Models for the processing and identification of faces

JOHN L. BRADSHAW* and GRAEME WALLACE†
Monash University, Clayton, Victoria 3168, Australia

Representations of human faces were constructed from Identi-Kit in a comparison of several possible models of information processing relating to facial recognition. There was no experimental evidence that faces were treated as unitary Gestalten, or that the component features were processed to any significant extent in parallel, even when the stimuli were presented as photographic positives in the normal upright mode. Instead, the best theoretical model to fit the data involved serial self-terminating processing, without replacement. Task difficulty, determining both processing time and number of errors made, was found to be a function both of the number of critical features present and the orientation (upright or inverted). Photographic negatives were handled in every way the same as positives. The more difficult the task, the stronger the evidence for a serial model and the greater the effects of practice. The latter suggests that, given time, Ss could handle inverted presentations as readily as upright ones, and that inversion may merely increase the difficulty in separating out the individual features for subsequent processing.

The human face can be viewed as a complex visual pattern subserving two purposes, communication and identification. Its complexity results from the number of features and their variations, such that extreme facial similarity is rarely encountered, despite the number of different individuals readily identifiable. That a learning process operates in the detection of critical identificatory features is perhaps evident from the difficulty which a European may initially have in distinguishing Asians and vice versa. On the other hand, the apparent speed with which a familiar face is identified suggests that it may act as a perceptual Gestalt or as a unitary "chunk" of information (Miller, 1956). In this way, an entire word may be responded to as fast as an individual letter is normally handled, as long as the word is a familiar one. It can be asked, therefore, whether this Gestalt or "chunk" quality of a face, whereby identification would perhaps occur through parallel rather than serial processing of the component features, operates only when the face is viewed in the normal mode, e.g., upright, or as a positive photographic image. If it is presented in an unfamiliar fashion, slower identification might result from recourse to a serial mode of operation. Yin (1969) noted that upside-down faces were difficult to recognize, and commented that Ss commonly adopt one of two strategies, either searching for a distinguishing feature or attempting to get an overall impression. These may perhaps be taken to reflect,

respectively, serial and parallel approaches in processing.

The question of how the facial features are identified and processed may seem to imply that their nature and number alone determine the speed and success of facial recognition. There will, however, be "feature" detectors, in a pattern-recognition sense, which process and encode the higher-order facial features. If so, conclusions concerning processing operations based on performance measures may relate to one level and not to the other. Feature detectors (pattern recognition) may operate in parallel to decode a particular facial feature, while the various members of the latter class are dealt with successively. However, until the nature of primary detectors is determined (Treisman, 1969), this objection can be levied against almost all studies of perceptual encoding processes which put forward models based upon observed performance differences, when the quantity and quality of various arbitrarily defined items in an array is varied. Indeed, in the context of identifying complex visual stimuli presented in an unfamiliar mode or orientation, one may conclude with Kolers and his associates (Kolers, 1968; Kolers & Perkins, 1969a, b) that there seems to be a higher-order mechanism of perceptual processing. This would be concerned with the recognition of geometrically transformed material, such as reversed or inverted script, and would function over and above the feature detectors operating at the level of specific orientations. This, of course, has important implications for the older classical theories of pattern recognition, such as template matching. Similar conclusions can be derived from the pioneering work of Rock (for review, see Rock, 1966).

The present study employed realistic facial representations constructed from Identi-Kit.[1] As each face is built up by superimposing seven or more overlays, corresponding to the major features (eyes, eyebrows, hair, mouth, chin, etc.), and a wide range of variations are available in each feature, an enormous number of different faces can be readily constructed, photographed, and appropriately displayed. A list-search paradigm was employed, whereby S scans down a column of list items for a prior-designated target. This is analogous to same-different reaction-time (RT) studies, where one or more storage items are previously presented, followed by a test item for comparison (Egeth, 1966; Sternberg, 1966, 1967). RT to positive or negative ("same" or "different") probes can give some information about how the comparison proceeds. This paradigm of successive presentations suffices for simple or overlearned stimuli, but with complex novel faces it was felt that too heavy a memory load would be imposed. Furthermore, the adoption of simultaneous comparison in list search was felt to approximate more closely the situations of identification in everyday life. Lists were constructed which contained paired items arranged vertically in a column, Ss' task being to detect targets where both members of the pair were identical, the other pairs containing dissimilar faces, or vice versa.

