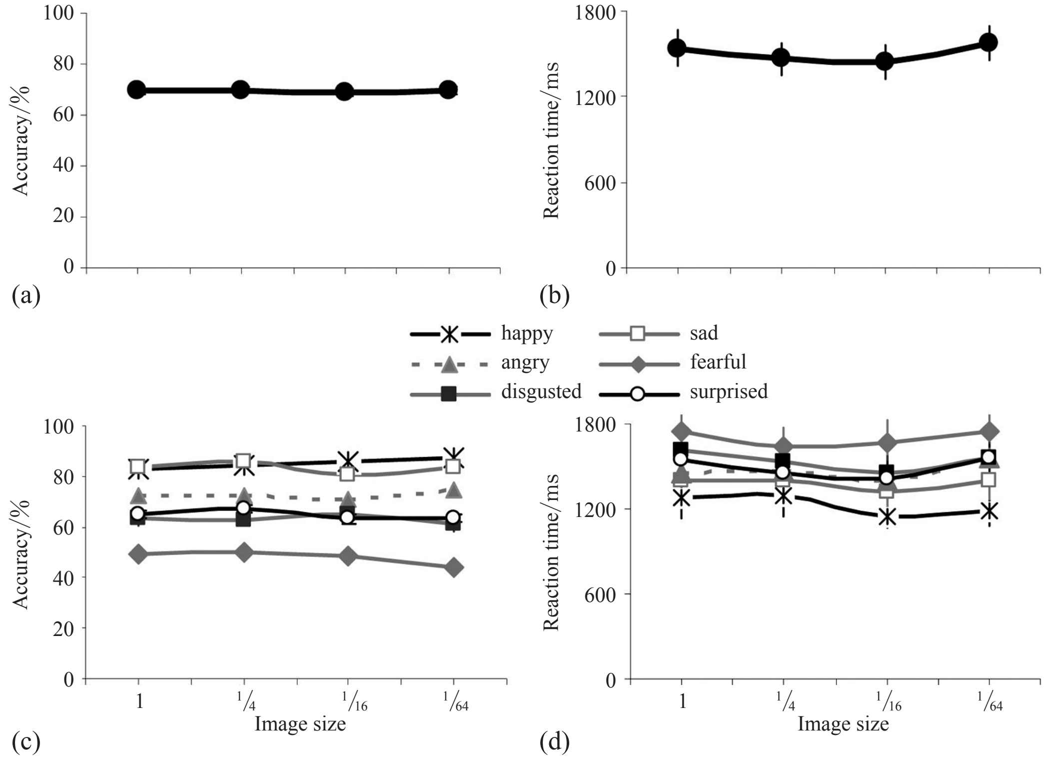

Figure 2. Mean accuracy (a) and (c) and reaction time (b) and (d) for expression categorization as a function of image size (viewing distance). Data presented in (a) and (b) are pooled across all the expressions. Different curves in (c) and (d) represent different facial expressions of emotion. Error bars represent SEM.

sizes, our participants showed indistinguishable categorization accuracy (see also tables 1 and 2) and reaction time for individual facial expressions presented with varying intensities.

However, in agreement with past research, both expression type and expression intensity had significant impact on categorization accuracy (expression type: $F_{3.4,85.26} = 60.29$, $p < 0.001$, $\eta_p^2 = 0.71$; expression intensity: $F_{1.24,30.99} = 50.54$, $p < 0.001$, $\eta_p^2 = 0.95$) and reaction time (expression type: $F_{3.5,87.45} = 15.14$, $p < 0.001$, $\eta_p^2 = 0.38$; expression intensity: $F_{1.09,27.28} = 23.76$, $p < 0.001$, $\eta_p^2 = 0.49$). A posteriori analysis (Bonferroni correction for multiple comparisons) further revealed that, among six tested expressions, participants showed the highest discrimination accuracy for happy and sad expressions, followed by angry expressions, and then by surprised and disgusted expressions. Fearful faces, on the other hand, induced the lowest discrimination accuracy (all ps < 0.01; figure 2c). In comparison with categorization accuracy, reaction time was less affected by individual facial expressions. Overall, happy and fearful faces tended to attract the fastest and the slowest reaction times, respectively (all ps < 0.05). There was little difference in reaction time while classifying sad, angry, disgusted, and surprised expressions (all ps > 0.05; figure 2d). Furthermore, increasing expression intensity from 20% to 100% led to monotonically increased expression categorization accuracy (tables 1 and 2) and decreased reaction time (all ps < 0.01).

Given relatively poor categorization accuracy for low-intensity expressive faces, full confusion matrices were computed to illustrate which expressions were mistaken for others and the impact of image size on expression categorization. For each displayed expression at a given size and intensity, we calculated the percentage of the trials in which the participant categorized the expression using each of the six expression labels. The averaged data across the participants were then analyzed with an ANOVA combined with a posteriori analysis.

As shown in tables 1 and 2, low-intensity happy, angry, surprised, disgusted, and fearful expressions were most likely to be mislabelled as sad expressions (Bonferroni correction for multiple comparisons, all *p*s < 0.01). For medium-intensity and high-intensity expressive faces, no systematic miscategorization bias was observed for happy, sad, angry, and surprised expressions (all *p*s > 0.05), but fear and disgust were often confused with surprised and angry expressions, respectively (all *p*s < 0.01). However, for an individual expression presented at a given intensity, image size had neither significant impact on categorization accuracy nor systematic impact on miscategorization bias (all *p*s > 0.05).

