Scan patterns during the processing of facial expression versus identity: An exploration of task-driven and stimulus-driven effects

George L. Malcolm  
Departments of Neurology, Ophthalmology and Visual Sciences, University of British Columbia, Canada

Linda J. Lanyon  
Departments of Neurology, Ophthalmology and Visual Sciences, University of British Columbia, Canada

Andrew J. B. Fugard  
Departments of Neurology, Ophthalmology and Visual Sciences, University of British Columbia, Canada

Jason J. S. Barton  
Department of Psychology, University of British Columbia, Canada

Perceptual studies suggest that processing facial identity emphasizes upper-face information, whereas processing expressions of anger or happiness emphasizes the lower-face. The two goals of the present study were to determine (a) if the distributions of eye fixations reflect these upper/lower-face biases, and (b) whether this bias is task- or stimulus-driven. We presented a target face followed by a probe pair of morphed faces, neither of which was identical to the target. Subjects judged which of the pair was more similar to the target face while eye movements were recorded. In Experiment 1 the probe pair always differed from each other in both identity and expression on each trial. In one block subjects judged which probe face was more similar to the target face in identity, and in a second block subjects judged which probe face was more similar to the target face in expression. In Experiment 2 the two probe faces differed in either expression or identity, but not both. Subjects were not informed which dimension differed, but simply asked to judge which probe face was more similar to the target face. We found that subjects scanned the upper-face more than the lower-face during the identity task but the lower-face more than the upper-face during the expression task in Experiment 1 (task-driven effects), with significantly less variation in bias in Experiment 2 (stimulus-driven effects). We conclude that fixations correlate with regional variations of diagnostic information in different processing tasks, but that these reflect top-down task-driven guidance of information acquisition more than stimulus-driven effects.

Keywords: face expression, face identity, scanpath, fixation


Introduction

When subjects inspect a visual scene, they often do so with not just a single fixation, but a series of fixations distributed among different regions of the scene. These shifts in fixation serve to re-direct both the fovea, the area of retina with the highest resolving capacity for spatial detail, and the locus of attention, which is often (but not always) correlated with the locus of fixation. Thus, shifts of fixations bring new perceptual data to the observer, while at the same time the perceptual data processed by the observer determines where subsequent fixations should be directed. This renders “seeing” an active process, with interplay between vision and eye movements (Henderson, 2003).

Although early studies of scanning behavior showed significant inter-subject variability in the way different observers scan the same stimulus (Noton & Stark, 1971), fixations are not distributed randomly. Rather, the interplay between eye movements and perception that lies behind the distribution of fixations likely reflects a process of acquiring information for a perceptual judgment (Deco & Schürmann, 2000; Rybak, Gusakova, Golovan,
Podlachikova, & Shevtsova, 1998). As depicted in some models (Itti & Koch, 2000; Rybak et al., 1998), this information acquisition can be guided by image properties, so that regions with high luminance or color contrast, borders, or motion are more likely to attract fixations (Mannan, Ruddock, & Wooding, 1996; Parkhurst, Law, & Niebur, 2002). However, it is not always the case that the locations of critical information for a perceptual decision correspond to the regions with the most prominent contrasts in low-level properties. It is also highly probable that the specific task or goal of the observer guides information acquisition (Chen & Zelinsky, 2006; Land & Hayhoe, 2001; Neider & Zelinsky, 2006; Yarbus, 1967). For a real-world example, if the task were to find one’s keys in a room an observer would search a completely separate region (e.g. a nearby desk) than if the task were to search for the exit of that same room (e.g. the far wall). The task or goal can influence the region of the image analyzed, independent of low-level visual saliency. Thus, models of scanning fixation patterns need to incorporate both bottom-up guidance by stimulus properties and top-down guidance by task and scene knowledge (Chen & Zelinsky, 2006; Henderson, 2003; Lanyon & Denham, 2005; Navalpakkam & Itti, 2005).