Specifically, it was asked which of three major models best accounted for performance in processing faces where they varied in the number of critical features present, and were shown either in the familiar mode of an upright positive photograph, or were inverted, or were photographic negatives, or both. The three models considered were parallel processing, serial self-terminating, and serial-exhaustive. The second model predicts that processing time will tend towards a linear function of the number of features to be encoded, one after the other, before a decision can be reached; at this point, processing stops. The other two would predict less of an association between performance time and the number of critical features present. With serial-exhaustive search, all features are encoded irrespective of whether a decision can be reached beforehand, and with parallel processing, all features are processed more or less simultaneously. RT could possibly increase under parallel search, as Sternberg (1966) pointed out, if the comparison times were independent, not equal, and the response was initiated only after termination of the slowest. It is also conceivable that with more features to process in parallel, the time required to
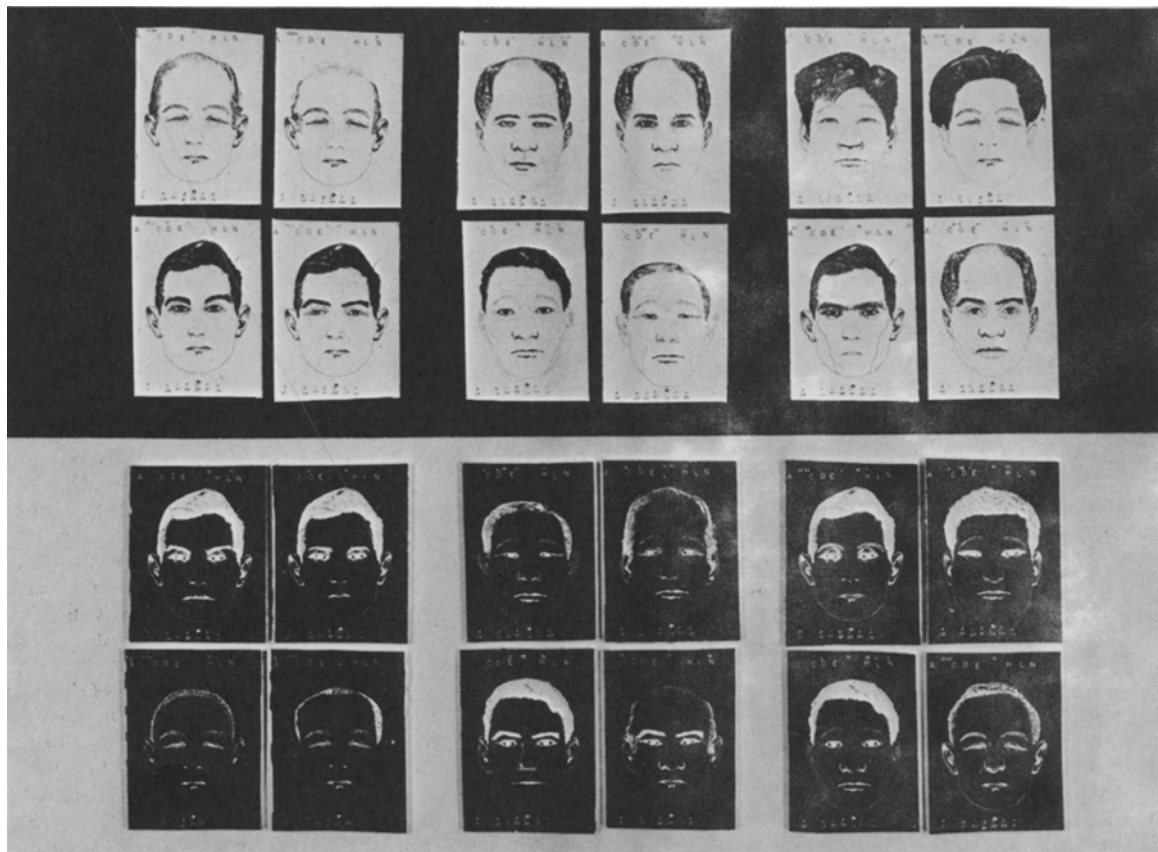
Fig. 1. Two examples each of paired list items with two, four, and seven features differing, in positive and negative mode. Faces constructed from Identi-Kit.

complete each feature increases proportionately (Bracey, 1969). However, these two variants of a parallel-processing model should perhaps be considered more as theoretical possibilities than as likely candidates.