Table 1. Confusion matrices of happy, angry, and sad expression categorization (in bold): percentage of participants selecting the expression labels, averaged across the stimulus set and participants.

Displayed expression			Categorized expression (%)					
	size	intensity (%)	happy	sad	angry	fearful	disgusted	surprised
Happy	1	20	**45.19**	37.50	3.85	9.62	5.77	0
		60	**99.04**	0.96	0	0	0	0
		100	**99.04**	0.96	0	0	0	0
	1/4	20	**48.08**	42.31	4.81	2.88	0.96	0.96
		60	**97.12**	1.92	0.96	0	0	0
		100	**98.08**	1.92	0	0	0	0
	1/16	20	**51.92**	36.54	1.92	4.81	2.88	1.92
		60	**97.12**	1.92	0	0	0.96	0
		100	**98.08**	1.92	0	0	0	0
	1/64	20	**52.88**	36.54	2.88	2.88	2.88	1.92
		60	**98.08**	1.92	0	0	0	0
		100	**97.12**	0	0	0.96	1.92	2.88
Sad	1	20	2.88	**77.88**	16.35	0	2.88	0
		60	0	**82.69**	0.96	4.81	11.54	0
		100	0.96	**89.42**	0	2.88	6.73	0
	1/4	20	3.85	**79.81**	6.73	1.92	7.69	0
		60	0	**90.38**	0	2.88	5.77	0.96
		100	0	**87.50**	0	4.81	4.81	2.88
	1/16	20	7.69	**73.08**	7.69	5.77	4.81	0.96
		60	0.96	**85.58**	0	0.96	11.54	0.96
		100	1.92	**87.50**	0.96	1.92	5.77	1.92
	1/64	20	5.77	**78.85**	7.69	0	4.81	2.88
		60	1.92	**88.46**	0	3.85	5.77	0
		100	0	**88.46**	0	6.73	3.85	0.96
Angry	1	20	0.96	46.15	**31.73**	9.62	10.58	0.96
		60	0	9.62	**86.54**	0	3.85	0
		100	0	0.96	**93.27**	0	1.92	3.85
	1/4	20	6.73	43.27	**29.81**	3.85	16.35	0
		60	0.96	5.77	**87.50**	0	5.77	0
		100	0	0.96	**94.23**	0.96	3.85	0
	1/16	20	6.73	44.23	**30.77**	4.81	14.42	0
		60	0.96	5.77	**86.54**	0.96	5.77	0
		100	0	0	**96.15**	0	2.88	0.96
	1/64	20	9.62	42.31	**31.73**	4.81	11.54	0
		60	0	3.85	**89.42**	0.96	4.81	0.96
		100	0.96	0.96	**94.23**	0.96	2.88	0

Table 2. Confusion matrices of fearful, disgusted, and surprised expression categorization (in bold): percentage of participants selecting the expression labels, averaged across the stimulus set and participants.

Displayed expression			Categorized expression (%)					
	size	intensity (%)	happy	sad	angry	fearful	disgusted	surprised
Fearful	1	20	5.77	75.96	4.81	**8.65**	4.81	0
		60	0	2.88	4.81	**65.38**	1.92	25.00
		100	0.96	1.92	3.85	**69.23**	7.69	16.35
	1/4	20	9.62	73.08	2.88	**9.62**	4.81	0
		60	1.92	1.92	0.96	**65.38**	4.81	25.00
		100	1.92	0	0.96	**74.04**	6.73	16.35
	1/16	20	15.38	59.62	5.77	**10.58**	6.73	1.92
		60	3.85	0	0.96	**61.54**	4.81	28.85
		100	2.88	0	1.92	**74.04**	4.81	16.35
	1/64	20	17.31	57.69	10.58	**7.69**	4.81	1.92
		60	3.85	2.88	2.88	**59.62**	10.58	20.19
		100	0	1.92	0.96	**70.19**	7.69	19.23
Disgusted	1	20	4.81	37.50	31.73	1.92	**24.04**	0
		60	0.96	1.92	14.42	2.88	**79.81**	0
		100	0	3.85	4.81	1.92	**89.42**	0
	1/4	20	7.69	29.81	30.77	1.92	**29.81**	0
		60	0	7.69	16.35	1.92	**72.12**	1.92
		100	0.96	7.69	4.81	0.96	**85.58**	0
	1/16	20	8.65	39.42	22.12	1.92	**27.88**	0
		60	0	6.73	17.31	0.96	**74.04**	0.96
		100	0.96	1.92	6.73	1.92	**88.46**	0
	1/64	20	4.81	33.65	34.62	1.92	**25.00**	0
		60	0.96	0.96	23.08	0.96	**73.08**	0.96
		100	0.96	2.88	3.85	3.85	**87.50**	0.96
Surprised	1	20	5.77	62.50	9.62	9.62	4.81	**7.69**
		60	1.92	0.96	0	4.81	1.92	**90.38**
		100	0	0.96	0.96	0.96	0	**97.12**
	1/4	20	16.35	57.69	4.81	4.81	7.69	**8.65**
		60	0	2.88	0	3.85	0	**93.27**
		100	0	0	0.96	1.92	0.96	**96.15**
	1/16	20	14.42	58.65	10.58	3.85	4.81	**7.69**
		60	0	0.96	0.96	5.77	0.96	**91.35**
		100	0	0	0.96	3.85	0.96	**94.23**
	1/64	20	16.35	54.81	10.58	6.73	4.81	**6.73**
		60	1.92	0	0	8.65	0	**89.42**
		100	0.96	0	0	2.88	0	**96.15**