How might studies on stimulus-driven and task-driven fixations in scenes inform us about fixations in faces? The eye movements made while subjects look at faces have been studied by many groups (Althoff & Cohen, 1999; Barton, Radcliffe, Cherkasova, Edelman, & Intriligator, 2006; Groner, Walder, & Groner, 1984; Luria & Strauss, 1978; Rizzo, Hurtig, & Damasio, 1987; Stacey, Walker, & Underwood, 2005; Walker-Smith, Gale, & Findlay, 1977). Analyzing which facial regions subjects scan and why those regions are fixated is complicated by the high variability of scanning patterns between subjects and across different faces (Walker-Smith et al., 1977), the complex and variable structure of faces, and the fact that multiple types of information are conveyed by faces, such as identity, expression, age, gender and direction of gaze. On the other hand, the fact that faces do convey multiple types of information can be exploited to determine how fixation patterns are driven by changing either the tasks or the diagnostic information available in a face image. There is substantial data regarding the location of information useful for perceptual decisions within faces (Shepherd, Davies, & Ellis, 1981), and recent work shows that this diagnostic information is located in different facial regions for different face-processing tasks (Gosselin & Schyns, 2001; Schyns, Bonnar, & Gosselin, 2002; Smith, Cottrell, Gosselin, & Schyns, 2005). Despite the known inter-subject variability in face scanning, could face perception fixation patterns be correlated with locations of diagnostic information relevant to specific forms of processing? If so, scanning fixations might be studied to learn more about the location of critical information in faces and how they are integrated into a perceptual decision.

In this report, we present the results of two experiments designed, first, to test the hypothesis that the distribution of scanning fixations while subjects view faces reflects the location of diagnostic information relevant to two specific discriminations. Second, the experiments examined whether the discriminations that resulted in separate regions of fixations were task- or stimulus-driven. In one discrimination, between different identities, prior data suggests that the most useful information is in the upper-face (Fisher & Cox, 1975; Langdell, 1978; Schyns et al., 2002), while in the other discrimination, between expressions of happiness and disgust, the data suggest that the lower-face contains the most critical data (Smith et al., 2005). Both discriminations were tested in two experiments.

In Experiment 1, we asked subjects to discriminate between two highly similar images, each created by morphing between two faces of both different identities and different expressions. Thus, in every trial of Experiment 1, there was both discriminable identity and expression information in the faces. Experiment 1 contained two blocks, one of which had subjects discriminate on the basis of identity, while in the other block subjects discriminated on the basis of expression. Because the same images were used in both blocks, any difference in fixation patterns between blocks would be task-driven. In Experiment 2, we again presented two highly similar images, but this time faces were created by either morphing between different identities with the same expression, or between different expressions in the same face. Subjects did not know whether the images on a given trial differed in expression or identity, but were asked merely to discriminate between the two faces. As there was no specific task to guide subjects, this experiment reflects mainly stimulus-driven processes.

If scanning fixations reflect diagnostic regions-of-interest, then in the identity discrimination trials there should be more fixations in the upper-face while in the expression discrimination trials there should be more fixations in the lower-face. If this distribution of fixations results from task demands, we should see a greater regional distribution of fixations in the task-driven experiment (Experiment 1) than in the stimulus-driven experiment (Experiment 2). If, however, scanning distribution in faces is guided by differences in the face stimuli, then the bias in fixation distribution between the upper-face and lower-face (for identity and expression discriminations, respectively) should be greater in the stimulus-driven experiment than in the task-driven experiment.

Methods

Participants

Eight subjects, 3 female, of mean age 31 years (range 25 to 41 years) participated in both experiments, all with normal or corrected to normal vision. All subjects gave informed consent in accordance with the principles of the Declaration of Helsinki and the institutional review boards.
of the University of British Columbia and Vancouver General Hospital approved the protocol.

Stimuli

We obtained from the Karolinska Database of Emotional Faces (Lundqvist & Litton, 1998) pairs of photos from eight different females, one for each showing happiness and one showing disgust. The eight females were paired to create four morph matrices. Each matrix was created using Fantamorph 3.0 (Abrosoft, www.fantamorph.com). When morphing between pairs of faces, corresponding reference points were selected on both faces. These points work as anchors between the selected face images, keeping relative properties of the chosen point in a particular location, or moving certain properties across two locations if the corresponding reference point is in a new location, while the remainder of the image morphs from one face into the other in a preset number of steps. We always morphed in 1% steps and used as many reference points as needed to remove any indication of manipulations from the resulting morphed faces (Figure 1).\(^1\)

The first morphs used different face-identity images with the same facial expression. This created two morph series of a 100 faces for each pair: one morph series with a happy expression, and one series with a disgusted expression. We then morphed across expression, keeping identity constant, in 1% steps from the happy identity morph to its corresponding disgusted identity morph (e.g. morphing from the 45:55 mix of identity 1 and identity 2 with the happy expression to the 45:55 mix of identity 1 and identity 2 with the disgusted expression). This created a 100 × 100 matrix of faces that differed in identity and/or expression.