Obviously, serial-exhaustive search is the least economical process. It might be expected that if the faces are in the normal positive upright mode, processing of these familiar Gestalten would approximate more closely to a parallel mode, while in the unfamiliar mode (negatives, inverted, or both), processing might revert to a serial form.

The effects of practice upon processing times for material presented in the familiar and in the unfamiliar modes were also examined. It was predicted that any initial superiority of material presented in the familiar mode would be less subject to improvement under practice. For this part of the study, a slightly different task was employed, searching for targets which differed in a matrix of identical pairs, instead of for targets which were identical pairs, the remaining list items differing.

METHOD

Faces were made up from Identi-Kit, a device used extensively in criminology for identification by the construction of facial likenesses. This kit, hired for the purpose, consists of a number of transparent overlays (approximately 7 x 5 in.), each displaying a single feature. Of the latter, seven were chosen, viz, nose, mouth, eyes, eyebrows, hair, chin shape, and age lines. Particularly obtrusive features, such as hats, beards, moustaches, and spectacles, were not employed. In this way, faces were constructed, photographed, and reduced to a 1 x 1½ in. format, in both positive and photographically negative versions. The major part of the experiment required the presentation of list-item material, consisting of paired faces differing from each other with respect to two, four, or all seven features. Faces, both in positive and negative form, were therefore constructed, 240 in all, there being 40 pairs differing as to each of the above three possible levels. The differing (critical) features were randomly varied. The paired items were then pasted in columns on cards, measuring

16 x 15 in., in five columns each of nine pairs. There was a separate card for positive and negative versions, and for each of the three levels of differing or critical features. The 45 paired items consisted of the 40 list items (faces different) and 5 target items (a pair of identical faces). The latter were pseudorandomly embedded in the list material, on average one to each of the five columns and never more than two. Positive and negative versions differed otherwise only as to the location of targets. The second part of the study, examining the effects of practice, required similar material. Here, however, list items consisted of identical pairs (40). The 5 target items were pairs whose members differed from each other in all seven features. Construction of lists, location of targets, and positive and negative versions were the same as before. Preliminary practice sheets to familiarize S with the task were also constructed, one for each type of material and mode of presentation (positive and negative, upright and inverted). They contained 5 list items and 1 target. For presentation under the

Fig. 3. List items different, targets same: Total number of targets and list items incorrectly identified as targets for the three types of material under the four modes of presentation.
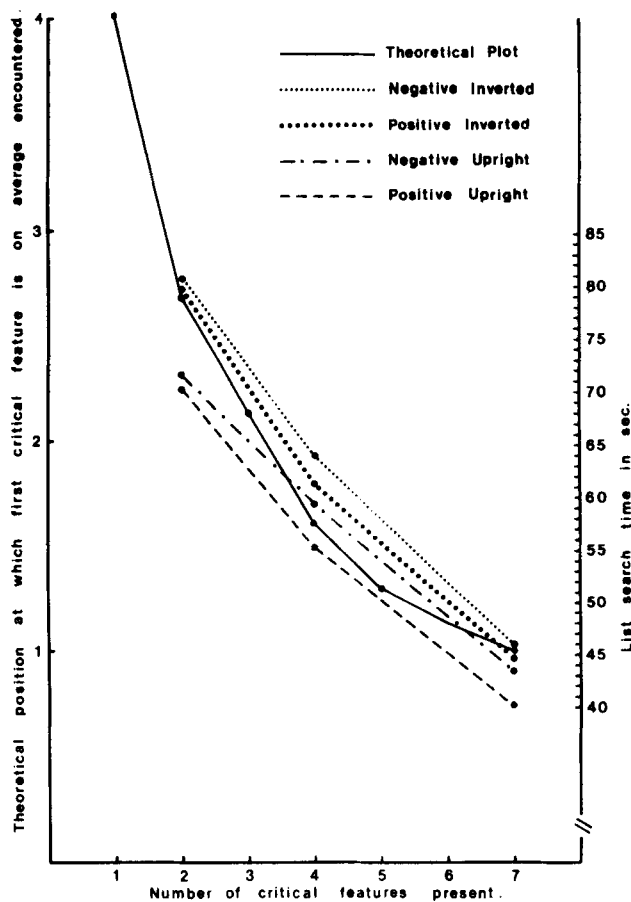
Fig. 2. Obtained times for the four modes of presentation, with three levels of critical features present, graphed onto a plot of the theoretical position at which the first critical feature is on average encountered. The latter assumes sampling without replacement.

