Laughery, K. R., Duval, G., & Wogalter, M. S. (1986). Dynamics of face recall. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. W. Young (Eds.), *Aspects of Face Processing* (pp. 373-387). Dordrecht, Netherlands: Martinus Nijhoff.

DYNAMICS OF FACIAL RECALL

K.R. LAUGHERY, C.DUVAL, M.S.WOGALTER

1. INTRODUCTION

The purpose of this chapter is to examine some of the dynamic properties of facial memory, particularly facial recall. Most research to date on face memory has employed recognition as opposed to recall procedures. One reason for this emphasis is that while facial recognition is a frequent, everday activity, facial recall or description of someone is much less common. Hence, the study of facial recognition processes has ecological validity. A second reason is essentially methodological. How does one adequately assess facial memory through recall? Having witnesses provide drawings presents the problem of artistic skills. Having witnesses provide verbal descriptions presents the problem of describing a complex spatial configuration. In both cases, the difficulties at the response level precludes knowing whether the witness possesses more facial information than shown in the resulting external representation.

In one applied context, witnesses recalling faces of criminals, research has been carried out on various techniques for constructing hard-copy representations from the witness's memory. Primarily, this research has focussed on the sketch-artist, identikit and Photofit. These techniques involved the witness subject: (1) recalling information about the face from memory, (2) describing the face to the artist or operator, (3) responding to questions posed by the artist or operator about the face, and (4) reacting to the constructed image to suggest changes. This work has been discussed elsewhere (Laughery & Fowler, 1980; Ellis, Shepherd & Davies, 1975).

In this chapter we will examine some of the dynamic properties of facial memory in two of these construction tasks - the sketch artist and the identikit. We are concerned with the memorial acts/processes occurring over time. Putting it differently: What is going on inside the witness's head while performing facial recall tasks? Towards this end, several behavioural indices will be presented and inferences drawn about the recall process. More specifically, we ask questions such as: "What strategies do witness subjects use when they try to remember and when they try to recall a face? Are there individual differences in strategy usage?"

Most of the data to be reported came from a large study carried out on sketch artist and identikit techniques (see Laughery & Fowler, 1980). The task procedure involved pairs of witness subjects meeting with an experimenter and a white-male target for a small group discussion lasting approximately eight minutes. The recall task immediately followed. One witness subject met with a sketch artist and the other with an identikit technician to construct images of the target. None of the subjects had ever done this task before. Generally, one target was used per pair of witnesses, but because of subject no-shows some sessions were held with only one witness subject. Witness subjects were told of the subsequent memory task before seeing the target; however, expectancy/intentionality has not been found to affect recall measures (Ellis, Davies & Shepherd, 1978).

A variety of data were recorded: (1) Construction sessions were tape recorded, transcribed, and timed, (2) Adjectives describing 23 facial parts/features were sorted to comprise an adjective dictionary, (3) Post-session questionnaires were completed by witness subjects following the construction, and (4) Judged similarity scores were obtained as a measure of performance.

Five analyses were carried out and concerned: strategies witness subjects reported using to remember faces; time-line data describing some of the frequency and time properties of feature production; time-line data describing feature sequencing during production; post-session responses describing features that were easy/difficult to remember/describe and vocabulary used in describing faces.

2. STRATEGIES USED TO REMEMBER THE FACE

This section deals with the strategies used by witness subjects to remember the target face. On the post-session questionnaire, witness subjects wrote down the methods and strategies that they used during initial exposure to remember the target face. The specific question asked: When you viewed the target, what did you do to help you remember the face? Since witness subjects knew in advance they would be attempting to recall the face, it seemed reasonable to assume they would employ some strategy to remember the target. Responses to the question yielded verbal statements from 121 witness subjects. In an effort to find a way to make some sense of these statements the following analysis was carried out. Twenty-seven additional subjects performed an unconstrained sorting task (Bookman & Arabie, 1972) on the statements. Prior to the sorting task, these subjects looked at a face and then wrote down how they would try to remember it. This first step was to acquaint them with the task. In the unconstrained method of sorting, each sorter can place responses into any number of similarity piles. Using the sortings, a nominal cluster analysis (Bhat & Hampt, 1976, p.61) was employed. This analysis identified an organisation consisting of three strategy clusters: (1) comparison or association with a known person ($n = 18$), (2) feature analytic ($n = 75$), and (3) forming a mental picture, wholistic processing ($n = 28$). These clusters are illustrated by the statements shown in Table 1.