3.2 *Analysis of gaze patterns in expression categorization*

To examine to what extent image size would affect participants' overall gaze behaviour in categorizing expression across all faces, two one-way ANOVAs were conducted with image size as the independent variable, and averaged number of fixations and fixation duration directed at each face as the dependent variables. The analysis revealed that reducing face size gradually decreased the number of fixations needed to classify facial expressions

($F_{1.7,42.81} = 48.5$, $p < 0.001$, $\eta_p^2 = 0.66$; figure 3a) but increased mean fixation durations ($F_{1.17,29.18} = 31.62$, $p < 0.001$, $\eta_p^2 = 0.56$; figure 3b). Also in agreement with past research (Guo 2012), different facial expressions tended to attract a different number of fixations ($F_{5,125} = 12.37$, $p < 0.001$, $\eta_p^2 = 0.33$; figure 3c). Participants directed the least and the most number of fixations to viewing happy and fearful faces, respectively (all ps < 0.05). They directed an indistinguishable number of fixations in classifying sad, angry, disgusted, and surprised expressions (all ps > 0.05).

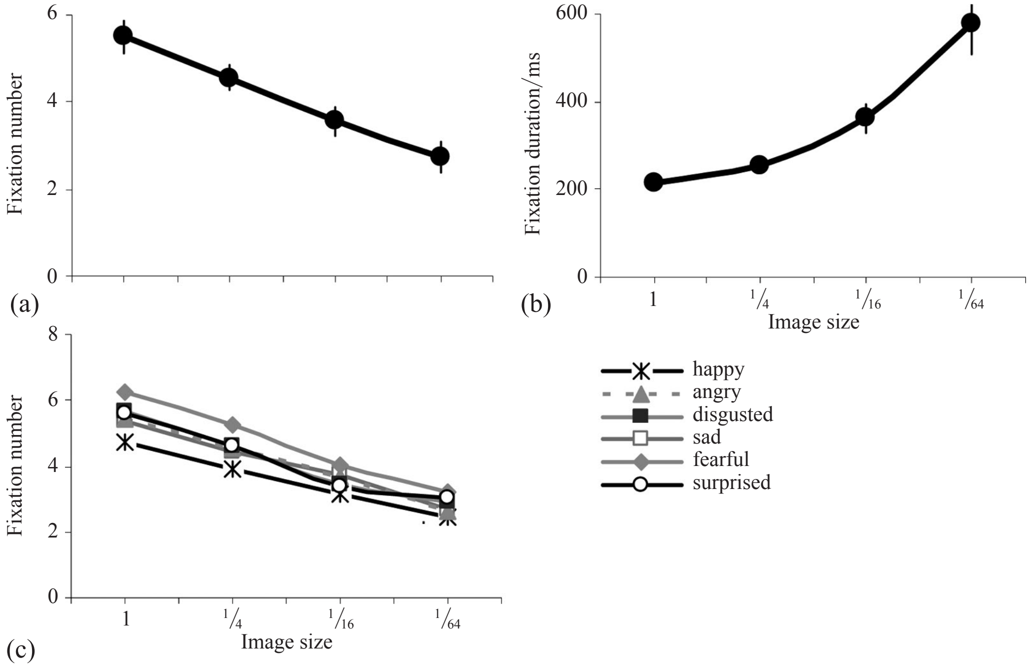

Figure 3. Number of fixations (a) and (c) and fixation duration (b) directed at the expressive face as a function of image size (viewing distance). Data presented in (a) and (b) are pooled across all the expressions. Different curves in (c) represent different facial expressions of emotion. Error bars represent SEM.

I then examined how image size would affect fixation distribution in viewing expressive faces. Early studies have demonstrated that, during the task of expression categorization, the vast majority of fixations were allocated at key internal facial features, such as eyes, nose and mouth (Eisenbarth and Alpers 2011; Guo 2012; Jack et al 2009; Sullivan et al 2007). However, as the size of these local facial features in the size-reduced face images (ie sizes 1/16 and 1/64) was smaller than the fovea region, participants did not have to gaze directly at these features to extract local facial information. Hence it was difficult to compare the proportion of fixations allocated at the same facial feature across different face sizes. I therefore calculated the proportion of fixations allocated within 1 deg radius around the face centre (~midpoint of nose body) for each face size. A one-way ANOVA showed that, with decreasing face size, participants directed increasingly higher proportion of fixations towards the central face area ($F_{2.31,57.63} = 111.3$, $p < 0.001$, $\eta_p^2 = 0.83$; figure 4a), indicating a stronger central bias in processing of smaller face images.