inverted mode, the cards were merely turned upside down. Ss scanned the cards while seated, the cards being positioned vertically in front of them at a distance of 18 in. Under these conditions, the code numbering present (measuring .5 mm in vertical dimensions) and situated immediately below each face, was completely illegible. It would not have been difficult to obliterate these code numbers before the experiment. However, Ss were unaware of their presence or purpose. Under the conditions employed, they were hardly even visible and, in any case, a strategy utilizing these identificatory feature codes would have been so slow as to be immediately apparent.

Material was presented in a varying counterbalanced form, repeating on the seventh S. In each case, the material for the study on practice, which contained list items same, was presented twice, once at the very beginning and once at the end, with the other material in its various orders sandwiched between. In this way, a

within-Ss design was employed. Instructions were read to Ss, emphasizing speed, and stressing that if a mistake was ever made they were to continue straight on and not return to correct it. The 18 Ss were paid volunteers either from the psychology department or from a local Teacher's College, 8 males and 10 females. Before each type of material, list, and mode of presentation, S examined the relevant prior-practice lists. In each experimental series S was required to categorize verbally each pair successively, as rapidly as possible while scanning down the list columns, by saying "same" or "diff." He was practiced beforehand in the use of the latter abbreviation, in an attempt to standardize articulation times. Ss' verbal responses were tape-recorded during the experiment, as well as being timed by E with a stopwatch. In this way, a check was made on scanning times, with the possibility of correcting for any unforeseen event occurring during a run. In fact, this proved unnecessary. It was also possible to analyze subsequently, and at
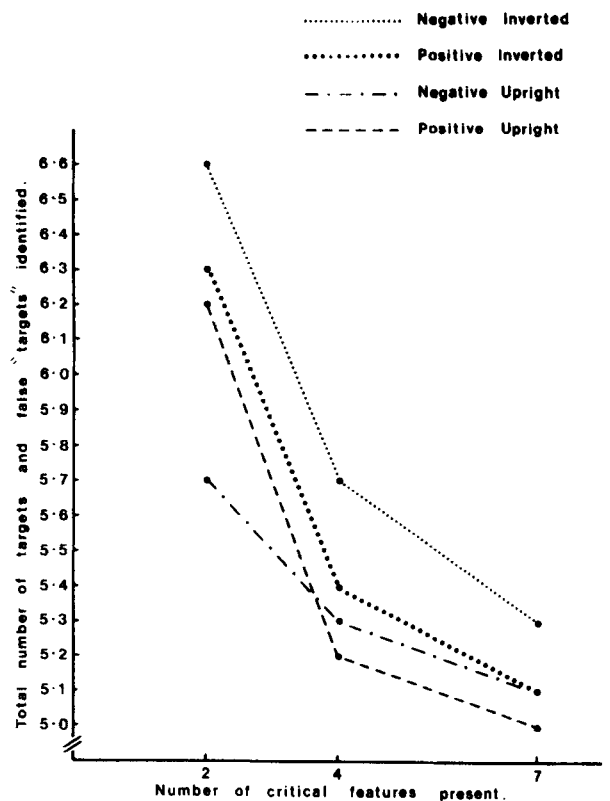
leisure, the number and location of targets, correct and false. Ss were told beforehand that targets in each list varied in number from about four to six, this deliberately being left vague, although in fact five were always present.