The issue of cluster reliability is important. To what extent are the witness's responses characterised by the clusters? An estimate of the overall reliability of the cluster structure would be helpful. A commonality score was developed that measured how much a given witness's response had in common with responses in each of the three clusters. Conceptually, commonality was based upon the extent to which a response was assigned to the same or other groupings in the sorting task. An analysis of these cluster commonality scores indicated that the above cluster structure accounted for 56% of the variance.

The issue of performance characteristics for persons using these strategy groupings was addressed by examining the quality of facial images produced. Another group of subjects provided ratings of goodness-of-fit (or similarity scores) between the images and photographs of the target from 102 of the 121 witnesses. The mean standardised similarity ratings as a function of strategy grouping is shown in Table 2.

An analysis of variance of the similarity scores for the three different clusters yielded a marginally significant result, $F(2,99) = 2.26$, $p < .07$. The fact that both clusters 1 and 3 involved visual imaging of the whole face prompted recalculating the analysis of variance with groups one and three combined. The result was significant though small, $F(1, 100) = 4.50$, $p < .05$. The means for this analysis are shown in Table 3. As can be seen

TABLE 1. Examples of statements describing strategies reportedly used to examine the target's face during study.

Cluster 1: Comparison or Association to Known Person

He looked very similar to a friend of mine - colouring, hair style, and general shape of face. I noted basic differences between my friend and the target - glasses, smooth skin, and no sideburns.
I tried to remember someone he reminded me of.
I compared his moustache and beard to a very close friend of mine and they were very similar. I tried not to forget my first and last impressions. I noticed how his face changed as he did different things - squint, smile, etc.
I tried to get a good fixed picture of him in my mind, and imagined certain features of his face with someone else I knew.

Cluster 2: Feature Analytic

I looked at each feature of the face and tried to remember the most noticeable thing about them.
I checked the length and width of the eyes, and their depth in the face; size of lips and position on face; and shape of the nose. Also, I checked wrinkles and identifying marks.
I tried to look at each of his features - his colouring, the shape of his face, his hair, his glasses, shape of his ears, and his upturned chin.
Looked at features of his face, how large was his nose, colour of eyes and hair, position of features.

Cluster 3: Forming Wholistic Mental Picture

I glanced at the target from the hair on his head to his chin. Then I shut my eyes and tried to visualise what he looked like. My second look at the target was another glance. By this time, I was able to close my eyes and see his face clearly.
I tried to visualise an image.
I looked at him, then got a visual picture in my mind, I kept this up for the entire time.
I looked at him, looked away, looked back; each several times to commit his face and stature to memory, and then to check the accuracy of the impression.

TABLE 2. Mean standardised similarity rating as a function of strategy grouping (lower score indicates higher similarity).

Cluster 1	Cluster 2	Cluster 3
.284	- .177	.276

TABLE 3. Mean standardised similarity rating for a two-strategy grouping (lower score means higher similarity).

Strategies 1 & 3 Wholistic/Gestalt	Strategy 2 Feature/Analytic
.280	- .177

in this table, as well as Table 2, recall performance was better for witnesses who reported using a feature analytic strategy than those reporting a wholistic or gestalt type of processing.

These two general groupings of witness subject responses, wholistic processing and feature analytic strategies, correspond to the views most commonly held to account for facial information processing (Bradshaw & Wallace, 1971; Nielsen & Smith, 1970). Indeed, Bartlett (1932) found two "natural groups" based on individual differences: (1) persons who relied mainly upon visual images, and (2) persons whose responses were predominantly determined by the use of language. We will return to this issue of wholistic versus feature processing in the discussion.

3. TIME-LINE DATA DESCRIBING FREQUENCY AND TIME PROPERTIES OF FEATURE PRODUCTION

Protocol analysis (Newell & Simon, 1972) involves a detailed accounting of the behaviour of an individual as the person works toward some specified goal. The purpose of such analysis is to provide information from which to infer the internal representations of the subject's knowledge as well as the processes employed by the person when performing the task. Both the content of the person's statements as well as the temporal properties of the process are examined.

A number of molar properties of the order and timing of feature descriptions during image generation were examined by a type or protocol analysis. The data was 64 witness subject protocols (34 sketches and 30 identikits) which consisted of tape recordings and transcripts. Twenty-five targets were described by one witness working with the identikit and a second witness working with the sketch artist. Nine of the remaining 14 targets were described only with the sketch, and the other five only with identikit. Summary statistics of this data are presented in Table 4.

