

How Do Monkeys Look at Faces?

Frederick K. D. Nahm and Amelie Perret

The Salk Institute for Biological Studies

David G. Amaral

University of California, Davis

Thomas D. Albright

The Salk Institute for Biological Studies

Abstract

■ Facial displays are an important form of social communication in nonhuman primates. Clues to the information conveyed by faces are the temporal and spatial characteristics of ocular viewing patterns to facial images. The present study compares viewing patterns of four rhesus monkeys (*Macaca mulatta*) to a set of 1- and 3-sec video segments of conspecific facial displays, which included open-mouth threat, lip-smack, yawn, fear-grimace, and neutral profile. Both static and dynamic video images were used. Static human faces displaying open-mouth threat, smile, and neutral gestures were also presented. Eye position was recorded with a surgically implanted eye-coil. The relative perceptual salience of the eyes, the midface, and the mouth across different expressive gestures was determined by analyzing the number of eye movements associated with each feature during static and dynamic presentations. The results

indicate that motion does not significantly affect the viewing patterns to expressive facial displays, and when given a choice, monkeys spend a relatively large amount of time inspecting the face, especially the eyes, as opposed to areas surrounding the face. The expressive nature of the facial display also affected viewing patterns in that threatening and fear-related displays evoked a pattern of viewing that differed from that recorded during the presentation of submissive-related facial displays. From these results we conclude that (1) the most important determinant of the visual inspection patterns of faces is the constellation of physiognomic features and their configuration, but not facial motion, (2) the eyes are generally the most salient facial feature, and (3) the agonistic or affiliative dimension of an expressive facial display can be delineated on the basis of viewing patterns. ■

INTRODUCTION

The ability to discern the meaning or importance of various sensory stimuli is an essential component in the life of sentient organisms such as primates. This is especially true in the area of social communication. Both human and nonhuman primates inhabit an environment in which the desires or intentions of other members as well as their own must constantly be communicated and perceived. In nonhuman primates, especially semiterrestrial old-world macaques who are heavily reliant upon visual communication (Marler, 1965; Altman, 1967), one of the primary means of social communication is through facial displays.

The comparative study of facial expressions began in 1872 with the publication of Charles Darwin's *The Expression of the Emotions in Man and Animals*. Many expressive facial gestures displayed by monkeys are characterized by specific movements and configurations of

the facial musculature and skin (van Hooff, 1962; Hinde & Rowell, 1962) and occur in stereotypical social encounters that can be used to determine both the emotional state and the behavioral intent of the gesturing monkey (Chevalier-Skolnikoff, 1973; Redican, 1975). For example, the open-mouth stare is frequently seen during an agonistic encounter, the fear-grimace when a subordinate is threatened by a more dominant animal, and the lip-smack prior to grooming or approaching. Facial displays also occur at different frequencies and intensities depending upon such factors as social status, age, and gender (Chevalier-Skolnikoff, 1973). Infant monkeys seldom display an open-mouth stare but are often seen lip-smacking or grimacing. This has been interpreted as indicative of the need at this early age to establish affiliative bonds. Although there are several hypotheses as to what determines the appearance of a particular facial expression (i.e., motivational states, protective responses, etc.), it is unequivocal that social communication in

higher nonhuman primates depends heavily upon vision and that facial expressions are important visual cues. Despite numerous studies of the morphology of facial expressions and the situations in which they occur, there is a paucity of quantitative data concerning the precise features of facial expressions to which percipients attend or how viewing patterns to different expressive facial displays might be related to the recognition of emotional or behavioral states. A more detailed analysis of the perceptual salience of expressive facial cues can, however, be obtained from laboratory studies using a variety of experimental techniques.

The first laboratory studies using facial expressions as visual stimuli employed static representations such as photographs or slides. These studies revealed that static images of monkey faces were effective in eliciting species-typical behavioral responses (Sackett, 1965; Overman & Doty, 1982). In an effort to determine which features are attended by monkeys viewing faces, Keating and Keating (1982) studied the visual scan patterns of monkeys viewing black and white photographic slides. The eye region was found to be a particularly strong attractor of eye fixation. Interestingly, the presentation of a small number of gesturing faces with prominent mouth gestures only marginally affected the percipient's pattern of fixations. More recently, Keating and Keating (1993) showed the recognition of familiarity of human facial images to be dependent upon the eye region, including the eyes proper and the brow. Wilson and Goldman-Rakic (1994) have also reported the eyes to be the preferential target of monkeys viewing conspecific facial images, in addition to other facial features and contours such as the ears, hairline, nose, mouth, and jawline. Although the eyes are undeniably important expressive cues, these data may underestimate the perceptual salience of other facial features since these previous studies were limited to a small number of static facial displays and, more importantly, under natural viewing conditions, facial gestures are inherently dynamic. In consideration of Altman's (1967) dictum that social communication is "a structure not of objects but of events," dynamic footage of facial expressions may be a more appropriate means of eliciting naturalistic viewing patterns to expression-dependent facial features.

The use of motion pictures of gesturing monkeys has in the past revealed that, given a choice, monkeys will choose to look at dynamic images over static ones (Butler, 1961). Video loops of "fearsome" pictures are known to be particularly effective in determining viewing preferences (Humphrey, 1972) and can be successfully used to communicate affect in the laboratory setting (Miller, 1967). Evidence from human studies has also revealed that the use of dynamic facial expressions can facilitate the accurate judgment of affect (see Ekman, Friesen, & Ellsworth, 1982). These studies suggest that dynamic images can be an important tool in the laboratory study of nonhuman primate visual systems and the recognition

of facial affect. Whether static or dynamic, however, it is clear that from an ethological perspective, faces play a central role in mediating primate social interactions.

In addition to providing both quantitative and qualitative evidence as to how monkeys visually scan facial displays, the study of viewing patterns to expressive facial displays has important implications for other related studies investigating the neurobiological mechanisms underlying face perception. Since Gross, Rocha-Miranda, and Bender (1972) first reported the existence of face-responsive cells in inferior temporal cortex of nonhuman primates, other neurophysiological studies have further documented face-responsive cells in the inferotemporal cortex of the macaque that are sensitive to particular facial features, orientation, relative configuration, and familiarity (Perrett, Rolls, & Caan, 1982; Desimone, Albright, Gross, & Bruce, 1984; Perrett, Mistlin, & Chitty, 1987; Yamane, Kaji, & Kawano, 1988). A small number of cells in the superior temporal sulcus have also been shown to respond differentially to faces exhibiting threatening and nonthreatening expressions (Hasslemo, Rolls, & Baylis, 1989). Interestingly, the inferotemporal cortex projects to the amygdaloid complex (Herzog & Van Hoesen, 1976; Turner, Mishkin, & Knapp, 1980; Iwai & Yukie, 1987), which also contains neurons sensitive to complex visual stimuli with behavioral significance. The most effective exemplars of this class of behaviorally significant visual stimuli in neurophysiological studies of the amygdala are facial displays (Nakamura, Mikami, & Kubota, 1992). Experimental lesions of the amygdala produce aberrations in social behavior such as failure to maintain social group contacts, absence of conspecific communication gestures, abolition of mother-infant bond, and change in social status (Rosvold, Mirsky, & Pribram, 1954; Dicks, Meyers, & Kling, 1969; Steklis & Kling, 1985; Zola-Morgan, Squire, Alvarez-Royo, & Clower, 1991). These data provide compelling evidence, as previously suggested (Brothers, 1990), that the mechanism by which the amygdala mediates social communication is via its role in the detection of emotional or behavioral salient visual cues associated with different facial expressions. The recent study by Adolphs, Tranel, Damasio, and Damasio (1994), showing impaired recognition of facial affect in a human subject with bilateral amygdala damage, strongly supports this view. The mechanisms by which damage to the amygdala effects such changes is not known, for despite the compelling nature of these behavioral and neurobiological findings, we have only a rudimentary understanding of the phenomenological properties of facial cues that are used to distinguish the emotional or intentional state(s) of a gesturing animal. Documenting the phenomenological salience of the face and its featural properties through the study of viewing patterns can provide us with important insights into these issues.

In the present series of experiments we set out to compare the viewing patterns of monkeys to both static

and dynamic conspecific and human facial displays. By presenting a variety of facial expressions differing in affective and behavioral significance, we have been able to investigate the relationship between facial expression and scanning patterns.