RESULTS

Scanning times for the various lists reflect the speed of processing for each type of material. Certain events, such as changes in fixation, articulation, location of target, etc., merely add a constant to the y-intercept. Figure 2 plots the average results obtained for the material, with list items different, targets same, under the various experimental conditions. The data were subjected to a four-way analysis of variance, and the following was found. The number of critical features present proved highly significant $(F = 130, df = 2,34, p < .001)$. The orientation of presentation (upright or inverted) proved no less significant $(F = 45, df = 1,17, p < .001)$, as did their interaction, Number of Features by Orientation $(F = 9.1, df = 2,34, p < .001)$. No other main effects or interactions approached significance. It seems, therefore, that the photographic negatives were processed as readily as the positives. The significant interaction
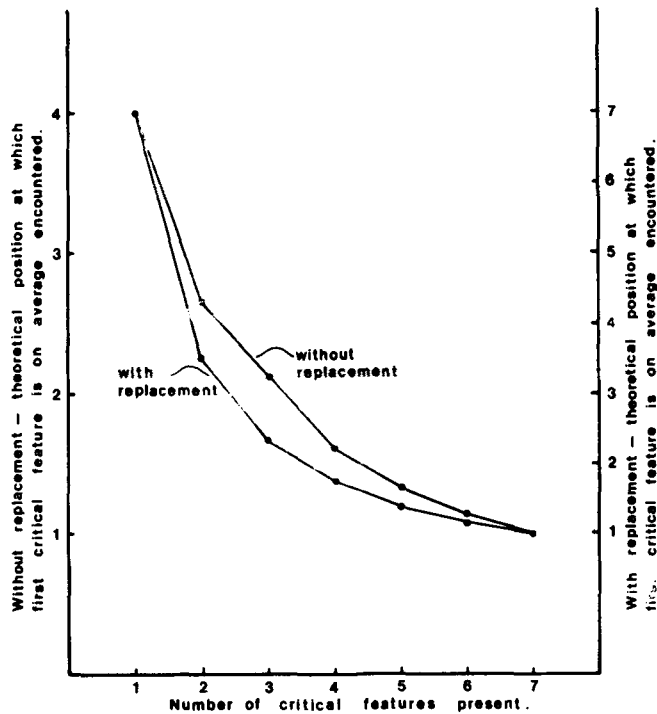
Fig. 4. A comparison between two serial self-terminating models, with and without replacement. Two curves are superimposed. They represent the theoretical position at which the first critical feature is on average encountered, for the various quantitites of such critical features.
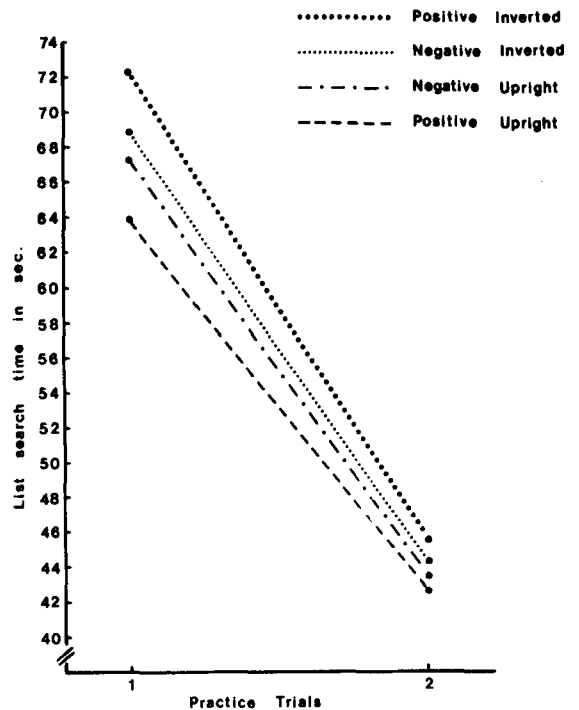

Fig. 5. Graph of practice effects for the study using list items same, targets different, under the four modes of presentation.

(Features by Orientation) implies a greater average slope for the more difficult (inverted) material, and consequently perhaps slower, feature-by-feature identification in processing. Conversely the flatter slope of the easier (upright) presentations could suggest a trend towards some degree of parallel processing. The implications of this and possible variant forms of a serial self-terminating model will be discussed later.