The frequency and time measures in this table quantify the following: (1) how many different features were described (Number of Feature Codes - NFC), (2) how many times features were described (Number of Feature Stops - NFS), (3) a measure reflecting the amount of "moving around" or feature refinement during construction (Ratio of Feature Stops to Feature Codes - RFSFC), (4) the duration of the image generation session (Total Elapsed Time - TET), (5) the Time per Feature Code - TFC), and (6) the Time per Feature Stop - TFS). Several t-tests were carried out to determine if the technique differences were statistically significant. It should be noted from this table that witness subjects using the sketch technique reported information on more features, switched features more often, refined the features more often, and spent more time describing features than witness subjects using the identikit. On the other hand, identikit witness subjects spent more time per feature on any one occasion (a feature stop). No difference between the two techniques was found for TFC, the time spent per feature. The RFSFC measure is essentially a token to type ratio. It is an indicant of how much "moving around" between features occurred in the generation session. A witness subject with RFSFC equal to 1, the lower bound of this measure, means he or she did not refine the description of a feature once it had been given. Though the RFSFC measure for identikit appears close to one, statistically it is significantly different from 1.0 (p .001).

Correlations were computed between the frequency and time measures and the goodness-of-fit of the constructed images. None of the correlations approached statistical significance.

TABLE 4. Time-line measures for both production techniques

	Sketch	Identikit	t	p
Frequency Measures				
Number of Feature Codes (NFC)	13.32	8.10	8.08	.001
Number of Feature Stops (NFS)	30.41	12.33	8.13	.001
Ratio of Feature Stops to Feature Codes (RFSFC)	2.26	1.48	6.33	.001
Time Measures				
Total Elapsed Time in sec. (TET)	2069	1368	4.63	.001
Time Per Feature Code (TFC)	156.0	171.9	1.07	.05
Time Per Feature Stop (TFS)	72.5	118.0	4.55	.001

The amount of time witness subjects spent describing the five most frequently mentioned features (the set was the same for both techniques) are presented in Table 5. This ordering of features by elapsed time corresponds, in part, to the ordering of parts of the face that were most helpful to facial recognition as reported by Goldstein and Mackenberg (1966). They determined that the upper portions of the face (hair-hairline and eyes) were more helpful to identification than the lower portions (mouth and chin). Other research has shown that the upper face features: (1) are selected first during reconstruction using the Photofit (Ellis, Shepherd & Davies, 1975), (2) are better recognised (Goldstein & Mackenberg, 1966; Fisher & Cox, 1975), (3) lead to more description with the face present (Shepherd, Ellis & Davies, 1977) or absent (Ellis, Shepherd & Davies, 1980), (4) lead to faster judgements for same/different comparisons of facial changes (Matthews, 1978), and (5) produced fewer confusions (Davies, Ellis & Shepherd, 1977). These results indicate the greater importance or salience of the upper features for identifying a face. Of course, this ordering may in part also reflect the complexity and distinguishability of these features. We will elaborate on this point later.

Table 5. Mean elapsed time to describe the most discussed features. (Mean times between features beginning at the same margin were not significantly different.)

Features	Mean time (sec)
Eyes	338
Hair/Hairline	334
Nose	230
Chin	134
Mouth-Lips	116

It should be noted that to some extent the above measures are almost certainly task driven. The longer times for sketches are partially determined by the fact that artists usually require more time to draw a feature than identikit operators require to select it. But the difference also reflects that the witness subjects in the artist situation are thinking and describing the features in greater detail than witnesses in the more wholistic identikit situation.

4. PRODUCTION SEQUENCING USING TIME-LINE MEASURES

The time-line data was also used to try to understand the progression of the recall/production effort. The purpose was to describe some of the molar properties of production sequencing in feature description. A measure was developed to quantify order properties. Three production sequencing variables were derived to represent the relative amount of new feature activity during the First 30%, Middle 40% and Last 30% phases of the construction process. More specifically, these measures reflect the degree to which subjects describe a quantity of new features relative to a quantity of feature stops within three specific time periods for each construction. In other words, they represent the relative amounts of new feature activity during the early, middle, and late phases of the construction process. To some degree, like the RFSFC measure, they describe how much "moving around" (amount of new feature activity vs new feature refinement) that occurred in the description.