RESULTS

Differences between Dynamic and Static Facial Displays

ANOVA comparisons between eye-movement data collected during the presentation of a dynamic image versus 1- and 3-sec static representations of that dynamic image were performed using the image type (dynamic, 1-sec static, and 3-sec static) as the nominal variable. There were three levels of the dependent variable: eyes, midface, and mouth. No significant differences were seen between the average viewing time (normalized for size) for the three major facial features (eyes, midface, and mouth) across the three image types (dynamic, 1-sec static, and 3-sec static).

Although neither the presence of motion nor the length of presentation consistently affected the viewing pattern to expressive facial displays, a lesser fraction of time was spent viewing the face during the 3-sec static presentations. ANOVA using image type as the nominal variable revealed a significant effect ($F(2, 213) = 27.274$, $p < 0.001$) for the total number of recorded eye positions off the viewing monitor. Using both Fishers PLSD and Tukey-Kramer post-hoc tests, the amount of time viewing off-screen positions during the 3-sec stimulus was found to be significantly greater than during the dynamic and 1-sec images at a probability level of 0.01. The number of eye positions recorded off the screen during presentation of the dynamic and 1-sec static images were not significantly different.

Scanning Patterns to Dynamic Facial Gestures

Figure 2A illustrates viewing time (msec) for the three major facial features (eyes, midface, and mouth) in each of the 11 dynamic images from Figure 1. Data are averaged across subjects ($n = 4$) and normalized for feature size. An 11×4 repeated measures ANOVA for each image (between-factor) and each feature (within-factor) revealed a significant feature effect ($F(2, 66) = 35.161$, $p < 0.001$). No significant image or image-by-feature effect was evident. The inset graph in Figure 2A shows the same data collapsed across all eleven dynamic images. Linear contrast analyses revealed that viewing time was significantly greater for the eyes as compared to midface ($F(1, 66) = 48.73$, $p < 0.001$) and mouth ($F(1, 66) = 56.46$, $p < 0.001$). There was no significant difference between the viewing time for the midface and that for the mouth ($F(1, 66) = 0.284$, $p = 0.596$).

Repeated measures ANOVA revealed a significant fea-

ture effect ($F(2, 66) = 120.44$, $p < 0.001$) when comparing the amount of time spent viewing facial features (eyes, midface, and mouth) versus other parts of the head (on-head) and other regions (on-screen). Subjects spent the majority of time scanning the face as opposed to other regions of the stimuli (Figure 3). Linear contrast analysis revealed these differences to be highly significant (facial features versus on-head: $F(1, 66) = 170.85$, $p < 0.001$; facial features versus on-screen: $F(1, 66) = 189.96$, $p < 0.001$). There was no significant difference between the viewing time for on-head and that for on-screen.

Scanning Patterns for 1- and 3-Sec Static Facial Expressions

Repeated measures ANOVA revealed a significant image ($F(9, 124) = 6.787$, $p < 0.001$), feature ($F(2, 124) = 52.047$, $p < 0.001$), and image-by-feature ($F(18, 124) = 3.952$, $p < 0.001$) effect for the 1-sec static images and a significant image ($F(9, 124) = 5.313$, $p < 0.001$), feature ($F(2, 124) = 28.552$, $p < 0.001$), and image-by-feature effect ($F(18, 124) = 2.660$, $p < 0.001$) for the 3-sec static images. That is, the viewing time was influenced by what images were presented and which features were present within these images. The data for the 1- and 3-sec static images are shown in Figure 2b and c, respectively. When collapsed across images, both 1- and 3-sec static image sets had a slightly different distribution of viewing times (eyes > mouth > midface) compared to the dynamic image (eyes > midface > mouth) set. For the 3-sec static set, the average time spent viewing the mouth versus the midface reached significance ($F(1, 124) = 0.897$, $p = 0.0058$).

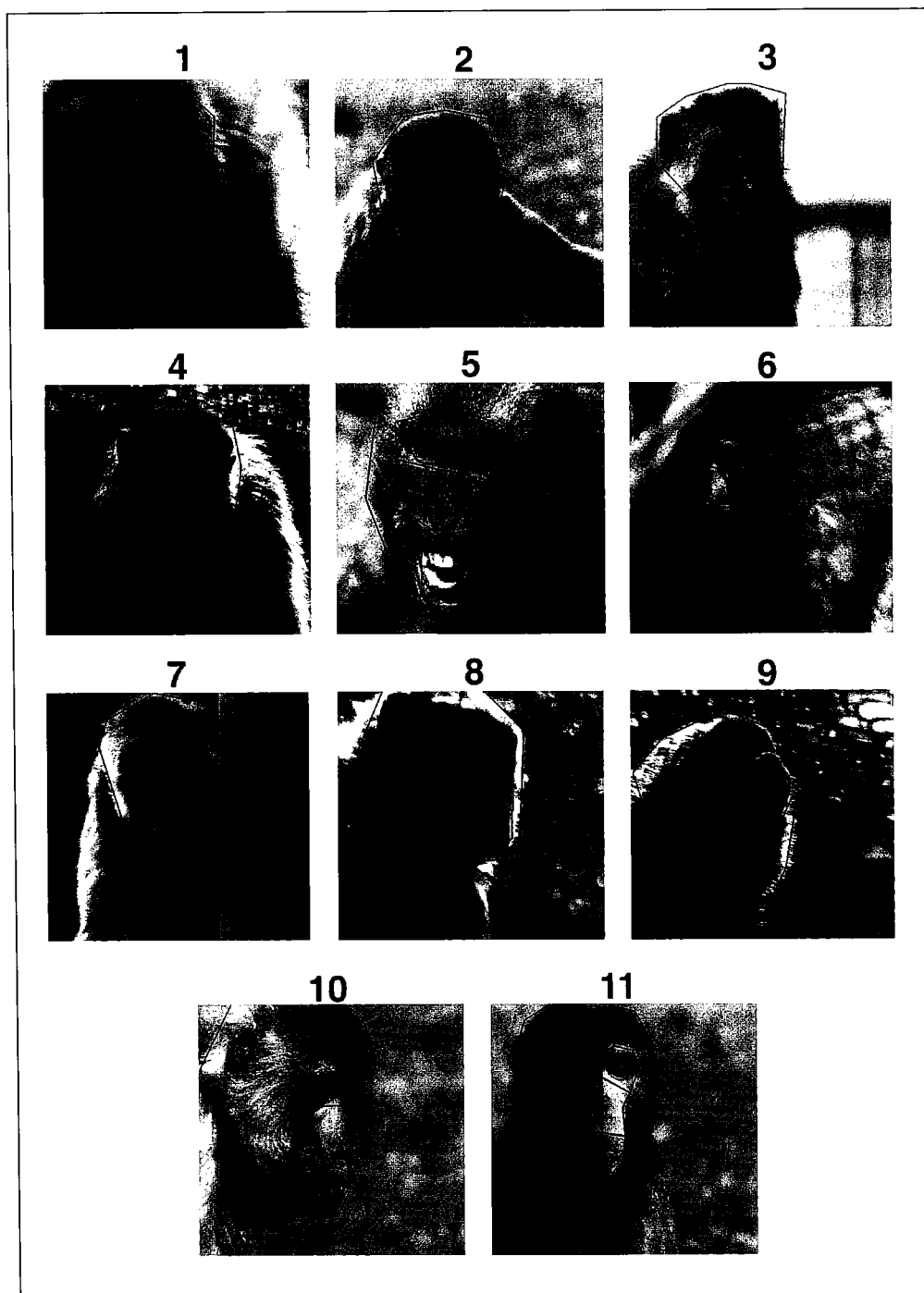
In cases where more than a single static image was culled from a dynamic image sequence, pairwise comparisons (unpaired t test) of viewing time for these associated static images revealed no significant differences. This was true for both 1- and 3-sec static presentations. Thus, average viewing patterns did not differentiate between the static images that were chosen to represent the 30-frame sequence.

A significant feature effect was observed for both 1- and 3-sec presentations when comparing the amount of time spent viewing all features of the face (eyes, midface, and mouth), other parts of the head (on-head), and other regions on the screen (1 sec: $F(2, 122) = 305.066$, $p < 0.001$; 3 sec: $F(2, 122) = 247.187$, $p < 0.001$). As with dynamic images, subjects spent significantly more time viewing features of the face than any other region of the visual image or computer display screen.