Each trial presented a single target face followed by two probe faces side-by-side. Subjects responded as to which of the two probe faces was most similar to the recently seen but no longer visible target face. This design was chosen to minimize the likelihood of subjects using strategies such as matching of low-level image properties, rather than engaging face-processing mechanisms to reach a perceptual decision. The target face was always one of the original, unmorphed sixteen faces chosen from the Karolinska database. The probe faces were taken from the morph matrix generated from the identity pair that had been used to create the matrix (Figure 2). (That is, if disgusted identity 1 was the target face, then the probe would be taken from the morph matrix generated from the mixtures of identity 1 and identity 2, and not from the matrices created from identity 3 and identity 4, identity 5 and identity 6, or identity 7 and identity 8.) For both experiments, our aim was to find probe pairs that would create a difficult discrimination task: if the task was too easy because the probes were too dissimilar, a single fixation may have sufficed to perform the discrimination. Having 100 × 100 matrices allowed us to carefully select faces from different regions of each matrix that, over several psychophysical pilot tests, yielded the desired accuracy rate of between 70 and 85%.\(^2\) Analysis of these pilot data showed no significant difference in accuracy between judgments of identity and those of expression in the probes used, in either Experiment 1 or Experiment 2.

Using Adobe Photoshop CS 8.0 (www.adobe.com), we placed all faces inside an oval aperture with black surround, eliminating hair cues that might have been used for identity tasks. Target faces were shown in color while probe faces were shown in gray-scale. Target faces spanned 25° in height and 23° in width, while the probe faces spanned 18° in height and 16° in width each (or 18° in height and 34° in width when placed side-by-side). These variations in hue and size between target and probes were meant to minimize reliance on low-level image cues.

Apparatus

We recorded eye movements with an Eyelink 1000 eyetracker sampling at 1000 samples/second with a mean spatial accuracy of 0.25°–0.5° (SR Research, www.sr-research.com). The eye tracker recorded the position and duration of eye movements during scanning and the input of subjects’ responses. Stimuli were shown on a 20” monitor placed 57 cm in front of the subject, in a room with standard dim illumination. Subjects placed their heads in a chin rest for the duration of each experiment to maintain viewing distance.

Procedure

Every experimental block began with the standard Eyelink calibrating procedure, which recorded eye position as subjects fixated a series of 9 dots arranged in a rectangular grid extending to 25° in height and 36° in width. Every trial began with a drift-correction, with the experimenter pressing the spacebar when they were fixating a central dot on the screen.

Half of the subjects began with Experiment 1 and half with Experiment 2. Within Experiment 1, half of the subjects began with the block testing discrimination of
identity and half began with the block testing discrimination of expression. There were, thus, four possible combinations in the testing order.

Each trial began with one of the sixteen (unmorphed) target faces, which remained visible for 3000 ms. Once the target face disappeared a 2° × 2° black square appeared in the center of the screen, which subjects were told to watch. The box remained on the screen until fixated, after which the box disappeared and two probe faces appeared side-by-side, separated by 2°. Probe faces remained visible until the subject made a perceptual decision and pressed a key indicating whether the left face or the right face was most similar to the target face.