The measure is a score that represents the Ratio of the number of New Feature Codes to the total number of Feature Stops during the First 30%, Middle 40%, and Last 30% of each subject's total elapsed time (TET). These scores are calculated differently from the RFSFC measure that was described earlier. When work on a feature overlapped an interval boundary, which was most of the time, one was added to the number of new feature codes in the interval where the code description began. However, the feature stop values added to the time intervals were proportioned on the basis of the time spent on the feature in the two intervals. For example, if a new feature description took five seconds in the first 30% interval and 15 seconds in the middle 40% interval, .25 stops and .75 stops were added to the first and middle intervals respectively. The ratio of new feature code to feature stop scores produced values that ranged from near 0 to slightly above 1.

A high production sequencing score indicates more complete effort on new features (features not previously described). A low score indicates more going back to features already described - an indicant of the refinement process. Mean production sequencing scores for both techniques are shown in Table 6. The first 70% of the identikit generation time is spent on new features. On the other hand, only the first 30% of sketch generation time is spent on new features; the latter 70% is spent on refinement. An analysis of variance confirmed this Technique by Time period interaction, $F(2,124) = 4.09$, $p < .05$.

TABLE 6. Mean production sequencing scores for time intervals broken down by recall technique.

Time Interval	Sketch	Identikit
First 30%	.842	1.083
Middle 40%	.479	.998
Last 30%	.195	.577

The production sequencing scores were correlated with the goodness-of-fit ratings. Only one significant relationship was found. For the middle 40% time period and the identikit constructions, there was a significant relationship with performance (similarity ratings), $r = .66$, (43% of variance). Because a low similarity rating indicates better performance, this positive relationship shows that identikit witness subjects were more accurate if they had lower production sequencing scores (indicating more refinement) during the middle 40% interval. More broadly, this relationship suggests that those witnesses using the identikit, who make preliminary selections of features early in the process and then return to refine the selections, will do better than witnesses who try to complete work on each feature and do little refinement. It should be noted, however, that not many of the identikit witnesses followed this more effective strategy. No comparable relationships were found for users of the sketch technique.

A cluster analysis was performed on the production sequencing data in order to explore the feature production process further. The analysis identified groups of individuals who share similar production strategies as represented by time-line/order measures. Using the 64 witness protocols that were coded for time-line, three groups of subjects were identified. Descriptions of these cluster groups are shown in Table 7.

TABLE 7. Descriptions of production sequencing cluster groups.

Group 1 production strategy (10 SK, 4 1DK):
There was a preliminary selection/definition of most of the facial features to be described, followed by both refinement of these features and the addition of some new features.

Group 2 production strategy (20 SK, 6 1DK):
These witness subjects made no special attempt to make preliminary selections of features. They worked on a group of a few features at a time which would then be refined until the witness subject was satisfied that the best representation had been obtained. They then continued the same pattern with a new set of features. The feature groupings were generally composed of facial parts in close proximity.

Group 3 production strategy (3 SK, 21 1DK):
These witnesses worked on one feature at a time until they were satisfied and then moved on to the selection of another feature.

Table 8 shows the time-line measures as a function of the three production sequencing strategy clusters. In particular, note that subjects in cluster 2 took the greatest total time to produce the construction, produced the greatest number of features, features stops, and tended to "move around" most frequently. In regards to these data, note how different clusters 2 and 3 are. It is interesting that membership in cluster 2 was dominated by sketch users and cluster 3 was dominated by identikit users. However, these two production strategies did not appear to differentiate greatly on goodness-of-fit scores (similarity ratings) as shown on the bottom row of the table.

Cluster 1 consists of fewer subjects than clusters 2 or 3 and slightly more than two-thirds worked with the sketch technique. As indicated by the RFSFC scores, these subjects did less refinement than subjects in cluster 2 and more than those in cluster 3. Their similarity scores, however, were high (poorer performance) compared to the other groups.

What then are some of the dynamics of facial recall? In the identikit construction task, most subjects work on features one at a time. Getting to and working on any given feature takes on average a longer time than with the sketch technique. With sketch artists, the most common strategy is to do some preliminary work on a number of features and then refine them.

To what extent are the production strategies task driven? Although the sketch and identikit techniques were represented in all three strategy groups, the pervasiveness of task demands seems apparent. A significant majority of subjects in clusters 1 and 2 worked with sketch artists and with the identikit in cluster 3.

Task influence may come about as a result of general procedures or training and experience of the artist and operator. These factors could lead to a systematic ordering of feature selections which in this sense may be beyond the witness's control. While these possibilities certainly exist,

TABLE 8. Time-line measures for three production sequencing strategy clusters.