Effect of Facial Expression on Scan Patterns

To facilitate the detection of expression effects on scan patterns, images were assigned to groups representing similar facial displays. These assignments were subjective

Figure 1. Gray-scale reproductions of representative video frames from each of the eleven (1 through 11) dynamic color video sequences. Three major facial features ("botheyes," "midface," and "mouth") were defined by the polygons in each image using a graphic editor. The region within the polygon enclosing the head, minus the area defined by the three features, is defined as on-head. The area outside of this polygon extends to the border of the video display and is defined as "on-screen." Open-mouth threat, images 1 through 4; fear-grimace, image 5; lip-smack, images 6 through 8; yawn, image 9; neutral, images 10 through 11.



and based on the resemblance of each image to stereotypical facial displays as described by Redican (1975). Facial displays were assigned to one of five categories: open-mouth threat (OMT, $n = 4$); lip-smack (LS, $n = 3$); yawn (Y, $n = 1$); fear-grimace (FG, $n = 1$); neutral (N, $n = 2$). When the individual dynamic images were grouped according to expression, there was no significant feature-by-expression effect ($F(8,78) = 1.63, p = 0.13$). For 1-sec dynamic images, these data are shown in Figure 4A. The dynamic images that comprise each expression category are designated in Figure 2 by the lines beneath common expressive gestures.

For 1-sec static images grouped according to expression, a significant feature-by-expression effect emerged ($F(8,134) = 3.98, p < 0.001$). No feature-by-expression effect was observed for 3-sec static images grouped according to expression. The data for 1- and 3-sec stimulus sets grouped by expression are shown in Figure 4B and C, respectively.

Table 1 summarizes the comparisons between the amount of time spent viewing the eye region versus the mouth for each type of presentation (dynamic, 1- or 3-sec static) within a particular expressive grouping. Across different types of image presentation, we ob-

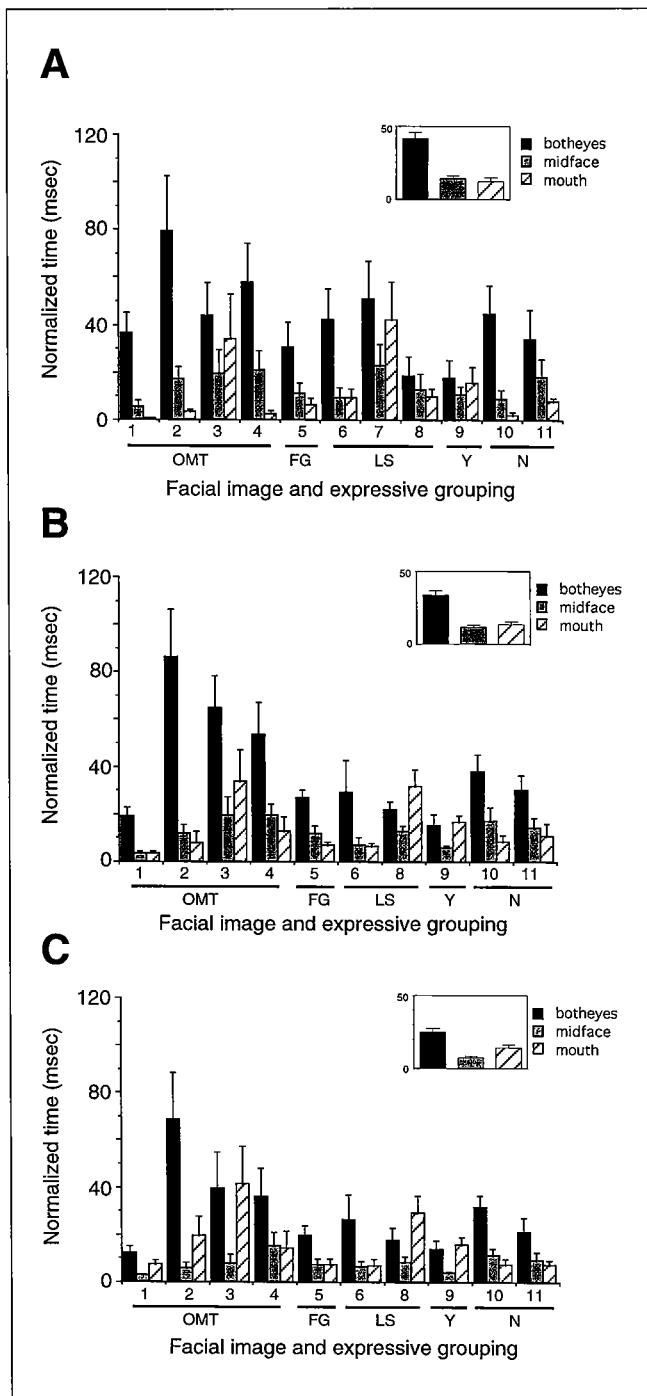


Figure 2. Average time (msec) subjects ($n = 4$) spent viewing each of the three major facial features (normalized by feature size) during the presentation of the 11 dynamic images (A), ten 1-sec static images (B), and ten 3-sec static images (C). The facial image numbers in parts (A), (B), and (C) correspond to the images shown in Figure 1. Images belonging to the same expressive grouping are denoted by a common line beneath the image labels. OMT = open-mouth threat, FG = fear-grimace, LS = lip-smack, Y = yawn, N = neutral. Inset graphs illustrate the same data collapsed across all images. Error bars indicate standard error of the mean.

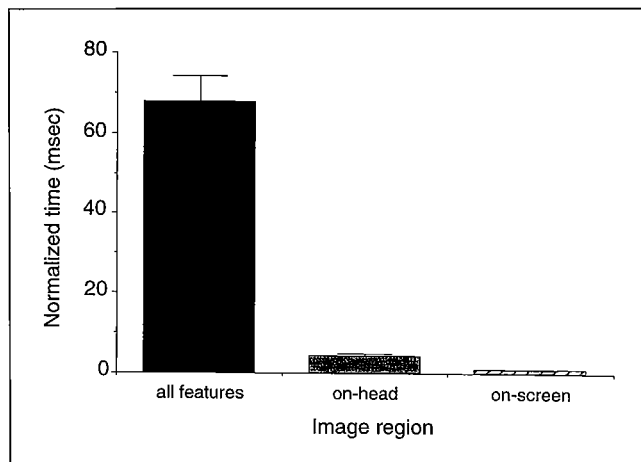


Figure 3. Average time (msec) four subjects visually inspected the face proper ("all features": botheyes, midface, and mouth), the area around the face ("on-head"), and other areas on the screen ("on-screen") collapsing across the 11 dynamic images. Viewing time normalized by feature size. Error bar indicates standard error of the mean.

served a consistent difference between the amount of time spent viewing the eye region versus the mouth of a neutral facial display. For both the dynamic and 1-sec static presentations of the open-mouth threat, the eyes were foveated significantly longer than the mouth. This was also true of the fear-grimace during the 1-sec static presentation. There were no significant differences between the amount of time spent viewing the eyes or mouth during the presentation of the lip-smack or yawn facial stimuli.

We also investigated, for the expressive groupings, whether the average amount of time spent viewing each feature varied as a function of image type. For example, does the presence of motion or the length of presentation affect the average amount of time a viewing monkey inspects the eye region during an open-mouth threat? Using expression as the nominal variable with three levels (eyes, midface, and mouth) of the dependent variable, the only significant difference we found was that the average amount of time spent viewing the midface during a 3-sec yawn was greater than that spent on the midface during the other presentation types. There was no difference in the average amount of time spent viewing the eyes during the dynamic, 1-sec static, or 3-sec static presentation of any other displays. This was also true for the average time spent viewing the midface and the mouth in these displays.