Figure 3 presents the overall number of targets named for the various conditions of list items different, targets same. Only rarely were targets ever missed, too few to justify a meaningful comparison. The scores instead reflect the extent to which Ss falsely identified ordinary list items as targets. The target data for the study concerned with the effects of practice, where the list items were identical pairs, were also discarded since all differences were quite minimal. When the data from Fig. 3 were subjected to a four-way analysis of variance, it was found that only two main effects and no interactions reached significance. Both were highly significant at p < .001 (number of critical features—F = 36, df = 2,34; and orientation of presentation—F = 20, df = 1,17). Their interaction just failed to reach the .05 level (F = 3.02, df = 2,34). The "detection" of more (false) targets is implied with the

more difficult material (either fewer critical features present or inversion or both). Since only five targets were in fact present, any "detections" exceeding that figure were inevitably erroneous. It can be seen from Fig. 3 that false targets ranged from near 0 to 1.6 for the various conditions. There is, then, once again a suggestion that the photographic negatives were treated in a similar fashion to the positives. It could perhaps be argued that because the material with the more critical or discriminating features is both processed faster and has fewer identified targets, there is therefore a casual link, the act of locating and identifying the targets influencing the obtained search rates. However, the maximum overall difference between any pair of target curves is only 1.6 targets, and within any one curve, 1.3. These values correspond to differences in search times of 41 and 35 sec, far too large an interval to account for this. Consequently it is suggested that a decrease in the number of critical features merely leads both to an increase in search time and a decrease in accuracy. Similarly, the less familiar the mode of presentation, the more the errors and the slower the processing.

Figure 4 compares processing speeds for material with list items same and targets different under two levels of practice,

when performed immediately before and repeated immediately after the other main task. When subjected to a four-way analysis of variance, it was found that Trial 2 was significantly faster than Trial 1, as expected (F = 69, df = 1,17, p < .001). As in both previous analyses, orientation reached significance (F = 16, df = 1,17, p < .01), and their interaction (Trial by Orientation) also proved just significant (F = 5.8, df = 1,17, p < .05). As before, no differences were found in processing times between the photographic negatives and positives. No other interactions approached significance. The significant interaction (Trial by Orientation) suggests that practice is of greater help to the more difficult material (inverted) and less to the more familiar, upright presentations, where presumably S cannot improve to the same extent. The difference in demand characteristics between the two types of task (search for targets where the pairs are different and where they are the same) makes it difficult to compare them directly. It is proposed to discuss this in more detail in the subsequent section.

DISCUSSION

Ss frequently commented upon the similarity of the Identi-Kit representations to typical human faces. This, in conjunction with its flexibility, suggests

that it might become a useful tool for a number of cognitive, perceptual, and information-processing studies.

The findings so far noted may be summarized as follows. There was no evidence that the faces were treated as unitary "Gestalten," or processed to any large extent in a parallel mode, though this might have been determined by the demand characteristics of the task. It might be tempting to ascribe the decrease in search rate with fewer critical features present merely to more frequent fixations. Certainly the latter were not controlled and, in the list-search paradigm, not directly controllable, though in principle perhaps detectable. However, with 45 paired items on the list and a search rate varying from 40 to 80 sec per list, depending on the type of material, there would be insufficient time for much more than one fixation per face. Irrespective of this problem, some type of serial model seems more appropriate, particularly with the less familiar orientations. Curiously, negative images were processed as readily as the positives. Orientation had a powerful effect, inverted presentations taking longer to process and resulting in slightly but significantly steeper functions. They were also more subject to practice, suggesting that, given time, one could handle inverted faces as readily as upright ones (cf. the findings on mirror script by Kolers, 1968, and Kolers & Perkins, 1969a, b). More precisely, perhaps, practice with inverted material might reduce the time required to locate and discriminate the unfamiliar individual features, one from another, for subsequent processing and comparison. The more difficult the task, i.e., the fewer the critical features present, or when inverted, or both, the higher the false alarm rate. This was felt to reflect, but not to determine, the search rates, since both were caused by the degree of task difficulty. In consequence, there was no evidence of a speed-accuracy tradeoff.

According to a serial self-terminating model, a decision can be reached and processing terminated as soon as the first distinguishing or critical feature is encountered. On the assumption, to be discussed shortly, of sampling without replacement, the average probability of this occurring at each sampling position can be calculated when one, two, or more critical features are present. If there is a total of N features, n noncritical and $N - n$ critical, the probability of encountering a critical feature at the first sampling position is given by

$$p_1 = \frac{N - n}{N}$$

and at the second by

$$p_2 = \frac{n}{N} \cdot \frac{N - n}{N - 1}$$

and at the third by

$$p_3 = \frac{n}{N} \cdot \frac{n - 1}{N - 1} \cdot \frac{N - n}{N - 2}$$

and at the fourth by

$$p_4 = \frac{n}{N} \cdot \frac{n - 1}{N - 1} \cdot \frac{n - 2}{N - 2} \cdot \frac{N - n}{N - 3}$$