Experiment 1

In this experiment, the probe faces were diagonally related in terms of the morph matrix (see Figure 2): that is, the pair differed from each other along both the identity and expression axes of the matrix. For example, one of the two probe faces might have a 45:55 mix of identity 1 and identity 2 with a 45:55 mix of happy and disgust, while the other probe face would have a 55:45 mix of identity 1 and identity 2 with a 55:45 mix of happy and disgust. Hence the first probe image would have slightly more identity 2 and a slightly more disgusted expression, while the second probe image would have slightly more identity 1 and a slightly happier expression. Similarly, morphs along the other diagonal could be used to create a probe pair in which one probe image would be slightly more identity 1 and slightly more disgusted, while the second probe image would be slightly more identity 2 and slightly happier. Thus for each of the four morph matrices, pairings could come from either diagonal (Figure 3). For each probe pairing, two probe image pairs were created: one with the first image on the left and the second on the right, and one with the positions reversed. Each probe image pair could be compared to either of two target faces at the end of the diagonal (e.g. happy identity 1, disgusted identity 1, happy identity 2, disgusted identity 2), creating 8 different trials for each morph matrix. Each trial was repeated twice, and with four different morph matrices this created 64 trials.

Subjects saw all these 64 trials twice, in both of the two different blocks. In one block, they were asked to indicate which of the probe faces was most similar in identity to
In this experiment, the probe faces were orthogonally related in the matrix: that is, they differed slightly along either the identity or the expression axes of the matrix, but not both. Therefore, unlike Experiment 1, here the face-pair stimuli provided subjects information on where to fixate, rather than the task, since they were asked to indicate merely which face was most similar in expression to the target face.

When creating the stimuli, a 45:55 mix, for example, of identity 1 and identity 2 with a 45:55 mix of happy and disgust could be paired with a 55:45 mix of identity 1 and identity 2 with the same 45:55 mix of happy and disgust. This would be a probe pair used in an identity trial, in which the first probe image is slightly more identity 1 and the second probe image slightly more identity 2, but expression is the same in both faces (Figure 3). Alternatively, a 45:55 mix of identity 1 and identity 2 with a 45:55 mix of happy and disgust could be paired with the same 45:55 mix of identity 1 and identity 2 but now with a 55:45 mix of happy and disgust. This would be a probe pair used in an expression trial, in which the first probe image is slightly more disgusted and the second probe image slightly more happy, but identity is the same. Unlike Experiment 1, which had two diagonal pairings per matrix, there are four possible orthogonal pairings in each morph matrix in Experiment 2: two along the horizontal (identity varying) rows and two along the vertical (expression varying) columns. This created 128 trials, 64 varying along the identity axis and 64 varying along the expression axis.

All 128 trials were randomly mixed in one block, with a rest period at the halfway point. Subjects were not instructed on any trial whether the probe faces varied on expression or identity. They were asked simply to indicate which of the two probes most resembled the target face.

### Analysis

Probe faces were divided along a horizontal boundary touching the lower eye crease: fixations above this line were classified as upper face, while those below the line were classified as lower face. Fixations were excluded if they fell outside the external outline of the face. (For examples of a trial and the eye movements made by one subject, see Figure 4).

We modeled the data using multilevel logistic regression (Agresti, 2002), which has precedence in eye-tracking experiment analysis (see Barr, in press), using the lme4 package (Bates, 2007) in R (www.r-project.org). Such models provide an efficient means to model item level responses clustered within subjects, eliminating the need to average across trials and thus provide more statistical power. We assessed the effect of dimension (identity versus expression) for both Experiment 1 (task-driven) and Experiment 2 (stimulus-driven) together in one model predicting fixation location (upper- or lower-face). The relationship between the variables in the final model we present is as follows:

\[
\Pr(y_{\text{simfe}} = 1) = \logit^{-1}(b_0 + b_1 \cdot \text{dimension}_{\text{simfe}} + b_2 \cdot \text{experiment}_{\text{simfe}} + b_3 \cdot \text{experiment}_{\text{simfe}} \cdot \text{dimension}_{\text{simfe}} + \text{subject}_s + \text{matrix}_m + \text{face}_f + \text{expression}_e + \epsilon_{\text{simfe}})
\]
where \( \Pr(ysimfe = 1) \) is read as “the probability that subject \( s \), for fixation \( i \), when given matrix \( m \), face \( f \), and expression \( e \), looks at the upper part of the face.” The variables \( b_0 \), \( b_1 \), \( b_2 \), and \( b_3 \) are the fixed effects, i.e. the group average effects, induced by the experimental manipulations (dimension\( simfe \) and experiment\( simfe \), which were given a treatment coding). The random effect term \( subject \) represents each subject’s deviation from the group average preference for upper versus lower face fixation. The other random effects terms (matrix\( m \), face\( f \), and expression\( e \)) represent subject-invariant, between-item variation in the proportion of upper versus lower face fixations. Finally, \( \varepsilon_{simfe} \) is the residual term, representing remaining unexplained variance.

