Processing multiple physical features in facial recognition

MICHAEL R. COURTOIS and JOHN H. MUELLER
University of Missouri, Columbia, Missouri 65211

Subjects made decisions about a series of faces, then had a recognition test. The study tasks required a decision about self-reference (does it look like you), an abstract trait (e.g., friendliness), a single physical feature (e.g., thickness of lips), or multiple physical features (which is the person's most distinctive feature). As in past research, single physical feature decisions led to poorer performance than abstract trait decisions. Of greater interest here, the multiple physical feature condition was significantly better than the single physical feature condition, though still slightly worse than the abstract trait condition. These results indicate that the number of features processed is as important to facial recognition as feature "depth," and that under some circumstances encoding "shallow" attributes can produce durable memory traces.

Several recent studies have examined the effects of various orienting tasks upon subsequent facial recognition (e.g., Bower & Karlin, 1974; Mueller, Carlomusto, & Goldstein, 1978; Patterson & Baddeley, 1977; Winograd, 1976). It has been generally found that memory for faces is better when the study task involves judgments about an abstract trait (e.g., honesty) instead of a physical feature, particularly when the physical feature is primarily a facial characteristic (e.g., nose length) rather than a feature of the whole body (e.g., weight). This pattern of results has been interpreted as consistent with the levels-of-processing view of memory (Craik, 1977; Craik & Lockhart, 1972). Such an explanation assumes that abstract trait decisions lead to "deeper" or more "elaborated" encodings than physical feature decisions, because the former require more processing than when an actual facial attribute is involved. While the results are consistent with an extrapolation of the depth model as it has been applied to verbal memory, they are oddly at variance with the commonplace expectation that concentrating on a facial feature will improve memory for the face. This paradox was the question that motivated the present study.

Abstract trait and physical feature decisions may differ in a manner analogous to the semantic-nonsemantic task manipulation used to test the depth model with verbal materials, but there may be other differences as well. One possibility is that an abstract trait decision requires the examination of numerous facial features, to match some complex prototype of the honest face, for example. In contrast, the physical feature task requires a decision about a single feature. This suggests

that the differences observed in the past studies of facial recognition and orienting tasks may have been due to the number of features scanned (as well as, or instead of, the depth of the features).

One implication of this argument is that processing multiple physical features should yield recognition performance comparable to abstract trait decisions. This could be accomplished within the incidental learning format by using a task that requires the subject to identify the most distinctive feature of each face, given a finite set of attributes, since several features would have to be systematically scanned. If this group were as good as the abstract trait group, that would be contrary to the basic depth model, which assumes physical feature encodings are transitory and generally not useful.

Thus, the present study compared single and multiple physical feature tasks to an abstract trait condition and a self-reference condition. As Rogers, Kuiper, and Kirker (1977) have shown for words (see also Keenan & Baillet, 1979), we have recently found that the decision "Does this look like you?" yields very good facial recognition (Mueller, Courtois, & Bailis, Note 1), with performance levels comparable to the abstract trait decision. This condition was included here for further examination of this type of processing, and as another reference point for "deep" processing, since a self-reference decision also likely entails the scanning of several features.

METHOD

Subjects

Sixty-four students from introductory psychology courses participated in partial fulfillment of course requirements. They were randomly assigned to one of the four treatments defined by study task: abstract trait, single physical feature, multiple physical features, or self-reference decisions.

Materials

The 120 black-and-white slides were made from pictures in

0090-5054/79/070074-04$00.55/0

college yearbooks. Half were of men and half of women, all Caucasians. Slides with distinctive features (e.g., eyeglasses) were not used. Three sets of 40 slides were formed, with each being used as the study set and the distractor set across sessions.

Procedures

The study slides were presented at a 5-sec rate, under the guise of a rating study. The subjects were tested in three group sessions, with all treatments imposed each time by varying cover sheet instructions. There were two versions of the abstract trait decision (friendliness and honesty) and two versions of the single physical feature decision (lip thickness and distance between the eyes), the alternative versions of each being used equally often. No differences were observed for the versions, so analyses to be reported involve pooling over this factor. Subjects with the multiple physical feature task were required to decide whether the person's most distinctive feature was the chin, lips, nose, eyes, or forehead, and to mark one of the five alternatives on an answer sheet. In order to make the response comparable in the other three groups, those subjects indicated their decisions in terms of a 5-point rating scale on the answer sheet. (Most orienting-task studies that we are aware of have simply required binary decisions. While a rating decision might seem to require somewhat more processing, there seems little reason to expect major changes in the pattern of performance for abstract vs. physical feature judgments, and none were found here.)