	Production Clusters		
Frequency Measures	1	2	3
Number Feature Codes (NFC)	11.7	13.5	7.6
Number Feature Stops (NFS)	22.7	33.7	11.5
Ratio of NFS to NFC* (RFSFC)	1.9	2.5	1.5
Time Measures			
Total Elapsed Time (TET)	1781	2187	1188
Time per Feature Code (TFC)	154	162	163
Time per Feature Stop (TFS)	78	70	108
Similarity	.345	- .419	- .361

* Calculated from means rather than individual scores

there were 13 subjects who worked with the technique that was not dominant in the three clusters. Hence, on the basis of these data alone there is subject control over the sequencing of work on features.

5. POST-SESSION RESPONSES TO QUESTIONS ASKING WHAT FEATURES WERE EASY DIFFICULT TO REMEMBER/DESCRIBE

As noted earlier, the post-session questionnaire was used to gain information about strategies the witness subjects used when studying the target. The post-session questionnaire was also used to gain information about which features witness subjects' thought were easy or difficult to remember or describe. The following questions were asked: (1) What parts of the face were easiest to remember?, (2) What parts of the face were difficult to remember?, (3) What parts of the face were easiest to describe?, and (4) What parts of the face were difficult to describe? Table 9 presents a summary of the results for the six most commonly mentioned features. Entries in the Table are the percent of total subjects who listed the feature in response to the particular question.

It is not our intent to make too much of these data, but they do contain some interesting observations. First, almost without exception, subjects reported that remembering faces was easier than describing them. This conclusion is based on two types of comparisons; the easy to remember vs easy to describe and the difficult to remember vs difficult to describe. The only exceptions occurred with the mouth/lips and eyebrows where more subjects listed them as more difficult to remember than difficult to describe. However, the percentages in these cells are close.

A second observation concerns the description of facial features. In the facial memory literature it has generally been concluded that people are not

TABLE 9. Percentage of witness subjects reporting features as easy or difficult to remember or describe collapsed across techniques.

	Easy to Remember	Easy to Describe	Difficult to Remember	Difficult to Describe
Eyes	43	38	35	45
Nose	36	26	37	43
Mouth/Lips	34	31	29	25
Eyebrows	26	20	14	12
Hair	49	46	8	26
Chin/Jaw	48	39	52	53

particularly good at describing faces (Shepherd, Davies & Ellis, 1978; cf. Davies, 1983). The results in Table 9 indicate that people's reports of ease or difficulty of description varies from feature to feature. More people find it difficult to describe the eyes, nose and chin/jaw features. However, more people report it easier to describe mouth/lips, eyebrows and hair. Indeed for hair the difference is almost twofold. Shepherd, Davies & Ellis (1981) in a review of the literature on cue saliency for faces point out that the most consistent finding across several methodological paradigms was that hair is the single most important feature. It is interesting to note in this regard that hair, more than the others, is probably the feature that we do describe verbally in our everyday activities. We talk about different hair styles or we describe to our barber or hair stylist how we want our hair to look. Hence, hair may be easier to describe because we are more practiced in such descriptions. This point is further addressed in the next section.

6. VOCABULARY IN DESCRIBING FACES

In our everyday memory, faces are not usually verbally described. On occasion, we may attempt to describe what someone looks like and we find we are at a loss for words. The sketch and identikit both require the recall of features, and this recall is usually accompanied by verbal description. In this section, some aspects of verbal description of faces are examined.

The interaction of the witness subject with the sketch artist produces more verbal description than with the identikit technician ($p < .01$). At the start, at least, sketch production is primarily a recall/generation task. There are no exemplars from which to choose as is the case with the identikit. Adjective words must be generated to describe the target. As the artist completes more and more of the sketch, the task increasingly becomes more recognition-like. On the other hand, the identikit produces an initial "lay-down" and becomes more of a recognition task early in the procedure. Hence, less verbal description is needed to communicate change with the identikit.

A correlation between the number of words produced by the witness subject during construction revealed no statistically significant relationship between number of words and goodness-of-fit scores ($p > .10$). This particular result may be due to the relatively large subject variance in this measure. This point can be illustrated by the following scenerio:

Some witness subjects who "seem" to have a good memory for a feature will give a description that is brief and concise, while others with good memories will attempt to elaborate at length. A similar point can be made about witness subjects who "seem" to have a poor memory: Some may not waste time (and words), and others may talk about features as if they are hoping some memory will emerge.