As the eyes, nose, and mouth regions in size-1 and size-¼ faces were large enough to attract direct fixations, a detailed comparison of fixation allocation at individual facial features was made between these two face sizes (figure 4b). A 2 (face size) × 3 (face region) ANOVA with normalized proportion of fixations directed at each facial region as the dependent variables showed a significant main effect of face size ($F_{1,25} = 17.76$, $p < 0.001$, $\eta_p^2 = 0.42$) and face

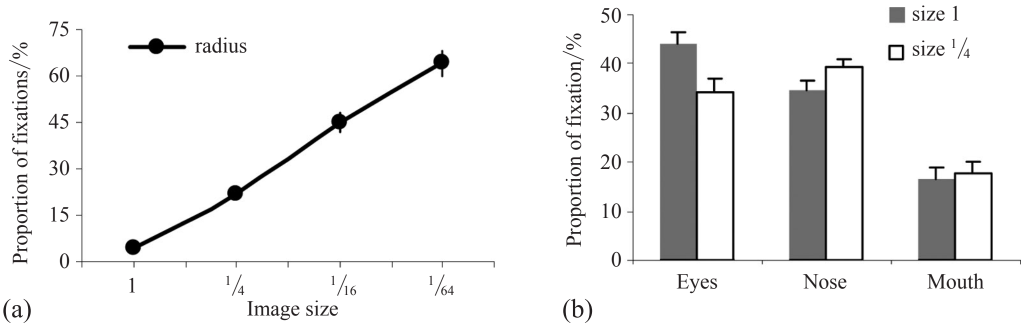

Figure 4. Normalized proportion of fixations allocated within 1 deg radius around the face centre (a) and at the eyes, nose, and mouth regions (b) during the task of categorizing facial expressions with varying face sizes (viewing distances). Error bars represent SEM.

region ($F_{2,50} = 18.32, p < 0.001, \eta_p^2 = 0.42$), and significant interaction between face size and face region ($F_{2,50} = 24.92$, $p < 0.001$, $\eta_p^2 = 0.5$). Specifically, among internal facial features in the larger faces (size 1), the eyes tended to attract the highest proportion of fixations, followed by the nose and then the mouth region (all ps < 0.01). However, among features in the smaller faces (size $^1\!/_4$), the eyes and nose attracted a similar proportion of fixations ($p = 0.26$), followed by the mouth region (all ps < 0.01). When comparing proportion of fixations directed at the same feature in faces of different sizes, participants directed more fixations at the eyes in size-1 faces ($p < 0.001$), more fixations at the nose region in size-$^1\!/_4$ faces ($p = 0.001$), but the same number of fixations at the mouth region ($p = 0.29$). It seems that, when the viewed faces were getting smaller, participants tended to reduce the amount of fixations directed at the eyes and redirected some of them towards the nose region (the central area of the face).

As the basic facial expressions have minimal overlap in transmitted facial information and different facial features can provide diagnostic information in recognizing different expressions (Smith et al 2005), people normally look more often at local facial regions that are most characteristic for each facial expression, such as the eyes in fearful faces and the mouth in happy faces (Eisenbarth and Alpers 2011; Guo 2012). To examine to what extent this expression-specific gaze allocation is affected by face size (size 1 vs size $^1\!/_4$), a 2 (face size) × 3 (face region) × 6 (expression type) ANOVA was conducted with a normalized proportion of fixations directed at each facial region as the dependent variables (figure 5). Expression intensity was not included as an additional independent variable, as a previous study has demonstrated that the proportional distribution of fixations at local features was unchanged for individual facial expression of varying intensities (Guo 2012). The analysis showed significant interaction between face region and expression type ($F_{10,250} = 10.77$, $p < 0.001$, $\eta_p^2 = 0.30$) and between face size and face region ($F_{2,50} = 11.96, p < 0.001, \eta_p^2 = 0.32$), but nonsignificant interaction between face size and expression type ($F_{5,125} = 1.08, p = 0.38$, $\eta_p^2 = 0.04$) or between face size, face region, and expression type ($F_{10,250} = 1.04$, $p = 0.41$, $\eta_p^2 = 0.04$).

The a posteriori comparison further revealed that, when categorizing size-1 faces, the eyes in happy faces were the least viewed, followed by the eyes in disgusted faces; the eyes in sad, angry, fearful, or surprised faces, on the other hand, were the most frequently viewed facial features (all ps < 0.05). For the nose region, the participants directed more fixations at the nose in disgusted, happy, and sad faces than that in angry, fearful, and surprised faces (all ps < 0.05). As for the mouth region, the mouth in happy and sad faces attracted the greatest and the least proportion of fixations, respectively (all ps < 0.01); the mouth in the faces of other expressions drew a similar number of fixations (all ps > 0.05).