The unannounced test followed immediately. The test sequence was composed of 40 old and 40 new slides, randomly ordered, presented one at a time at a 5-sec rate. Subjects marked an answer sheet to indicate whether each slide had been used in the study set. No constraints were placed on guessing or the number of "old" responses allowed. Subjects were informed of the number of old and new slides in the test sequence.

The subjects were administered the Test Anxiety Inventory (Spielberger, Gonzalez, Taylor, Algaze, & Anton, 1978) prior to the study phase, and after the test phase they completed the State Anxiety Inventory (Spielberger, Gorsuch, & Lushene, 1970). These questionnaires were used to obtain further evidence on the relationship between anxiety and facial recognition (Mueller, Bailis, & Goldstein, in press). At the end of the experiment, the Betts Questionnaire Upon Mental Imagery and the Gordon Imagery Scale (Richardson, 1969) were given in an effort to determine the effect of imaginal ability on facial recognition. Finally, subjects were asked a series of debriefing questions requiring them to rate their anxiety during the experiment, perceived benefits of the study tasks, and general confidence in "old" and "new" responses on the test, using a 5-point scale.

RESULTS

The primary results are summarized in Table 1. In general, the single physical feature group did worse than the other three, and while the abstract trait group was the best overall, it was not always significantly better than the self-reference and multiple physical feature groups.[1]

The task main effect was not significant for false alarm rate $[F(3,60) = 2.12, MSe = .012]$, nor for criterion level (β) $(F < 1)$. The task effect was significant for hit rate $[F(3,60) = 3.38, MSe = .017]$. The l.s.d. test value for differences between means was .09, showing the only significant $(p < .05)$ difference to be between the two extreme cases (abstract trait and single physical feature). The task effect was also significant

Table 1
Mean Recognition Performance by Orienting Task

	Abstract	Self	Single	Multiple
			Physical Feature	
Hit Rate	.84	.77	.69	.77
False Alarm Rate	.12	.13	.21	.15
d'	2.39	2.10	1.51	1.94
β	2.74	2.31	2.07	1.81
A'	.92	.89	.83	.88

for d' $[F(3,60) = 4.35, MSe = .502]$. The l.s.d. test value was .50, showing the single physical feature group to be significantly worse than both the abstract trait and the self-reference groups, with no other reliable differences.

The nonparametric signal detection measure $(A'$; Grier, 1971) also showed a significant task main effect $[F(3,60) = 5.64, MSe = .004]$. The l.s.d. value was .045, so the single physical feature group was worse than each of the others, with no other differences significant.

Table 2 shows the correlations between recognition performance and the anxiety and imagery questionnaire scores. Two scores were computed from the Betts questionnaire, a full-scale score over various modalities, and the visual subscale score alone. As can be seen, none of the correlations was significant (for the .05 level, r would have to exceed .24). This may be a result of the lack of an adequate range in the scores, and selection of extreme cases for anxiety or imagery might show differences. The lack of a correlation between imagery ability and facial recognition seems odd, but this is the second experiment in which we have obtained this outcome (Mueller et al., Note 1). In fact, perhaps this is the modal outcome with regard to relationships between self-report measures of imagery and performance on various tasks (cf. Hiscock, 1978).

DISCUSSION

The initial hypothesis seems confirmed: Processing multiple shallow features can lead to good recognition performance. The present results indicate improved performance when multiple physical features are processed compared to a single feature. Performance still did not completely match that achieved when

Table 2
Correlations Between Recognition Performance, Anxiety Scores, and Mental Imagery Scores

	Hits	False Alarms	d'	β
Test Anxiety				
Worry	−.03	.04	−.04	−.15
Emotion	.04	.12	−.05	−.20
State Anxiety	.01	.06	−.05	−.07
Imagery				
Betts (Total)	−.03	−.01	−.06	−.13
Betts (Visual)	.07	−.03	.04	−.07
Gordon	.08	−.17	.15	.02

abstract traits were processed, but the remaining difference was not statistically significant.