We assessed whether the models were a significant fit to the data by using log-likelihood ratio tests, calculated as \(-2(l_1 - l_0)\), where \( l_0 \) and \( l_1 \) denote the maximized log-likelihood of two models to be compared, one of which (\( l_1 \)) comes from the model which has the predictors of interest; the other (\( l_0 \)) comes from a model which differs only by not having these variables. This statistic has a null distribution approximating that of \( \chi^2 \), with degrees of freedom obtained from the difference in the number of parameters, so a \( \chi^2 \)-test is used to assess whether a predictor contributes significantly to the model’s fit. We also use the Wald statistic, calculated from the slope estimate and its standard error, to test whether the slope differs significantly from zero (see Agresti, 2002).

**Results**

We began our exploration of fixation count with a model containing only a fixed effect of the intercept and all the random effects. The random effect with the largest variance was that for subjects (\( SD = 1.60 \)). From the items, the face provided most subject-invariant variance (\( SD = 0.11 \)), followed by matrix and expression (both \( SD = 0.05 \)).

There were main effects of both experiment [\( \chi^2(1) = 21.8, p < 0.001 \)] and dimension [\( \chi^2(1) = 23.2, p < 0.001 \)]. We tested the effect of the originally hypothesized interaction between experiment and dimension, namely that there would be more fixations in the upper-face for the identity condition than for the expression condition for the task-driven experiment, and that the difference in upper-versus-lower fixations between dimensions would be less for the stimulus-driven experiment. The interaction

---

**Figure 4.** Examples of scanning fixations during processing of probe pairs of faces in single trials. The top image was during a task-driven identity task, the middle image was during a task-driven expression task, and the bottom image was during a stimulus-driven task. The red lines connect fixations in the order that they occurred, while the blue rings correspond to fixations made, with larger rings indicating longer durations.

**Figure 5.** The thick red lines show group average effects, computed from the fixed effects. The thin black lines show estimates for each individual, computed from the fixed effects and the subject random intercept. (a) and (b) show estimated probabilities of fixating in the upper-half of the face for the two experiments.
was significant \[ \chi^2(1) = 42.5, p < 0.001 \]. The slope for the interaction term was significantly different to zero (slope = 0.56, 95% CI = 0.39–0.73, Wald \[ z = 6.5, p < 0.001 \]). Figures 5a and 5b present the mean fixation probabilities computed from the model.

We also modeled the data in the two experiments separately. For the stimulus-driven experiment, there was only a trend to an effect of dimension \( \chi^2(1) = 2.91, p = 0.09 \). The effect was much larger for the task-driven experiment \( \chi^2(1) = 62.8, p < 0.001 \). For the task-driven experiment, the probability of a fixation on the upper-half of the face was significantly higher for identity processing than for expression processing (slope = 0.48, 95% CI = 0.36–0.59, Wald \[ z = 7.9, p < 0.001 \]), whereas there was a slight opposite trend for the stimulus-driven experiment (slope = −0.11, 95% CI = −0.23 to 0.02, Wald \[ z = −1.7, p = 0.09 \]).

**Discussion**

Previous research has shown that the diagnostic information relevant to processing identity is concentrated more in the upper face, while that relevant to processing emotional expressions, in particular happiness and disgust, is concentrated more in the lower face (Gosselin & Schyns, 2001; Schyns et al., 2002). Consistent with the assertion that the distribution of scanning fixations reflects a drive to gather information directed towards a perceptual decision, our data showed a significant shift from more scanning of the upper- than lower-face, to more scanning of the lower- than upper-face, when observers switched from the task of processing identity to the task of processing expression, despite the fact that the stimulus sets of both tasks were identical (containing simultaneous changes along both identity and expression axes). However, when the task was simply to detect any difference, with one set of stimuli differing along the identity axis alone and the other differing along the expression axis alone, there was no change in the mean distribution of scanning between the upper- and lower-face. These results suggest that a scan pattern does change in a manner that reflects the distribution of diagnostic information within a face, but that these patterns are driven more by the task that is driving information acquisition than by stimulus differences.