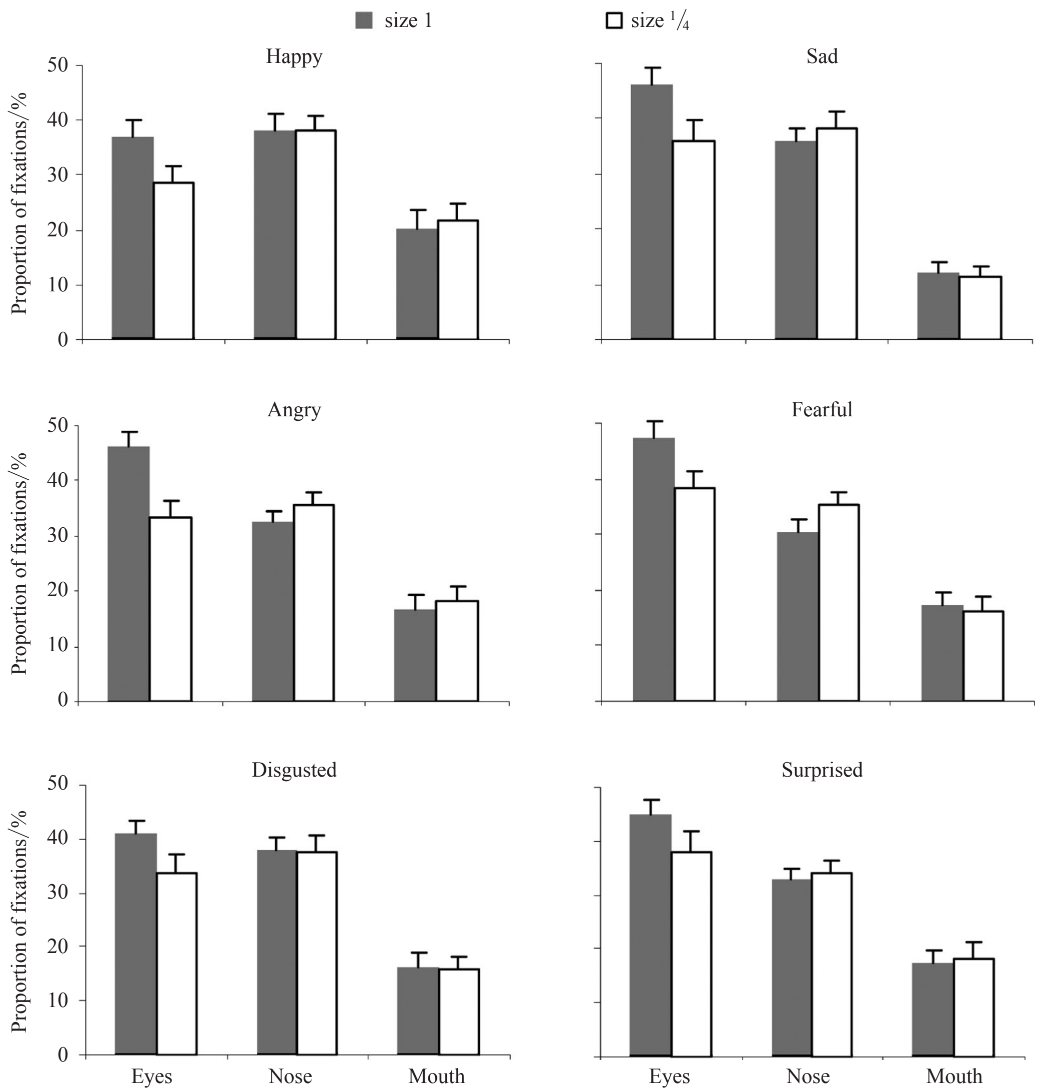

Figure 5. Normalized proportion of fixations directed at the eyes, nose, and mouth regions when categorizing different facial expressions with varying face size (viewing distances). For each expression, data sampled from different intensities were collapsed together. Error bars represent SEM.

When categorizing the smaller size-$\frac{1}{4}$ faces, the eyes in the happy faces were still the least viewed, but the eyes in surprised and fearful faces tended to attract slightly more fixations than the eyes in angry and disgusted faces (all *p*s < 0.05). There was no difference in the number of fixations directed at the nose region in different facial expressions (all *p*s > 0.05). For the mouth region, the mouth in happy and sad faces still attracted the greatest and the least proportion of fixations, respectively (all *p*s < 0.01), and the mouths in the faces of other expressions drew a similar number of fixations (all *p*s > 0.05). Although there were some variances, this expression-specific gaze allocation was largely preserved in size-1 and size-$\frac{1}{4}$ faces. As far as individual facial features were large enough to attract direct fixation, participants tended to look more often at the mouth region in happy faces than in other expressive faces, and at the eyes in fearful and surprised faces than in other facial expressions.

4 Discussion

Invariant facial expression recognition is a key to effective social interactions. Previous studies have demonstrated viewing-angle-invariant facial expression categorization, such as comparable recognition accuracies between frontal and profile view for common expressions such as happiness, sadness, fear, anger, disgust, and surprise (Kleck and Mendolia 1990; Matsumoto and Hwang 2011; see also Hess et al 2007). This study further showed that our visual system is capable of recognizing facial expressions fairly invariant of face size or viewing distance within given limits. In a self-paced facial expression categorization task participants showed indistinguishable categorization accuracy and reaction time when the presented face size was varied to mimic viewing distance in typical social interactions (ranging from arms length to 5 m; figures 2a and 2b). Furthermore, this viewing-distance-invariant categorization was likely to be independent of expression category and displayed intensity. Although our categorization performance is expression-dependent (eg people often have the most accurate and fastest identification performance for happiness, but are least accurate in recognizing fearful expressions) and increases with the increasing expression intensity (Gao and Maurer 2010; Guo 2012), the categorization performance to the same expression displayed with the same intensity was not varied across the tested face sizes (figures 2c and 2d; tables 1 and 2). It should be noted that the high recognition accuracy for sad expressions (as shown in figure 2c) could be partly due to categorization bias of labelling low-intensity ambiguous expressive faces as sadness (tables 1 and 2). Future studies could address to what extent this categorization bias affects facial expression recognition.