Though no relationship was found between the total number adjectives and quality of image, we were curious about the construction time differences for individual features shown in Table 5. Are there differences in the number of adjectives used to describe the different features? Table 10 shows the proportion of the total number of adjective descriptors during the construction phase that was used for several features. The table also presents results from a study by Ellis, Shepherd & Davies (1980) using the Photofit. Despite differences in the categorisation of features between studies (e.g. the features mouth, lips and face shape were categorised differently), note the similarity in the results. Note also that there are considerably more adjectives used to describe hair. This result may be related to the earlier point about practice in describing hair and may further indicate a richer vocabulary for this feature.

One of the findings that has shown up repeatedly in research on facial memory concerns a sort of top-down ordering of feature importance. Ellis, Shepherd and Davies (1975) found that subjects constructing the Photofit preferred to select the features in the order: hair-forehead, eyes, nose, mouth and chin. Ellis, Davies & Shepherd (1977) noted that subjects took much more time selecting hair and eyes than other features. Upper features have been shown to be better recognised than lower features (Fisher & Cox, 1975). In addition, Matthews (1978) subjects make same/different judgements on simultaneously presented pairs of identikit faces with different combinations of features. Reaction times indicated that changes to the hair, eye, and chin region were detected the fastest, while eyebrows, and nose and mouth changes were slowest. Thus, in general, the upper features of the face seem more important to memory, which is consistent with the results reported here for more processing time (Table 5) and verbal descriptors (Table 10).

We also examined those adjectives used most often (50 times or more) by witness subjects across all features. Table 11 shows the frequency of

TABLE 10. Proportion of descriptors allocated to various facial features (adapted from Ellis, 1984).

Ellis, Shepherd & Davies, 1980		Present Research	
Feature	Proportion	Feature	Proportion
Hair	.27	Hair	.23
Eyes	.14	Eye-Eyelashes	.15
Nose	.14	Nose	.13
Face structure	.13	Eyebrows	.09
Eyebrows	.08	Mouth-Lips	.09
Chin	.07	Chin	.07
Lips	.06	Face shape	.05
Mouth	.03	Facial hair	.04

TABLE 11. Frequency of adjectives used in the description of all faces along with root word count per million (A = 50–100 per million, AA = 100 per million or over) from Thorndike & Lorge (1944).

Frequency	Adjectives	Thorndike-Lorge Count
735	long, longer	AA
639	straight, straighter	AA
556	thin, thinner	AA
540	wide, wider	AA
405	round, rounded, rounder	AA
351	thick, thicker	A
272	short, shorter	AA
260	full, fuller	AA
251	small, smaller	AA
248	dark, darker	AA
245	curly	14
232	curve, curved	36
222	part, parted	AA
216	large, larger	AA
201	prominent	25
196	down	AA
191	big, bigger	AA
174	square, squared	AA
166	narrow, narrower	AA
150	smooth	A
156	turn-	A
145	light, lighter	AA
141	point, pointed	49
134	high, higher	AA
131	bushy	4
121	wave, wavy, wavier	2
117	cover-	AA
114	slant-	AA
108	move-	AA
107	oval	5
105	heavy, heavier	AA
102	flat, flatten, flatter	A
99	deep-	AA
95	angle, angular	33
78	brown	AA
76	average	A
76	broad, broader	A
75	stick-	AA
70	low, lower	AA
69	arch, arched	38
63	blocky, block-	A
63	crease, creases, creasing	3
61	normal	41
60	fat, fatter	AA
53	medium	23

adjectives used in the description of all faces along with the root word language count from Thorndike & Lorge (1944). There are two things to note about this list. Most of the adjectives were size and shape measurement terms (e.g. long, straight, thin, wide, round, and short). Also, most were very high frequency terms in the language. In general, they are feature independent; that is, these adjectives are general-use, relative-size descriptors that could be used to describe most geometric shapes, such as parts of buildings. In an imaginary word association test, it seems unlikely that a facial feature would be given to most of them. This generalisation must be qualified because there are a few terms that appear to be frequently associated with hair (e.g. curly, wavy, and bushy). A few of the adjectives might be regarded as prototypical words - average, normal, and medium. Are the features for which these words are used actually prototypical features or are subjects just "filling in the blanks" in their memory? Probably there is some of each. The important point from these data is that there does not appear to be much linguistic richness in our vocabulary to describe faces.