Viewing Patterns Elicited by Static Human Faces

Although primary emphasis in this study was on the nature of viewing patterns to conspecific facial gestures, we also explored the extent to which differential scan patterns could be evoked by static human faces. We used three different categories of expressions: neutral (frontal

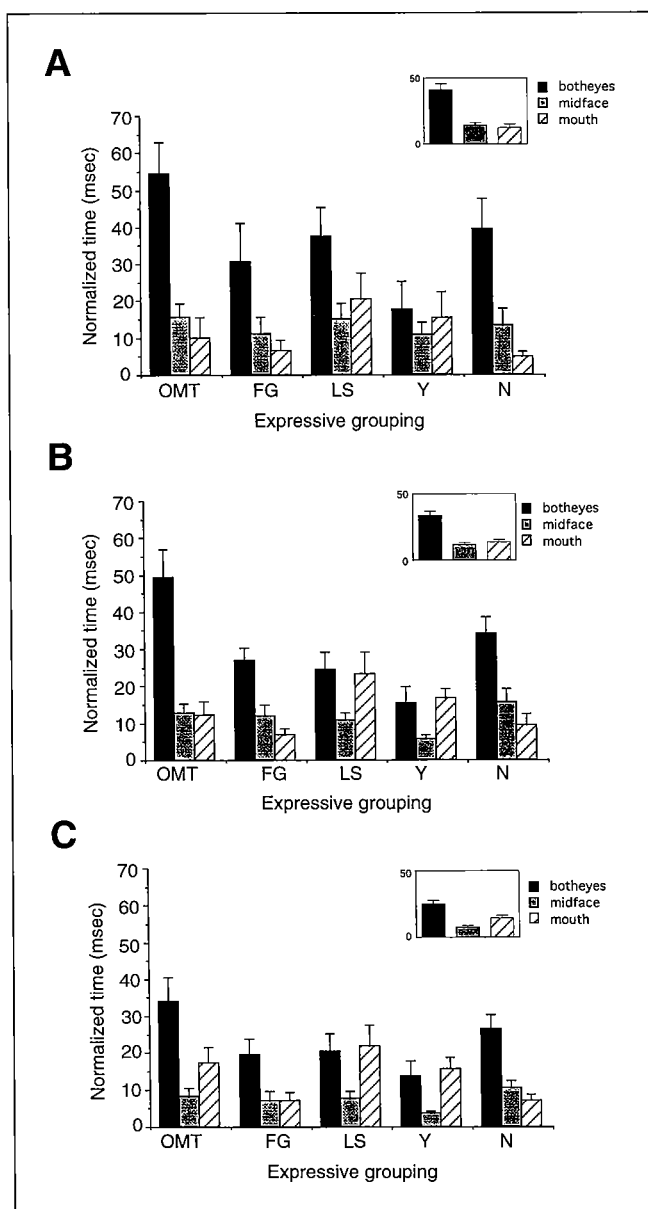


Figure 4. Average time (msec) four subjects visually inspected the three major facial features (normalized by feature size) during the presentation of 11 dynamic images (A), ten 1-sec static images (B), and ten 3-sec static images (C), which are grouped according to expression. The facial images which comprise these expressive groupings are shown in Figure 2. Abbreviations as in Figure 2. The inset graph illustrates the average time (msec) the four subjects visually inspected the three major facial features (normalized by feature size), collapsed across the five expression categories. Error bars indicate standard error of the mean.

and profile), open-mouth, and smile. Repeated measures ANOVA revealed a significant feature effect for the 1-sec static ($F(2, 60) = 45.122, p < 0.001$), and the 3-sec static human faces ($F(2, 60) = 24.953, p < 0.001$). When collapsing across images, the profile of responses to both 1- and 3-sec static human faces resembled those for both dynamic and 1-sec static monkey images (eyes \gg midface $>$ mouth).

When grouped according to expression, the amount

Table 1. Unpaired Comparison of Average Time Spent on Eyes versus Mouth for Each Stimulus Set and Each Expression.

Stimulus set	Expression ^a	Eyes vs. Mouth	
		<i>t</i> value	<i>p</i>
Dynamics	OMT	4.384	< .001
	FG	2.192	.0709
	LS	1.602	.1233
	Y	.203	.8459
	N	4.067	.0012
1-sec statics	OMT	4.365	< .0001
	FG	5.158	< .001
	LS	0.138	.8913
	Y	-0.277	.7844
	N	4.380	< .001
3-sec statics	OMT	2.074	.0437
	FG	2.498	.0256
	LS	-0.179	.8594
	Y	-0.412	.6844
	N	4.445	< .001

^a OMT = open-mouth threat, FG = fear grimace, LS = lipsmack, Y = yawn, N = neutral.

and pattern of viewing time was significantly affected by both features ($F(2, 74) = 35.063, p < 0.001$) and the interaction between features and expression ($F(4, 74) = 3.937, p = 0.0059$) in 1-sec images. A significant feature effect ($F(2, 74) = 17.48, p < 0.001$) was observed during the presentation of 3-sec human faces. No other significant effects were seen.

When comparing the amount of time spent viewing the screen region (face, head, and on-screen), both 1- and 3-sec presentations of human images elicit viewing patterns similar to the dynamic images. Subjects spent a significantly greater length of time viewing the features of the face than any other region of the screen (1 sec: $F(2, 60) = 108.66, p < 0.001$; 3 sec: $F(2, 60) = 80.144, p < 0.001$).

Contour Plots

To provide a more quantitative sense of how monkeys look at faces, we constructed contour plots from recorded eye positions and superimposed these plots on the appropriate facial images. Only data for the static presentations were plotted since the correspondence

between a superimposed contour plot and the appropriate dynamic facial stimuli is confounded by the fact that the facial display on the viewing monitor varies over time.

Figure 5A through D depicts four facial images, each representing a different expression category defined in this study (A, open-mouth threat; B, fear-grimace; C, yawn; and D, neutral). Superimposed upon these images are contour plots derived from the data collected during the 3-sec presentation of each image. It is important to note that these contour plots reflect the total number of eye positions, not the data normalized for area of the feature in the image, which was used for statistical analyses.

DISCUSSION

The goal of this study was to address three questions related to the perception of facial displays by rhesus

monkeys. First, is there a difference between the manner in which monkeys view real-time dynamic versus static facial displays? Second, when given a choice to freely inspect two-dimensional images of conspecific and human faces, to what extent do monkeys visually explore the face and which facial features are most often inspected? Third, do facial displays differing in expressive content evoke different patterns of viewing? By presenting both real-time and static facial displays, we have shown that temporal changes of these facial features are not an important determinant in the evocation of stereotypical viewing patterns. Our study was designed to investigate the fixation patterns to facial images presented in a controlled laboratory setting and thus leaves open the question of what extent facial motion may yet affect the general state of orientation and arousal or the generation of a particular behavioral response.

In regard to which features were most often visually

Figure 5. Reproductions of four static video frames, each representing a different expressive grouping: (A) open-mouth threat, (B) fear-grimace, (C) yawn, (D) neutral. Contour plots superimposed on each image represent the total number of eye-position data points collected from four subjects over the course of 25 three-second static presentations. The location of the contour plots represents the horizontal and vertical eye-position data relative to the facial image, whereas the contour interval reflects the relative density of recorded eye-position measurements.



inspected, the eyes were the predominant target of monkeys viewing high-resolution monkey and human facial displays. Since our data using human faces may have been confounded by the monkeys' familiarity with the individual in the video-displayed image, we will confine our discussion of expression and viewing patterns to the data using conspecific facial displays. The major finding of our study, accordingly, was that the expressive nature of a facial display affects the relative amount of time spent foveating different facial features, and the viewing patterns defined by these foveations can segregate facial displays along agonistic and affiliative dimensions.

Viewing Patterns for Static versus Dynamic Facial Expressions

All previous studies that have investigated the viewing patterns of monkeys to facial images have used static photographic slides (Keating & Keating, 1982) or other forms of static images such as photographs (Wilson & Goldman-Rakic, 1994). Since expressive facial gestures are dynamic by nature, we investigated the differences between viewing patterns to real-time facial displays and static representations of the same gesture. When comparing the amount of time spent viewing a particular feature, there were no statistically significant differences among the different presentation types. The finding that facial motion did not consistently affect viewing patterns to expressive facial displays was unexpected. For example, subjects spent an equivalent amount of time viewing the eyes during the dynamic, 1-sec static, and 3-sec static presentation of a fear-grimace. This was also true for the amount of time spent viewing the midface and the mouth. This invariance of viewing time to presentation type was observed for all of the facial displays, as well as when comparing across presentation type using the data grouped according to expression. The only difference we observed was that the amount of time spent viewing the midface in the dynamic yawn was significantly greater than 1- or 3-sec static viewing.