Using bandpass filters to manipulate spatial frequency bands of expressive faces, early research has found that mid-peak and high-peak spatial frequencies are needed to discriminate happiness, sadness, and fear from neutral faces, suggesting that these expressions can be recognized only within relative proximity (Goren and Wilson 2006). By reducing face size from 2 deg to 0.07 deg and presenting part of a bandpass-filtered face through 'bubbles' protocol, Smith and Schyns (2009) lately revealed a gradient of decreasing expression recognition sensitivity to the increasing viewing distance in which only happy and surprised expressions are suited for longer distance recognition. In this study I used relatively large faces (ranging from 2 deg to 16.5 deg) and found consistent recognition performance for individual facial expressions, even when the facial expression was displayed at very low (20%) intensity. This is reasonable as the six tested common expressions with varying intensities are routinely present during our social interactions with others, and often occur within a distance of a few metres, ensuring that fine facial details are available to the perceiver.

By presenting part of an intensified expressive face in isolation (ie through a masking or 'bubbles' protocol), past studies have observed that participants could solely rely on different facial parts to recognize basic facial expressions (Calvo and Nummenmaa 2008; Smith et al 2005). For instance, the lower half of the face is more informative for labeling happy expressions, whereas the upper half is better for detecting fear and surprise, suggesting that different facial features can transmit diagnostic information in recognizing different expressions. Recent eye-tracking studies further examined how individual facial expressions affected gaze allocation at the key internal facial features, and found that people tend to look more often at local features that are most characteristic for each facial expression, such as eyes in sad faces and mouths in happy faces (Eisenbarth and Alpers 2011). However, probably because the expressive cue (especially low-intensity expressive cues) from a single facial feature is often ambiguous and unreliable for accurate expression categorization (Jack et al 2009; Kohler et al 2004), participants rarely categorize an expression (even at peak intensity) after fixating at only a single characteristic facial region. Instead, they often analyze facial information sampled from the diagnostic region (eg mouths in happy faces) in conjunction with that from other key internal facial features (eg eyes and nose in happy faces) before labeling the expression (Guo 2012).

In the present study such 'holistic' but also expression-specific gaze patterns seemed to be size invariant as far as individual facial features in the viewed faces were large enough to attract direct gaze. When the faces were presented in size-1 or size-¼ faces (equivalent to viewing a face at arms length or 1.2 m away), irrespective of the displayed expression and its intensity, our participants often scanned all key facial features (ie eyes, nose, and mouth) and gazed relatively more at the most characteristic local feature for each expression, such as the mouth in the happy face (in comparison with the mouth in other facial expressions) and the eyes in the fearful and surprised faces (in comparison with the eyes in other facial expressions; figure 5). Although the nose and surrounding region could transmit informative cues to detect disgusted and sad expressions of varying intensities (Guo 2012; Smith et al 2005), it might be surprising that the nose attracted a considerable number of fixations in face size 1 (less than the eyes but more than the mouth) and size ¼ (similar as the eyes but more than the mouth). This could be partly due to the current experimental setup which was designed to mimic a natural viewing condition. As the face was presented at 10 deg away from the initial central fixation point, the first two saccades were more likely to land at the face centre (Hsiao and Cottrell 2008), which in turn increased the probability of the nose being fixated.

Small faces, on the other hand, promoted central fixation bias, even when individual facial features could still attract direct gaze (such as in size $^1/_4$). That is, participants showed an increasing tendency to gaze at the central face region (nose body region) with the decreasing face size, regardless of the displayed expression and its intensity (figure 4). As the centre of the face is an optimal viewing position for perceiving and integrating expressive cues from surrounding local facial features in all directions, the stronger central fixation bias in smaller faces could also be interpreted as a holistic gaze behaviour for processing facial expressions. This is further supported by the prolonged fixation duration when viewing smaller faces (figure 3b), as integrating expressive cues from nearby multiple local facial features would need extra processing time. It is worth pointing out that the degree of this central bias in smaller faces (size $^1/_{16}$ and $^1/_{64}$) could be underestimated in this study. As the face was presented at 10 deg away from the initial central fixation point, the first saccade to the face centre could be overshot or undershot (eg landing at the edge of the face). This also explains the reason participants sometimes made 2 or 3 fixations when viewing the smallest size-$^1/_{64}$ faces.