7. SUMMARY AND CONCLUSIONS

In this paper we have explored some of the properties of facial memory in the context of a memory recall task. The task consisted of generating from memory a hard-copy representation of a face using two techniques employed in law enforcement - the sketch artist and the identikit. In addition to the constructed facial images, the results included tape recordingsa of each session and a post session questionnaire. Time-line and vocabulary data were compiled from the tape recordings. The constructed images were assessed by having a separate group of subjects rate similarity between the images and photographs of the actual faces.

An analysis of acquisition strategies was carried out by examining subjects' responses to a question in the post-session questionnaire. The question asked what they did during exposure to the target person to remember his face. A cluster analysis revealed two general strategies which we have characterised as wholistic and feature analytic. These results are consistent with previous findings that have identified similar strategies for processing facial information. The number of subjects in these two categories, 46 in the wholistic and 75 in the feature analytic, indicated a sizeable majority who processed the facial information at the level of features. This result does raise a question about some earlier conclusions of Woodhead, Baddeley & Simmonds (1979) that most subjects process faces at a wholistic level. On the other hand, Woodhead et al. based their conclusions on the failure of a feature-oriented training procedure to improve performance on a facial recognition task which may be better served by a wholistic strategy. It should be noted that subjects in the current study were told in advance of seeing the target person that they would be constructing a facial image; a procedure that may have induced many to examine the face at a feature level. This point is supported by the fact that constructions by the feature analytic group were rated better than those of the wholistic group. Hence, it seems appropriate to conclude that people may process facial information either wholistically or at a feature level, depending on the purpose for which the information will be used. The observation that faces are normally processed wholistically may be due to the fact that facial information is typically used for recognition purposes.

Protocal analyses using time-line and feature description data revealed a number of interesting findings. In addition to sketch and identikit differences, there were significant differences in the amount of time spent on different features. Features at the top of the face, hair and eyes, received the greatest description time, with a decreasing amount of time for features lower in the face. As noted earlier, this top-down emphasis is consistent with previous findings showing greater importance or salience for the top features.

A cluster analysis of the time-line feature data revealed differences in strategies employed by subjects in the construction task. These strategies differed in terms of how much "moving around" or refinement was involved. One group of subjects was characterised by a procedure in which some preliminary work was done on each feature early in the construction, and the remaining time was used for refinement. At the opposite end of the refinement continuum was a group who completed work on each feature before moving on to the next. In between these groups was a number of subjects who worked on several features during a time period, and after completing work on them moved to another set of features. In general, the strategy of working on one feature at a time was associated with the identikit technique, while the two strategies involving refinement were associated with the sketch. These results would indicate the construction strategies may be task driven. However, there were many exceptions which indicate the strategies are at least in part task independent.

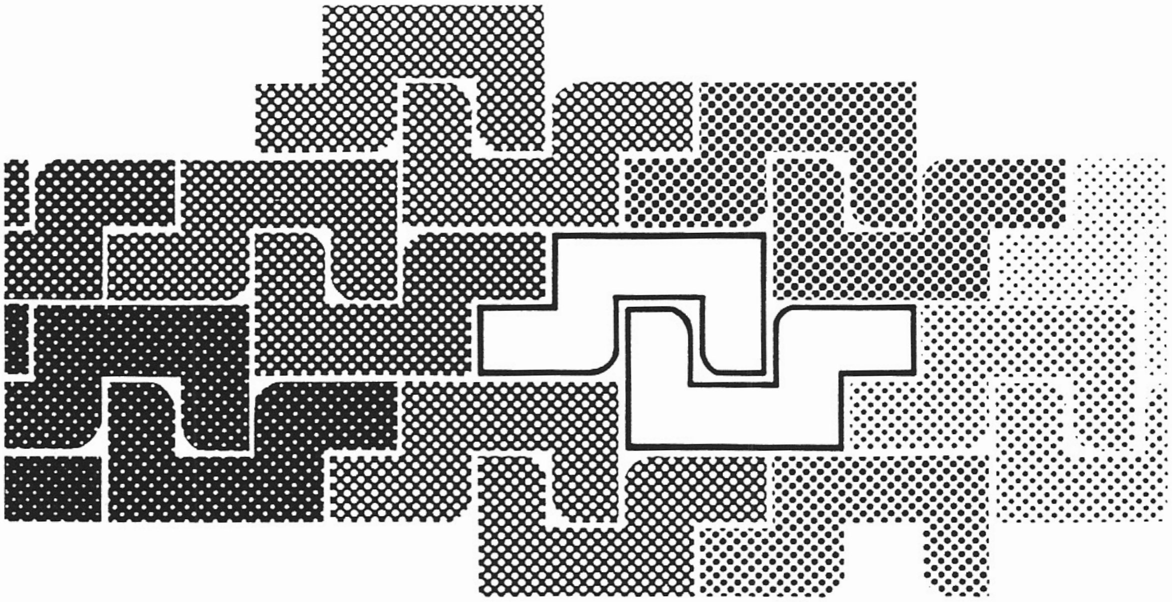
Differences in feature processing were also revealed in the answers to other questions in the post-session questionnaire. Subjects were asked to note which features were easy or difficult to remember or describe. Not surprisingly, most features are easier to remember than to describe (or more difficult to describe than to remember). This finding is consistent with other work that has indicated people having difficulty verbally describing faces. This difficulty may be due in part to lack of practice, since most facial memory activity involves recognition. An exception in the current findings was hair, which was frequently listed as easy to describe. Since hair is a feature that we probably do describe more than others, this result is consistent with the practice notion mentioned above.