The importance of task in determining fixation distribution has been noted in a number of recent studies, both of natural scene perception (Land & Hayhoe, 2001; Tatler, Baddeley, & Vincent, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006) and the processing of natural object arrays (Chen & Zelinsky, 2006). Our results extend the conclusions about the critical role of task in scanning behavior from scene processing to the recognition of objects, in our case faces.

Direct contrasts between top-down task relevancy effects and bottom-up stimulus saliency effects have also been previously made. When subjects search for a target in an array of objects, their eye movements can be distracted by an object given high saliency by being colored red; however, when they are cued to the target object, so that search is guided by prior knowledge, this effect is minimized (Chen & Zelinsky, 2006). When subjects are engaged in a task like making a sandwich or a cup of tea, their gaze is directed more to task-relevant objects and “influenced very little by the ‘intrinsic salience’ of objects” (Land & Hayhoe, 2001). When subjects shown natural scenes are prompted to extract specific information (e.g. count the pedestrians in a street scene), task and scene context are more accurate predictors of fixations than a model based on low-level visual saliency alone (Torralba et al., 2006).

Our face study differed from these scene studies in two important ways. First, our experiments did not pit task-driven effects against stimulus-driven effects in the same trials: rather, we dissociated them so that in one experiment task information varied while the stimuli did not, while in the other experiment stimulus properties varied while task information did not. This allowed us to ask, given the different distributions in diagnostic information for expression and identity in our design, whether scanning patterns are driven by the task, by the stimulus, or by both.

Second, the stimulus variations in our faces likely did not reflect large shifts in saliency from the upper- to the lower-half of the faces, but a shift in the locus of diagnostic information. As such, this may represent not so much a salience effect but a type of knowledge-driven effect that others have called “scene-schema knowledge”, a “generic semantic and spatial knowledge about a particular type of scene” (or object in our case) that can guide visual search and scanning (Henderson, 2003). If scene-schema knowledge had been effective in guiding face scanning, then we speculate that early perceptual data from the initial segments of scanning in Experiment 2 should have led to internal hypotheses regarding whether the faces differed in expression or identity. These early hypotheses should then have led to increased scanning of the relevant face half of these subtly differing morphs to arrive at a decision regarding which morph was more similar to the target. Our results show, however, that this type of stimulus-based knowledge is not as effective as “task-related knowledge” in generating shifts in oculomotor search.

Many different approaches have been employed to study the distribution of diagnostic information in faces, including tests of recognition of portions of faces, tests of recognition while selective components are distorted, multidimensional scaling studies (for review, see Shepherd et al., 1981), and, more recently, the ‘Bubbles’ technique, in which recognition is measured for frequency-filtered faces viewed through randomly positioned
apertures (Gosselin & Schyns, 2001; Schyns et al., 2002; Smith et al., 2005). While most of the previous studies did not examine the influence of task, the ‘Bubbles’ studies advanced the field by showing how diagnostic information changed as individuals switched between judgments of identity, expression, and gender. Our data indicate that fixations also show an effect of task, and that this effect is consistent with the shift in diagnostic information from the upper- to the lower-half as the task changes from processing identity to expression. The presence of this shift is predicted by concepts of the role of fixations in acquiring information for perceptual decisions (Deco & Schurmann, 2000), and affirms that eye movements can serve as a useful probe for the spatial distribution of diagnostic information in perceptual tasks.

Acknowledgments

JB was supported by a Canada Research Chair and a Senior Scholar Award from the Michael Smith Foundation for Health Research. This work was supported by NIMH 1R01 MH069898, CIHR MOP-77615, and CIHR MOP-85004. This work was originally presented at the meeting of the Vision Sciences Society, Sarasota, May 2007.

Commercial relationships: none.
Corresponding author: George L. Malcolm.
Email: g.l.malcolm@sms.ed.ac.uk.
Address: Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh, Scotland EH8 9JZ.

Footnotes

1 All subjects in the pilot and experiment confirmed after participation that the morphed faces appeared normal, without any visible sign of manipulation.
2 The final set of pilot data was obtained from six subjects, four of whom also performed the eye movement experiments.

References