Taken together, although the gaze pattern in recognizing facial expressions would change according to the viewed face size, the underlying gaze strategy might remain as the holistic processing. Specifically, in order to integrate all the featural information into an individual representation of an expressive face as a whole, people tend to scan all key internal facial features in large faces or to fixate at the central face region in smaller faces. However, the central bias in fixation distribution (especially at an early viewing stage) has been documented in a wide range of scene-viewing studies regardless of task demand, presentation format, natural scene category, and image quality (eg Judd et al 2011; Tatler 2007; Tseng et al 2009), suggesting it could be an inherent characteristic of human oculomotor behaviour. As the scene centre may be an optimal location for early scene perception, the central fixation bias could be more evident for smaller scenes, including nonface images. Hence it remains unclear to what extent the proposed holistic face processing strategy could account for the enhanced central fixation bias in the smaller faces. Future studies could systematically compare gaze behaviour in viewing of face and nonface images to examine how the cognitive process underlying central fixation bias is affected by different scene categories.

Our findings of gaze allocation in viewing expressive faces of varying sizes could further advance the understanding of cognitive mechanism underlying invariant facial information recognition. Computational studies have proposed two basic approaches for achieving size invariance: normalization and extracting invariant features (Biederman 1987; Wiskott 2006). The normalization approach suggests that invariant recognition is based on an internal transformation to normalize retinal image of an object to a standard size. The invariant features approach, on the other hand, suggests that invariant recognition is based on some object features which are invariant to the size of an object in the visual field. As for facial expression recognition, considering the substantial variability in local facial regions across different individuals, different situations, and different intensities to express the same facial expression [eg angriness could be associated with frowning, the outer brow raised, visible teeth, the lower lip depressed, lips tightly closed (Kohler et al 2004)], it is not surprising that the gaze behaviour of our participants was changed systematically according to the face size to extract and process expressive facial cues. Therefore, normalization seems to be a dominant approach to achieve size-invariant representation of facial expressions of emotion.

It remains to be seen to what extent the current findings can be generalized to different contexts, such as different face presentation formats and task demands (eg brief fovea presentation with fixed duration). Furthermore, to control for potential gender, culture, or race bias in social cognition, this study recruited only Western Caucasian females and tested their expression categorization performance towards Western Caucasian faces. To minimize the potential interaction between identity and expression processing and to keep the testing time at a reasonable length (~1 h per testing session), this study used photographs of expressive faces of only 4 models, which inevitably had limited variability within each expression category. Although categorization of six basic facial expressions is not heavily influenced by these variables (Bruce and Young 1998), it would be interesting to examine whether similar size-invariant expression recognition and associated gaze behaviour sill exists when viewing a large variety of realistic faces with differences in identity, race, and age.

In conclusion, this study aimed to examine how face size (simulating viewing distance within typical social interaction range) would affect facial expression categorization performance and associated gaze allocation. Regardless of the presented facial expressions and their intensities, the participants showed indistinguishable categorization accuracy and reaction time across the tested face sizes, suggesting a size-invariant categorization process for facial expressions of emotion within given limits. This size-invariant expression recognition might be linked with a holistic and expression-specific gaze strategy. Smaller faces would attract a stronger central fixation bias to efficiently gather facial cues from surrounding features. Furthermore, if individual facial features were large enough to attract direction gaze, people would scan all key internal facial features but fixate more often at the local feature that is most characteristic for individual facial expression.

Acknowledgments. I would like to thank Kelly-Anne Dudgeon, Carrie Tooth, and Rachel Denman for helping with data collection.

References

Biederman I, 1987 "Recognition-by-components: a theory of human image understanding" *Psychological Review* **94** 115–147

Bruce V, Young A, 1998 *In the Eye of the Beholder* (New York: Oxford University Press)

Calvo M G, Nummenmaa L, 2008 "Detection of emotional faces: salient physical features guide effective visual search" *Journal of Experimental Psychology: General* **137** 471–494; doi:10.1037/a0012771

de Gelder B, 2009 "Why bodies? Twelve reasons for including bodily expressions in affective neuroscience" *Philosophical Transactions of the Royal Society, Series B* **364** 3475–3484; doi:10.1098/rstb.2009.0190