The vocabulary used to describe facial features also revealed some interesting results. The number of descriptors used for different features reflected the top-down importance or salience of feature and was consistent with earlier results reported by Ellis, Shepherd & Davies (1980). Perhaps even more interesting, however, are the vocabulary items used to describe the facial features. Most of the adjectives are high frequency words that describe size and shape. Moreover, they are words that are not specifically associated with facial features; but instead might be used to describe any structure. A few words such as curly and wavy might be associated with facial features, but these terms are exceptions. This outcome also helps to explain the difficulty people have in describing faces. In addition to the lack of practice, our vocabulary is not well suited to the description.

One of the issues that cuts across the above analyses and findings is the extent to which strategies and processes employed by subjects are task driven or subject determined. Several of the findings are unbdoubtedly influenced by the memory construction task. As already noted, strategies employed by subjects to remember the target person were probably influenced by the fact that they were informed that they would subsequently be doing the construction task. Similarly, the greater time spent on the hair and eyes as well as the number of vocabulary items used in describing these features may be because the features are more complex and take longer to construct - especially in the sketch technique. A third example of task determined strategies is the greater "moving around" or refinement that occurs in the sketch technique. This result is probably influenced by the sketch artists who usually try to get some preliminary outline information about the different features before filling in details.

These task driven strategies and processes to some extent reflect adaptive characteristics of memory. However, not all of the subjects' behaviour is determined by the task. Vocabulary items used to describe the features are more a function of the language experience of the subjects. Also, there were individual differences in the strategies employed. Thirty-eight percent of the subjects used a wholistic processing strategy during initial exposure to try to remember the face. The time-line data indicated that strategies during construction varied between subjects within each of the construction techniques. Hence, the dynamics of or processes employed in facial memory are both a function of the task to be performed and the experience and capabilities of the person.

In the opening paragraph of this paper the problem of response competence was raised as a methodological problem in examining facial recall. The goodness-of-fit of the recalled face (the constructed image) cannot be regarded as an accurate reflection of the subject's memory. Other significant sources of error or variance are important. These sources include the limitation of the identikit foils to represent features, the skills of the artist or operator, and the verbal skills of the witness subject. Each of these factors limit how good the representation can be. Virtually everyone who has attempted the construction task notes that his/her memory is better than the constructed image. But these observations should not preclude the use of recall procedures to study memory for faces. Protocol analyses such as those presented here provide information from which memory processes and strategies can be inferred. In the present study, these analyses have identified task driven as well as subject determined strategies.



Aspects of Face Processing

Edited by

Hadyn D. Ellis, Malcolm A. Jeeves,
Freda Newcombe and Andy Young

NATO ASI Series

Series D: Behavioural and Social Sciences - No. 28

Aspects of Face Processing

edited by:

Hadyn D. Ellis

Department of Psychology
University of Aberdeen
Old Aberdeen
Scotland, U.K.

Malcolm A. Jeeves FRSE

Psychological Laboratory
University of St. Andrews
St. Andrews, Fife
Scotland, U.K.

Freda Newcombe

Neuropsychology Unit
The Radcliffe Infirmary
Oxford
U.K.

Andy Young

Department of Psychology
University of Lancaster
Bailrigg, Lancaster
U.K.

1986 **Martinus Nijhoff Publishers**
Dordrecht / Boston / Lancaster
Published in cooperation with NATO Scientific Affairs Division