In our study, the presence of motion in facial displays was not essential for the evocation of stereotypical viewing patterns nor did it affect in any consistent way the amount of time spent exploring facial features defined in our study. From these data we conclude that the emotive properties of expressive facial displays, as measured by viewing patterns, are not related to the changing nature of the face or the generation of a facial display but to the static physiognomic characteristics of the face and their relative configuration. Our conclusions should not be taken as evidence that dynamic facial displays play no role in the study of nonhuman primate visual mechanisms of behavior; facial motion, and the variation in facial cues may indeed serve as an important cue in orienting, attention, and determining the type and degree of behavioral response to facial displays. Further

studies using dynamic facial images as stimuli in behavioral testing paradigms are needed to investigate these effects.

Which Facial Features Are Most Often Inspected?

When freely inspecting both dynamic and static conspecific facial images, as well as static human faces, subjects spent a significantly longer percentage of time viewing the face, as opposed to other features on the head or other regions of the viewing monitor. This was true despite variances in the size of the head on the viewing monitor or the amount of body and limbs present in the video images. From these data we conclude that central facial features constitute an inherently compelling visual image, and when given a choice, viewing monkeys will preferentially explore the face rather than other parts of the body or surrounding objects. Within the face proper, the eyes were consistently inspected by viewing monkeys, as were other facial features such as the midface and the mouth. This was true for both conspecific and human faces. Even in our study, which utilized high-resolution images providing fine detail of the mouth, teeth, and other features of the face and contours of the head, the inspection of the eyes occupied an overwhelming proportion of subjects' viewing time when the data was normalized for the size of the feature subtended on the display monitor. These findings are thus consistent with previous studies using both monkey (Keating & Keating, 1982; Wilson & Goldman-Rakic, 1994) and human faces as visual stimuli (Keating & Keating, 1993).

The amount of time the monkeys' eyes were recorded off the viewing monitor was minimal but was significantly greater during the presentation of the 3-sec static images. Because there was no difference between this value for the 1-sec static and dynamic images, it seems unlikely that the presence of motion can account for this difference. That is, it is not the case that the dynamic images were simply more visually compelling. A more parsimonious explanation for these findings is that 3-sec displays fatigue monkeys' immediate attentive capacity when viewing faces. From these data we conclude that stimulus presentations of 1 sec in duration are sufficiently long enough to assess monkeys' viewing patterns to expressive two-dimensional facial displays.

Viewing Patterns for Different Expressive Groupings

Although the eyes were the most salient feature, the relative amount of time spent viewing the eyes, midface, and mouth was dependent upon the nature of the facial display, and a number of interesting and unexpected effects of facial expression on viewing patterns emerged. Before turning to this discussion, a number of methodological points deserve mention.

Although the assignment of facial displays into expressive categories was done entirely by visual inspection, the subjective classification of displays into these categories was not problematic because the assignments were based on relatively unambiguous similarities in facial configuration. A caveat to this argument is the fact that a significantly shorter amount of time was spent viewing image number 1 of four representing the open-mouth threat expression. Although the overall pattern of viewing for the four open-mouth threat expressions were similar, this finding suggests that some of the stimuli contained visual properties affecting viewing patterns for which we did not control. Inspection of open-mouth threat image number 1 reveals a number of differences, such as the relative magnification of the head or coloration of the fur, that may have accounted for this difference. In short, although further studies are needed to investigate the effect that these and other visual properties have on viewing patterns to facial images, we do not believe that the general validity of our conclusions are compromised since the data from which these conclusions are drawn were based on the averaged viewing time across images within an expressive grouping. A more general problem related to the expressive groupings defined in our study is that the images might be emotionally ambiguous for the monkeys viewing them since the behavioral context in which these facial displays are normally observed is absent. Despite the lack of behavioral context normally provided by naturalistic settings, visual inspection patterns in a controlled experimental setting help to quantitate the relative importance of different facial features as visual stimuli. Further studies are required to investigate the effect that rewarding or aversive behavioral situations might have on viewing patterns observed in this study.

Alternatively, the absence of behavioral context can be used to our advantage, since the disinhibited monkey, not subject to the behavioral consequences of prolonged fixation, may not necessarily change what it looks at, but the duration of the looking. Neuroethological studies have repeatedly shown that gaze aversion is one of the most frequently observed means for avoiding agonistic social interactions (Altman, 1967; Chance, 1962). In our study, the monkeys looked directly at the eyes for a significantly longer proportion of time than other regions of the screen. Thus, the subjects in our study were not averting their gaze, even to frontal open-mouth threat displays. These findings are similar to those of Humphrey (1972), who showed that "interest" is a more powerful determinant of monkeys' viewing patterns than the rewarding or aversive cue associated with visual images. That is, the perceptually compelling nature of the face seems capable of overriding the learned behavioral response. There is thus considerable reason to believe that the amount of time inspecting the eyes in the laboratory setting does not reflect a qualitative

change in viewing pattern. Rather, the monkeys simply look longer at what they furtively examine in the social context.

The major influence that we observed related to facial expression was the relative amount of time spent viewing specific facial features such as the eyes and the mouth within a particular expressive grouping. During the presentation of the dynamic and 1-sec static open-mouth threat stimuli, a significantly larger percentage of time was spent on the eyes as compared to the mouth. Also, significantly more time was spent viewing the eyes relative to the mouth in the 1-sec static fear-grimace and the 1- and 3-sec static neutral display. Although this difference was not significant for the dynamic fear-grimace or the 3-sec fear-grimace, values for both presentation types tended toward the pattern for the open-mouth threat. For the lip-smack and yawn facial displays, monkeys spent relatively more time exploring the mouth, and during no static or dynamic presentation of these facial expressions was the amount of time spent viewing the eyes significantly greater than the time spent viewing the mouth. One potential explanation for these data is that the mouth region is simply less perceptually compelling in the open-mouth threat and fear-grimace stimuli than in the lip-smack or yawn stimuli. We do not favor this argument, however, since the fear-grimace image that we used contains one of the most vivid mouth and teeth configurations in our entire stimulus set. There are, however, a number of interesting and perhaps more reasonable interpretations of these data.

First, visual inspection of the mouth region in facial displays denoting agonistic encounters might be inherently aversive for viewing monkeys. This aversive quality could then be used to trigger the appropriate set of behavioral reactions given the agonistic nature of the facial display. Second, the configuration of the mouth in facial displays might provide reliable visual clues of an agonistic nature that require only brief visual inspection. Lastly, the functional explanation we favor as to why the eyes are watched so closely is that information pertaining to the presence and/or direction of threat is best betrayed by the eyes. In the open-mouth threat, the fear-grimace, and to a lesser extent the neutral facial images used in our study, the direction of gaze was consistently directed toward the viewing monkey. Taken together with the predominance of eye positions recorded during the presentation of these visual stimuli, this suggests that viewing patterns to agonistic or socially neutral encounters are largely influenced by the perceptually compelling nature of direct eye contact.

In conclusion, it is clear from an ethological perspective that the open-mouth threat and the fear-grimace belong to the class of facial gestures that are used during agonistic social encounters. On the other hand, the lip-smack is perhaps the most affiliative gestural display (Chevalier-Skolnikoff, 1973; Redican, 1975), and the yawn

has been interpreted as a gesture related to the release of tension or to appeasement (Redican, 1975). Although our stimuli were a priori assigned to five expressive categories, this was exclusively based on their perceptual similarity, not their presumed emotional salience. Unexpectedly, the characteristic viewing patterns to expressive facial displays that we have found can further segregate these stimuli, a posteriori, along the more superordinate dimensions of agonistic or affiliative. This is to our knowledge the first evidence that the emotional and behavioral salience of a facial expression can be delineated on the basis viewing patterns—that is, how monkeys look at the face.

METHODS

Animal preparation

Subjects

Four adult female rhesus monkeys (*Macaca mulatta*) weighing between 6 and 9 kg were used in this study. The monkeys were housed, unrestrained, in separate cages except during the experimental sessions, wherein they were restrained in a standard primate chair. The procedures employed in these experiments were approved by the Salk Institute Animal Care and Use Committee and conform to USDA regulations and NIH guidelines for the humane care and use of laboratory animals.