Du S, Martinez A M, 2011 "The resolution of facial expressions of emotion" *Journal of Vision* **11**(13):24, 1–13; doi:10.1167/11.13.24
Ekman P, Friesen W, 1976 *Pictures of Facial Affect* (Palo Alto, CA: Consulting Psychologists Press)
Eisenbarth H, Alpers G W, 2011 "Happy mouth and sad eyes: scanning emotional facial expressions" *Emotion* **11** 860–865; doi:10.1037/a0022758
Fang F, Clapham P J, Chung K C, 2011 "A systematic review of interethnic variability in facial dimensions" *Plastic and Reconstructive Surgery* **127** 874–881; doi:10.1097/PRS.0b013e318200afdb
Gao X, Maurer D, 2009 "Influence of intensity on children's sensitivity to happy, sad, and fearful facial expressions" *Journal of Experimental Child Psychology* **102** 503–521; doi:10.1016/j.jecp.2008.11.002
Gao X, Maurer D, 2010 "A happy story: Developmental changes in children's sensitivity to facial expressions of varying intensities" *Journal of Experimental Child Psychology* **107** 67–86; doi:10.1016/j.jecp.2010.05.003
Goren D, Wilson H R, 2006 "Quantifying facial expression recognition across viewing conditions" *Vision Research* **46** 1253–1262; doi:10.1016/j.visres.2005.10.028
Guo K, 2012 "Holistic gaze strategy to categorize facial expression of varying intensities" *PLoS ONE* **7**(8), e42585; doi:10.1371/journal.pone.0042585
Guo K, Mahmoodi S, Robertson R G, Young M P, 2006 "Longer fixation duration while viewing face images" *Experimental Brain Research* **171** 91–98; doi:10.1007/s00221-005-0248-y
Guo K, Meints K, Hall C, Hall S, Mills D, 2009 "Left gaze bias in humans, rhesus monkeys and domestic dogs" *Animal Cognition* **12** 409–418; doi:10.1007/s10071-008-0199-3
Guo K, Tunnicliffe D, Roebuck H, 2010 "Human spontaneous gaze patterns in viewing of faces of different species" *Perception* **39** 533–542; doi:10.1068/p6517
Hager J C, Ekman P, 1979 "Long-distance of transmission of facial affect signals" *Ethology and Sociobiology* **1** 77–82; doi:10.1016/0162-3095(79)90007-4
Hall E T, 1966 *The Hidden Dimension* (Garden City, NY: Doubleday)
Hess U, Adams R B, Kleck R E, 2007 "Looking at you or looking elsewhere: The influence of head orientation on the signal value of emotional facial expressions" *Motivation and Emotion* **31** 137–144; doi:10.1007/s11031-007-9057-x
Hsiao J H, Cottrell G W, 2008 "Two fixations suffice in face recognition" *Psychological Science* **9** 998–1006
Jack R E, Blais C, Scheepers C, Schyns P G, Caldara R, 2009 "Cultural confusions show that facial expressions are not universal" *Current Biology* **19** 1543–1548; doi:10.1016/j.cub.2009.07.051
Judd T, Durand F, Torralba A, 2011 "Fixations on low-resolution images" *Journal of Vision* **11**(4):14, 1–20; doi:10.1167/11.4.14
Kirouac G, Doré F Y, 1985 "Accuracy of the judgement of facial expression of emotions as a function of sex and level of education" *Journal of Nonverbal Behavior* **91** 3–7; doi:10.1007/BF00987555
Kleck R E, Mendolia M, 1990 "Decoding of profile versus full-face expressions of affect" *Journal of Nonverbal Behavior* **14** 35–49; doi:10.1007/BF01006578
Kohler C G, Turner T, Stolar N M, Bilker W B, Brensinger C M, Gur R E, Gur R C, 2004 "Differences in facial expressions of four universal emotions" *Psychiatry Research* **128** 235–244; doi:10.1016/j.psychres.2004.07.003
Lee Y, Matsumiya K, Wilson H R, 2006 "Size-invariant but viewpoint-dependent representation of faces" *Vision Research* **46** 1901–1910; doi:10.1016/j.visres.2005.12.008
Loftus G R, Harley E H, 2005 "Why is it easier to identify someone close than far away?" *Psychonomic Bulletin & Review* **12** 43–65; doi:10.3758/BF03196348
Lundqvist D, Flykt A, Öhman A, 1998 "The Karolinska Directed Emotional Faces—KDEF", CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9
Matsumoto D, Hwang H S, 2011 "Judgements of facial expressions of emotion in profile" *Emotion* **11** 1223–1229; doi:10.1037/a0024356
Palermo R, Coltheart M, 2004 "Photographs of facial expression: Accuracy, response times, and ratings of intensity" *Behavior Research Methods, Instruments, & Computers* **36** 634–638; doi:10.3758/BF03206544

Rutishauser U, Tudusciuc O, Neumann D, Mamelak A N, Heller A C, Ross I B, Philpott L, Sutherling W W, Adolphs R, 2011 "Single-unit responses selective for whole faces in the human amygdale" *Current Biology* **21** 1654–1660; doi:10.1016/j.cub.2011.08.035

Smith F W, Schyns P G, 2009 "Smile through your fear and sadness: Transmitting and identifying facial expression signals over a range of viewing distance" *Psychological Science* **20** 1202–1208; doi:10.1111/j.1467-9280.2009.02427.x

Smith M L, Cottrell G W, Gosselin F, Schyns P G, 2005 "Transmitting and decoding facial expressions" *Psychological Science* **16** 184–189; doi:10.1111/j.0956-7976.2005.00801.x

Sullivan S, Ruffman T, Hutton S B, 2007 "Age differences in emotion recognition skills and the visual scanning of emotion faces" *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences* **62** 53–60; doi:10.1093/geronb/62.1.P53

Tatler B W, 2007 "The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions" *Journal of Vision* **7**(14):4, 1–17; doi:10.1167/7.14.4

Tseng P, Carmi R, Cameron I G M, Munoz D P, Itti L, 2009 "Quantifying center bias of observers in free viewing of dynamic natural scenes" *Journal of Vision* **9**(7):4, 1–16; doi:10.1167/9.7.4

Vassallo S, Cooper S L, Douglas J M, 2009 "Visual scanning in the recognition of facial affect: Is there an observer sex difference?" *Journal of Vision* **9**(3):11, 1–10; doi:10.1167/9.3.11

Wiskott L, 2006 "How does our visual system achieve shift and size invariance?", in *23 Problems in Systems Neuroscience* Eds J L van Hemmen, T J Sejnowski (New York: Oxford University Press) pp 322–340