Shortly after this experiment was underway, we discovered that a similar experiment had been performed by Winograd (1978). Winograd compared an abstract trait condition to a distinctive feature condition, where subjects picked the most outstanding feature of each face from among nine alternatives, and a single physical feature condition. His results indicated even more clearly that the number of features studied is the critical determinant, as his distinctive feature group performed at a level virtually identical to the abstract trait group. Perhaps the greater improvement in his study was due to the use of nine rather than five features examined in the multiple physical feature conditions. Alternatively, perhaps there was some additional benefit our abstract trait group received from making rating decisions rather than binary judgments.

The results of these two experiments require at least a modification of the depth interpretation of orienting task effects in facial recognition. At the very least, it is necessary to accept the fact that encodings of physical features can lead to permanent memory traces. Of course, the retention intervals used here are not long, so it could be argued that shallow encodings merely sufficed, in a manner analogous to immediate free recall vs. final free recall, and would not have held up given a longer interval before the test.

Acceptance of this argument presents several problems, though. For one thing, the intervals here were long enough to produce the usual deficit for the single physical feature condition. Therefore, something must be occurring with multiple feature encoding to survive this delay. One possibility is that elaborative encoding can occur with shallow features, yielding durable memory traces, at least at this retention interval. However, such elaborative encoding is not generally considered likely with shallow features. All in all, the explanation of the results of these two experiments is perhaps more straightforward simply in terms of the number of features encoded. By this analysis, facial recognition is good for the abstract trait, self-reference, and multiple physical feature conditions here because those decisions require attending to more aspects of the face during study.

REFERENCE NOTE

1. Mueller, J. H., Courtois, M. R., & Bailis, K. L. *Self-reference in facial recognition.* Paper read at the annual meetings of the Midwestern Psychological Assoc)Eds.), Chicago, 1979.

REFERENCES

Bower, G. H., & Karlin, M. B. Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology,* 1974, **103**, 751-757.

Craik, F. I. M. Depth of processing in recall and recognition. In S. Dornic (Ed.), *Attention and performance* (Vol. 6). Hillsdale, N.J: Erlbaum, 1977.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior,* 1972, **11**, 671-684.

Grier, J. B. Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin,* 1971, **75**, 424-429.

Hiscock, M. Imagery assessment through self-report: What do imagery questionnaires measure? *Journal of Consulting and Clinical Psychology,* 1978, **46**, 223-230.

Keenan, J. M., & Baillet, S. D. Memory for personally and socially significant events. In S. Dornic (Ed.), *Attention and performance* (Vol. 8). Hillsdale, N.J: Erlbaum, 1979.

Mueller, J. H., Bailis, K. L., & Goldstein, A. G. Depth of processing and anxiety in facial recognition. *British Journal of Psychology,* in press.

Mueller, J. H., Carlomusto, M., & Goldstein, A. G. Orienting task and study time in facial recognition. *Bulletin of the Psychonomic Society,* 1978, **11**, 313-316.

Patterson, K. E., & Baddeley, A. D. When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory,* 1977, **3**, 406-417.

Richardson, A. *Mental imagery.* New York: Springer, 1969.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology,* 1977, **35**, 677-688.

Spielberger, C. D., Gonzalez, H. P., Taylor, C. J., Algaze, B., & Anton, W. D. Examination stress and test anxiety. In C. D. Spielberger & I. G. Sarason (Eds.), *Anxiety and stress* (Vol. 5). Washington, D.C: Hemisphere/Wiley, 1978.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. *Manual for the state-trait anxiety inventory.* Palo Alto, Calif: Consulting Psychologists Press, 1970.

Winograd, E. Recognition memory for faces following nine different judgments. *Bulletin of the Psychonomic Society,* 1976, **8**, 419-421.

Winograd, E. Encoding operations which facilitate memory for faces across the life span. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory.* London: Academic Press, 1978.

NOTE

1. Effects described as significant involve $p < .05$ or better.