Surgical Preparations

After initial acclimation to the primate chair, the subjects were prepared for fixation training using surgical procedures that have been previously published (Dobkins & Albright, 1994). All surgical procedures were performed under aseptic conditions using barbiturate anesthesia (sodium pentobarbital, 25 mg/kg i.v. bolus, followed by continuous infusion at a rate of 3.5 mg/kg/hr).

A stainless steel head-post was affixed to the skull in the four experimental subjects using stainless steel skull screws, orthopedic plates (Synthes), and dental acrylic. A search coil for measuring eye position was surgically implanted beneath the conjunctiva in one eye of all four subjects using the method of Judge, Richmond, and Chu (1979). Postsurgical follow up is described in Dobkins and Albright (1994).

Behavioral Fixation Training

The subjects were trained on a simple fixation task using conventional behavioral procedures (Dobkins & Albright, 1994). The purpose of this task was to direct the subjects' gaze to the center of the screen. During the stimulus-on period, the subject was freely allowed to visually explore the entire extent of the computer monitor. A typical fixation trial consisted of the following

sequence. The trial was initiated by the onset of the fixation target at the center of the screen. If the monkey failed to fixate the target within the designated period of time (usually 5 sec), the target disappeared for approximately 1 sec and a new trial was initiated. Fixation was successful if maintained for 250 msec within a fixation window of 1.0 to 2.0°. Following this short fixation period, the fixation target was then turned off, the fixation window enlarged to encompass a large area (60°), and the appropriate test visual stimuli were presented for either 1 or 3 secs. This large fixation window during the stimulus-on period allowed us to measure both eye movements within the area defined by the stimulus and off the viewing monitor. The latter provided an indication of whether or not the viewing monkey was making saccades away from the monitor during stimulus presentations. After termination of the stimulus period the animal received a small drop of juice, and a variable intertrial delay ensued until the next trial was initiated.

Animal subjects were placed on water-restricted diets, and behavioral control was achieved using a liquid reward. Behavioral control, stimulus presentation, and data acquisition was automated using a PDP 11/73 microcomputer. The fixation target was generated using a VGA computer graphics board resident in a 386 personal computer.

Acquisition of Video Stimuli

Two-hundred minutes of unedited video footage of free-ranging monkeys (*M. mulatta*) were collected over the course of three days at the California Regional Primate Research Center, Davis, California. All filming was conducted in three separate 1.5-acre outdoor corrals housing 120 to 150 mixed-sex social groups of rhesus macaques of varying age and size. This environment provided an excellent opportunity to film a wide range of species-specific facial displays in a naturalistic social setting. An effort was made to film only extremely close-up views of behaving monkeys so that the face region occupied the extent of the viewfinder. Although we were able to film some varieties of facial displays between monkeys, most of the stereotypical displays we subsequently used in these experiments were acquired when the monkey displayed at the cameraman or the accompanying animal technician.

Color video footage was acquired with a production-quality Sony Betacam recorder fitted with a 12:1 zoom lens. A total of eight 30-min Betacam videotapes (3M, Betacam PB 30) were collected over the course of the three days of filming. Each of these master tapes was then transferred onto standard VHS medium along with a window dub (frame counter) using a Sony Betacam Player Recorder (BVW-75). These VHS tapes were inspected for video footage containing facial displays, and selected segments were noted by their corresponding frame numbers. Using these frame numbers, the selected

segments were located on the original Betacam tapes and copied onto a 1-in master videotape (3M, C Format) using a Sony 3100 1-in Type C Format Player Recorder. This master videotape thus contained the entire collection of video footage from which all stimuli were subsequently chosen (see below). The entire collection of video clips stored on the 1-in master was then transferred to a double-sided optical laser disc (Panasonic, TQ-FH332) using a Panasonic Laser Disc Recorder (Panasonic, TQ-3031F).

Stimulus Presentation

Experimental Apparatus

Video images stored on an optical laser disc were played back using an optical laser disc player (Panasonic TQ-3032F). The double-sided disc we employed is capable of storing 108,000 individual frames. At a playback rate of 30 frames/sec, 30 min of continuous motion is stored on each side, a total of 60 min for a double-sided disc. Any stored video image or video image sequence could be accessed by simply specifying the optical disc frame address(es) (see below).

The video playback options were under the control of the PDP 11/73. Using a RS-232C computer interface card installed in the laser disc player and on-line control commands, we specified such options as the direction of play (forward, reverse) and frame address.

The output of the laser disc player was a composite NTSC video signal that was first sent to a real-time digital frame grabber (VIP-8000, IEV Corp.) residing in a personal computer. The digital conversion of the NTSC signal by the frame grabber allowed the video signal to be presented on a high-resolution graphics monitor. The frame grabber board was also equipped with VGA capabilities, which allowed the integration of real-time video with computer-generated graphics (such as a fixation point). This composite video signal was displayed on an RGB video monitor for viewing by the experimental subject (NEC Multisync XL 19-in color video monitor, RGB, 60 Hz, 712 lines, noninterlaced). The displayed images occupied the entire screen of the viewing monitor.

Visual Stimulus Sets

The color video images stored on the optical laser disc were a collection of facial displays previously edited from a larger library of raw video footage. From the laser disc collection of video images, 11 facial displays differing in expression and identity of the displayer were chosen as stimuli for this experiment. Each dynamic stimulus was 1-sec in duration and consisted of 30 consecutive video frames. The five facial displays included in this stimulus set were open-mouth threat, fear-grimace, lip-smack, yawn, and neutral. There were a total of six monkey identities exhibiting one or more of these facial

displays. The identity of these monkeys was also unknown to all the experimental monkey subjects from which eye-position data were collected. In cases where there was a significant amount of motion over the course of 30 frames, several static images from that particular dynamic sequence were selected to represent the corresponding dynamic facial display. A minimum of one and a maximum of three static images were culled from each of the 11 dynamic stimuli. These static images were presented for 1 and 3 sec durations. In one case, no clear static image could be obtained from the dynamic sequence (dynamic image number 7).

In addition to presenting static and dynamic facial displays of conspecific monkeys, static images of human facial displays were used. Three human facial expressions were employed: neutral, open-mouth, and smile. Two of four human subjects from which these images were collected were familiar to viewing monkeys. A total of 10 stimuli comprised this visual data set.

Eye-Movement Recording

Subjects' eye positions were continuously monitored using the Robinson (1963) magnetic search-coil technique. Each subject sat in a standard primate chair and viewed the stimulus monitor from a distance of 60 cm. The visual angle subtended by the viewing monitor was 40°. Offset and gain of the horizontal and vertical eye position signals were calibrated prior to each experimental session by placing a fixation target at various eccentricities and modifying the gain and phase of the eye-coil monitoring system so that the eye positions were centered on each target. Horizontal and vertical eye positions were stored every 8 msec, with a maximum standard error of approximately 0.25°.

Dynamic and static facial displays were presented in a random trial order. Eye positions were collected during each trial from the period beginning 500 msec prior to fixation point onset and ending 500 msec after stimulus offset. These data were stored on-line using the PDP 11/73 microcomputer and were later transferred to magnetic tape and reformatted for analysis on a Silicon Graphics Iris Elan R-4000. The intertrial interval was a minimum of 2 sec in duration.

On average, five trials of each stimulus (dynamic and static) were presented during each experimental session. Over the course of the entire experiment, each subject saw five sets of these five trial presentations. Thus, each stimulus was presented 25 times.

Eye-Movement Data Analysis

Superimposing the eye-movement data on a visual image gives a qualitative impression of the viewing patterns to different facial images. In order to provide a more quantitative analysis of viewing patterns, we adopted and modified a method used by Keating and Keating (1982).

This method consists of dividing each image into regions representing major facial features. For this study we defined three major regions: (1) both eyes, (2) midface, and (3) mouth. Two other minor regions consisted of (1) the area surrounding the face proper (on-head) and (2) all other regions outside of the head and face proper but within the boundaries of the video monitor.