and so on until probability reaches zero. The mean position (C) of such an encounter for any number of critical features present will then be given by the expression

$$C = 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3 + \ldots$$

This is plotted in Fig. 2, where the scales have been equated to accommodate the theoretical and empirical functions. The relationship between the empirical processing times and the theoretical position at which the first critical feature is an average encountered is a simple one, if it is assumed that the latter function can also represent the average time taken to process the requisite number of features. Since the critical features for each stimulus pair of faces are chosen randomly, it does not matter whether processing times for the individual features differ, or whether the direction of search is random or systematic. Taking into account the fact that the theoretical plot is based on seven points, and each set of empirical data on three alone, the fit appears fairly close, the inflection in the empirical data being slightly less than that of the theoretical.

One can, however, calculate the theoretical plot on the assumption of replacement, where S can be thought of as having no effective memory for the previously tested features, so that each may be sampled several times in succession (cf. Restle, 1969). If c is taken to represent the proportion of critical features present (e.g., 2/7, 4/7, etc.), the mean number of noncritical features encountered, n, is given by the formula $n = (1 - c)/c$. Consequently the average position, C, at which the first critical feature is encountered is n + 1, or $C = 1 + (1 - c)/c$. Both functions, with and without replacements, are geometric distributions and can be directly compared in Fig. 5, where the scales have been appropriately adjusted. It would seem that sampling with replacement gives a worse fit to the empirical data, with the moderate inflection of the latter. Indeed, it is not

unlikely that, given serial processing, S would remember which features he had compared and discarded. A second possibility is that S's sampling is redundant, seeking confirmation in a second, third, or further critical feature before making a decision. However, if this is so, it could prove quite arbitrary where the final limit is placed. Moreover, putting the criterion arbitrarily at two generates an irregular function quite different from the empirical data, and from either of the two theoretical functions so far considered.

It was noted earlier that the demand characteristics of the two tasks, as defined by the nature of the targets to be located, might differ. Certainly in everyday life we are commonly required to locate a familiar face from a crowd of unfamiliar ones. We do not normally look for a different face from among identical ones. Moreover, in the study on practice effects, the list-item pairs possessed all identical features, and the target-item pairs differed in all seven. The optimal strategy would have compared a single feature. However, even after practice, scanning times proved slower than in the other task when the list items had all seven features differing. Had the former task contained targets with fewer features differing, even longer times might have been expected.

In conclusion, there is some evidence that we may perhaps recognize faces by a rapid sequential self-terminating scan, without replacement of rejected unimportant features, at least in an experimental situation like the one employed. It is, however, still possible that its nature and demands were such as to emphasize the features at their individual componental level, and consequently to impose a serial strategy where normally parallel processing could operate.

REFERENCES

BRACEY, G. W. Two operations in character recognition: A partial replication. Perception & Psychophysics, 1969, 6, 357-360.

EGETH, H. E. Parallel versus serial processes in multidimensional stimulus discrimination. Perception & Psychophysics, 1966, 1, 245-252.

KOLERS, P. A. The recognition of geometrically transformed text. Perception & Psychophysics, 1968, 3, 57-64.

KOLERS, P. A., & PERKINS, D. N. Orientation of letters and errors in their recognition. Perception & Psychophysics, 1969a, 5, 265-269.

KOLERS, P. A., & PERKINS, D. N. Orientation of letters and their speed of recognition. Perception & Psychophysics, 1969b, 5, 275-280.

MILLER, G. A. The magical number seven plus or minus two, or some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-97.

RESTLE, F. Mathematical models and thought:

A search for stages. In J. F. Voss (Ed.), *Approaches to thought.* Columbus: Merrill, 1969. Pp. 142-151.

ROCK, I. *The nature of perceptual adaptation.* New York: Basic Books, 1966.

STERNBERG, S. High speed scanning in human memory. Science, 1966, 153, 652-654.

STERNBERG, S. Two operations in character recognition: Some evidence from reaction-time experiments. Perception & Psychophysics, 1967, 2, 45-53.

TREISMAN, A. M. Strategies and models of selective attention. Psychological Review, 1969, 76, 282-299.

YIN, R. K. Looking at upside-down faces. Journal of Experimental Psychology, 1969, 81, 141-145.