The first step in defining these regions was the conversion of the video stimuli into a digital format. This was accomplished by converting and saving each frame (30 frames for a dynamic sequence) in a targa format using a PC-based digital frame grabber. A graphic editor was developed for tracing closed lines around each of the major facial features in an RGB image. These digitized boundaries were used to define the position and pixel density of each feature. This procedure, which was performed on each frame for each dynamic stimulus, provided an accurate method of delimiting the spatial location of the major facial features over time. A gray-scale static representation of each (i.e., a single representative frame from each sequence) of the 11 dynamic images with the major facial features defined by this technique is shown in Figure 1.

After determining the position of facial features using this technique, we were able to analyze the eye-fixation data in a number of interesting ways. The horizontal and vertical coordinates of eye-position data points were scaled to correspond to pixel coordinates in the video display. This digitization of eye position in stimulus coordinates allowed us to determine the facial feature associated with all recorded eye positions. The number of eye positions associated with each feature was subsequently converted into units of time (msec). To account for the varying size of features within and between different stimuli, the time spent viewing each feature was normalized by pixel density associated with each feature. Furthermore, to allow comparisons between stimulus presentation lengths of 1 and 3 sec, viewing time is rescaled as the number of milliseconds per trial, per second. All results are normalized for the size of the feature and presented as the number of milliseconds per trial, per second, and per 1000 pixels.

Although the fixation point was not present during the presentation of each visual stimulus, there was a period of time immediately following the disappearance of the fixation point, coinciding with the onset of the visual stimuli, where the eyes of viewing monkeys were still directed to the center of the screen. Average latency for initiating voluntary saccades is approximately 200 msec (Becker, 1991). For this reason eye positions recorded during the first 200 msec following fixation spot offset were not included in the data analyses. All data were averaged across blocks of five presentations. A total of five blocks were shown to each subject.

Contour plots were constructed using a 15×15 matrix that subtended the area bounded by the visual stimulus. The height of each zone corresponds to the

number of eye positions associated with each zone, summed over 25 presentations for each stimulus.

Statistical tests were performed (as indicated) using analysis of variance (ANOVA), single pairwise comparisons (t tests), Fisher PLSD and Tukey-Kramer post-hoc tests, and linear contrast analysis using Statview version 4.01 and SuperANOVA version 1.11 implemented on a Macintosh II personal computer. The dependent variable in all cases was the average viewing time (msec) per trial, per 1000 pixels, per second.

Acknowledgments

This work was supported by Graduate Fellowships from the McDonnell-Pew Center for Cognitive Neuroscience at San Diego (FKDN), NSF Dissertation Research Improvement Grant BNS-9013202 (FKDN, DGA), and NIMH grant MH41479 (DGA). A portion of this work was conducted at the California Regional Primate Center (NIH RR00169). We wish to thank Theodore Bullock, Fraser Wilson, Gene Stoner, Bill Mason, and Peter Marler for helpful comments and Jennifer Costanza for her superb technical assistance.

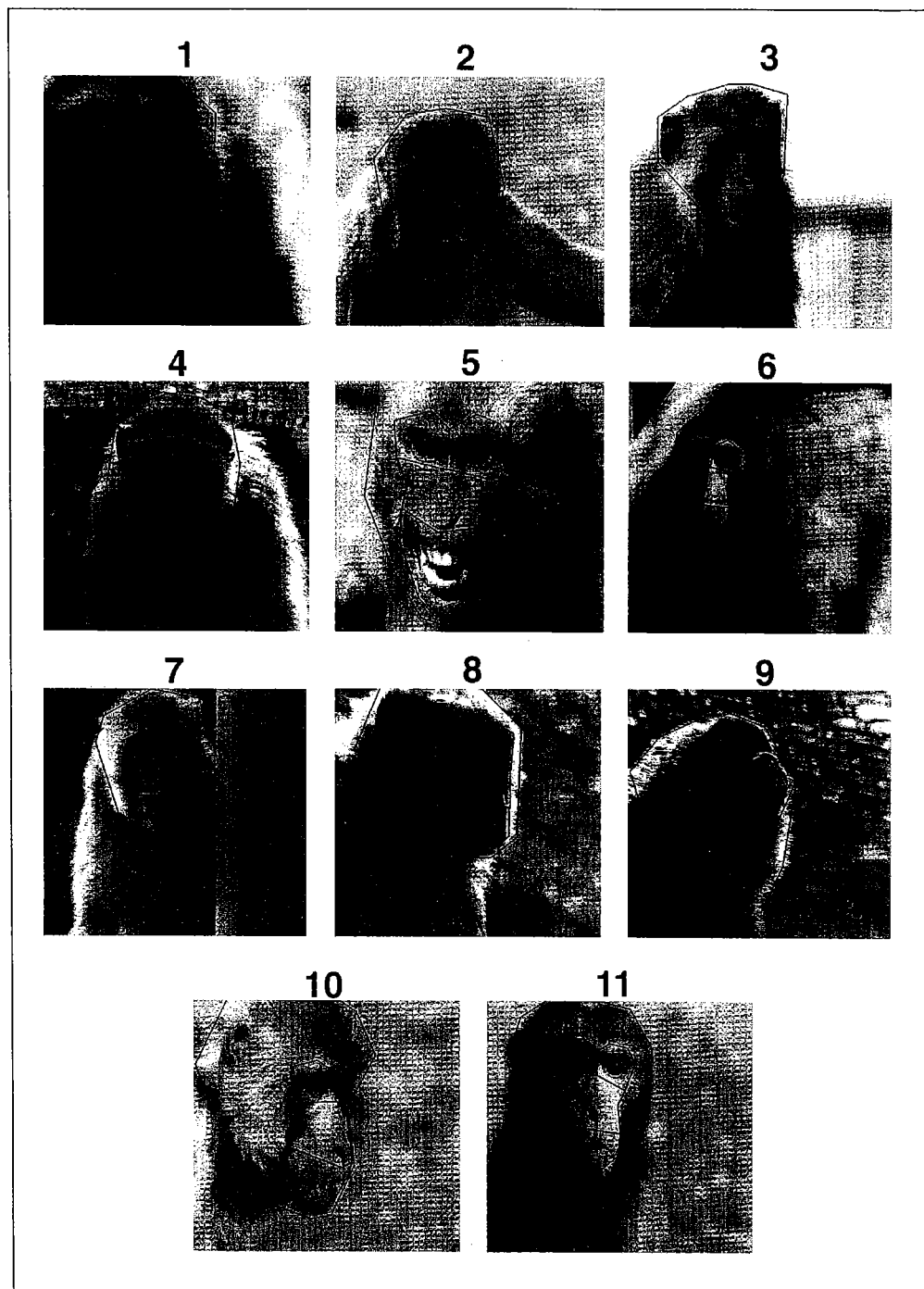
Reprint requests should be sent to Frederick K. D. Nahm, M.D., Ph.D. Harvard Medical School Neurology Program, Beth Israel Deaconess Medical Center, East, Dept. of Neurology, 330 Brookline Avenue, Boston, MA 02215.

REFERENCES

- Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, 372, 669-672.
- Altman, S. (1967). The structure of primate social communication. In S. Altman (Ed.), *Social communication among primates*. Chicago: University of Chicago Press.
- Becker, W. (1991). Saccades. In R. H. S. Carpenter (Ed.), *Vision and visual dysfunction*, Vol. 8, "Eye movements," Ch. 5. Boca Raton: CRC Press.
- Brothers, L. (1990). The neural basis of primate social communication. *Motivation & Emotion*, 14, 81-91.
- Butler, R. A. (1961). The responsiveness of rhesus monkeys to motion pictures. *Journal of Genetic Psychology*, 98, 239-245.
- Chance, M. R. A. (1962). An interpretation of some agonistic postures; the role of "cut-offs" acts and postures. *Symposia of the Zoological Society of London*, 8, 71-89.
- Chevalier-Skolnikoff, S. (1973). Facial expression of emotion in nonhuman primates. In P. Ekman (Ed.), *Darwin and facial expression*. New York, London: Academic Press.
- Darwin, C. (1872). *The expression of the emotion in man and animals*. London: J. Murray.
- Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience*, 4, 2051-2062.
- Dicks, D., Meyers, R. E., & Kling, A. (1969). Uncus and amygdala lesions: Effects on social behavior in the free-ranging rhesus monkey. *Science*, 165, 69-71.
- Dobkins, K. D., & Albright, T. D. (1994). What happens if it changes color when it moves?: Neurophysiological experiments on the nature of chromatic input to macaque visual area MT. *Journal of Neuroscience*, 14(8), 4854-4870.
- Ekman, P., Friesen, W. V., & Ellsworth, P. (1982). Methodologi-

- cal decisions. In P. Ekman (Ed.), *Emotion in the human face*. Cambridge: Cambridge University Press.
- Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology*, 35, 96-111.
- Hasselmo, M. E., Rolls, E. T., & Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioral Brain Research*, 32, 203-218.
- Herzog, A. G., & Van Hoesen, G. W. (1976). Temporal neocortical afferent connections to the amygdala in the rhesus monkey. *Brain Research*, 115, 57-69.
- Hinde, R. A., & Rowell, T. E. (1962). Communication by postures and facial expressions in the rhesus monkey (*Macaca mulatta*). *Proceedings Zoological Society of London*, 138, 1-21.
- Humphrey, N. K. (1972). 'Interest' and 'pleasure': Two determinants of a monkey's visual preferences. *Perception*, 1, 395-416.
- Iwai, E., & Yukie, M. (1987). Amygdalofugal and amygdalopetal connections with modality-specific visual cortical areas in macaques (*Macaca fuscata*, *M. mulatta*, and *M. fascicularis*). *Journal of Comparative Neurology*, 261, 362-387.
- Judge, S. J., Richmond, B. J., & Chu, F. C. (1979). Implantation of magnetic search coils for measurement of eye position: An improved method. *Vision Research*, 20, 535-538.
- Keating, C. F., & Keating, E. G. (1982). Visual scan patterns of rhesus monkeys viewing faces. *Perception*, 11, 211-219.
- Keating, C. F., & Keating, E. G. (1993). Monkeys and mug shots: cues used by rhesus monkeys (*Macaca mulatta*) to recognize a human face. *Journal of Comparative Psychology*, 107(2), 131-139.
- Marler, P. (1965). Communication in monkeys and apes. In I. Devore (Ed.), *Primate behavior: Field studies of monkeys and apes*. New York: Holt-Rinehart and Winston.
- Miller, R. E. (1967). Experimental approaches to the physiological and behavioral concomitants of affective communication in rhesus monkeys. In S. Altman (Ed.), *Social communication among primates*. Chicago: University of Chicago Press.
- Nakamura, K., Mikami, A., Kubota, K. (1992). Activity of single neurons in the monkey amygdala during performance of a visual discrimination task. *Journal of Neurophysiology*, 67(6), 1447-1463.
- Overman, W. H. Jr., & Doty, R. W. (1982). Hemispheric specialization displayed by man but not macaques for analysis of faces. *Neuropsychologia*, 20, 113-128.
- Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, 47, 329-342.
- Perrett, D. I., Mistlin, A. J., & Chitty, A. J. (1987). Visual neurons responsive to faces. *Trends in Neuroscience*, 10, 358-364.
- Redican, W. K. (1975). Facial expressions in nonhuman primates. In L. A. Rosenblum (Ed.), *Primate behavior: Developments in field and laboratory research*, vol. 4. New York: Academic Press.
- Robinson, D. A. (1963). A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE Transactions on Biomedical Engineering*, 10, 137-145.
- Rosvold, H. E., Mirsky, A. F., & Pribram, K. H. (1954). Influence of amygdectomy on social behavior in monkeys. *Journal of Comparative Physiology and Psychology*, 47, 173-178.
- Sackett, G. P. (1965). Responses of rhesus monkeys to social stimulation presented by means of colored slides. *Perceptual and Motor Skills*, 20, 1027-1028.
- Steklis, H. D., & Kling, A. (1985). Neurobiology of affiliative behavior in nonhuman primates. In M. Reite & T. Fields (Eds.), *The psychobiology of attachment and separation* (93-129). New York: Academic Press.
- Turner, B. H., Mishkin, M., & Knapp, M. (1980). Organization of the amygdalopetal projections from modality-specific cortical association areas in the monkey. *Journal of Comparative Neurology*, 191, 515-543.
- van Hooff, J. A. R. A. M. (1967). The facial displays of the catarrhine monkeys and apes. In D. Morris (Ed.), *Primate ethology*. Chicago: Aldine Publishing Company.
- Wilson, F. W. W., & Goldman-Rakic, P. S. (1994). Viewing preferences of rhesus monkeys related to memory for complex pictures, colours and faces. *Behavioural Brain Research*, 60(1), 79-89.
- Yamane, S., Kaji, S., & Kawano, K. (1988). What facial features activate face neurons in the inferotemporal cortex of the monkey? *Experimental Brain Research*, 73, 209-214.
- Zola-Morgan, S., Squire, L. R., Alvarez-Royo, P., & Clower, R. P. (1991). Independence of memory functions and emotional behavior: Separate contributions of the hippocampal formation and the amygdala. *Hippocampus*, 1(2), 207-220.

Figure 1. Gray-scale reproductions of representative video frames from each of the eleven (1 through 11) dynamic color video sequences. Three major facial features ("botheyes," "midface," and "mouth") were defined by the polygons in each image using a graphic editor. The region within the polygon enclosing the head, minus the area defined by the three features, is defined as on-head. The area outside of this polygon extends to the border of the video display and is defined as "on-screen." Open-mouth threat, images 1 through 4; fear-grimace, image 5; lip-smack, images 6 through 8; yawn, image 9; neutral, images 10 through 11.



and based on the resemblance of each image to stereotypical facial displays as described by Redican (1975). Facial displays were assigned to one of five categories: open-mouth threat (OMT, $n = 4$); lip-smack (LS, $n = 3$); yawn (Y, $n = 1$); fear-grimace (FG, $n = 1$); neutral (N, $n = 2$). When the individual dynamic images were grouped according to expression, there was no significant feature-by-expression effect ($F(8,78) = 1.63, p = 0.13$). For 1-sec dynamic images, these data are shown in Figure 4A. The dynamic images that comprise each expression category are designated in Figure 2 by the lines beneath common expressive gestures.

For 1-sec static images grouped according to expression, a significant feature-by-expression effect emerged ($F(8,134) = 3.98, p < 0.001$). No feature-by-expression effect was observed for 3-sec static images grouped according to expression. The data for 1- and 3-sec stimulus sets grouped by expression are shown in Figure 4B and C, respectively.

Table 1 summarizes the comparisons between the amount of time spent viewing the eye region versus the mouth for each type of presentation (dynamic, 1- or 3-sec static) within a particular expressive grouping. Across different types of image presentation, we ob-

between a superimposed contour plot and the appropriate dynamic facial stimuli is confounded by the fact that the facial display on the viewing monitor varies over time.

Figure 5A through D depicts four facial images, each representing a different expression category defined in this study (A, open-mouth threat; B, fear-grimace; C, yawn; and D, neutral). Superimposed upon these images are contour plots derived from the data collected during the 3-sec presentation of each image. It is important to note that these contour plots reflect the total number of eye positions, not the data normalized for area of the feature in the image, which was used for statistical analyses.

DISCUSSION

The goal of this study was to address three questions related to the perception of facial displays by rhesus

monkeys. First, is there a difference between the manner in which monkeys view real-time dynamic versus static facial displays? Second, when given a choice to freely inspect two-dimensional images of conspecific and human faces, to what extent do monkeys visually explore the face and which facial features are most often inspected? Third, do facial displays differing in expressive content evoke different patterns of viewing? By presenting both real-time and static facial displays, we have shown that temporal changes of these facial features are not an important determinant in the evocation of stereotypical viewing patterns. Our study was designed to investigate the fixation patterns to facial images presented in a controlled laboratory setting and thus leaves open the question of what extent facial motion may yet affect the general state of orientation and arousal or the generation of a particular behavioral response.

In regard to which features were most often visually

Figure 5. Reproductions of four static video frames, each representing a different expressive grouping: (A) open-mouth threat, (B) fear-grimace, (C) yawn, (D) neutral. Contour plots superimposed on each image represent the total number of eye-position data points collected from four subjects over the course of 25 three-second static presentations. The location of the contour plots represents the horizontal and vertical eye-position data relative to the facial image, whereas the contour interval reflects the relative density of recorded eye-position measurements.

