Orienting task and study time in facial recognition

JOHN H. MUELLER, MICHAEL CARLOMUSTO, and ALVIN G. GOLDSTEIN
University of Missouri, Columbia, Missouri 65201

In two experiments, subjects were required to make various decisions about photographs of faces and were then given an unannounced recognition test. The deep processing decision involved a personality trait (e.g., generosity), another task involved an attribute solely determined by some physical feature of the face (e.g., height of forehead), while the third type of judgment involved a physical attribute of the body rather than the face alone (e.g., weight). In Experiment 1, subjects who had 15 sec per face during study performed better on the recognition test than subjects who had 5 sec, but there was no Rate by Task interaction. In both experiments, the Face condition showed fewer hits and lower d' than the Body and Deep groups, but there was no significant advantage for the Deep group over the Body group.

It has been shown that the subject's activities during study will determine how well words are remembered. Orienting tasks that require processing of the semantic features lead to better retention than do tasks that require attention to sensory features (e.g., Hyde & Jenkins, 1973). The levels-of-processing view of memory (Craik & Lockhart, 1972) explains this result by arguing that retention is improved by (1) deep processing as opposed to shallow, where "deep" corresponds to semantic content and "shallow" refers to physical features, and (2) by rehearsal which "elaborates" the stimulus trace rather than merely "maintaining" it (Craik & Tulving, 1975).

Several recent studies obtained similar results for the effect of orienting tasks on recognition memory for facial photographs (e.g., Bower & Karlin, 1974; Patterson & Baddeley, 1977; Strnad & Mueller, 1977; Winograd, 1976). Even though the semantic-nonsemantic dimension does not seem as directly applicable to unfamiliar facial stimuli, subjects do seem to remember faces better following decisions about some abstract trait (e.g., honesty) compared to some physical trait (e.g., gender). To our knowledge, only one previous study has failed to produce this advantage for deep decisions (Hall, Note 1).

The present study was part of an effort to further identify tasks that facilitate facial recognition. A previous study (Mueller, Bailis, & Goldstein, Note 2) examined whether or not the orienting task was differentially effective when the judgment involved a comparison to an absolute criterion (e.g., "Is the person taller than 5 ft 6 in.") vs. a subjective criterion (e.g.,

"...taller than you"). Rogers, Kuiper, and Kirker (1977) found that self-reference decisions led to better memory for words. However, although Mueller et al. found better performance following abstract judgments than after shallow tasks, there was no main effect of reference point nor any Level by Reference interaction.

It was noted by Mueller et al. (Note 2) that the nature of the physical attributes used for shallow judgments about facial photographs had differed across studies. In some cases, the physical feature was an attribute of the face per se (e.g., nose length), while in other cases it was a property of the body (e.g., weight) rather than the face alone. The one study which failed to find an advantage for abstract tasks (Hall, Note 1) focused on details that might be described for a police artist, thus it involved many purely facial features. Traits such as honesty seem clearly "abstract," but features like height are in a sense also abstract compared to interocular distance, in that height is not immediately visible in the photograph. Thus, while both facial and body features have been mixed as ostensibly shallow tasks in previous research, it seems desirable to make a systematic comparison of these two types of physical-feature tasks and to compare each to the "deeper" judgments.

The other factor examined in Experiment 1 was exposure duration during the study phase. Hall used a rather slow rate of presentation (7.5-sec exposure and 12.5-sec interslide interval) compared to the other studies (typically about 5-sec exposure and 1.0-sec interslide interval). Longer exposures could attenuate the depth advantage, if subjects automatically process more thoroughly when the task does not require the full exposure period. Although it is possible to separate exposure rate and interslide interval (cf. Evans, 1977; Weaver, 1974), it was decided to systematically vary the former to determine its impact on orienting task effects in facial recognition.

EXPERIMENT 1

Method

Design and Subjects. Seventy-two students from introductory psychology courses served as subjects as part of course requirements. The experimental design involved the factorial combination of three orienting tasks (deep, face, body) and two study rates (5 sec, 15 sec), with men and women randomly distributed in the six groups.

Materials. The stimuli were 100 black-and-white slides of faces from college yearbooks, half male and half female, all Caucasians. Pictures with distinctive cues (e.g., glasses) were not included, yielding a rather homogeneous set of stimuli. These were divided into two sets, with each half being used about equally often as the study set. [The stimuli were exactly the same as those used by Mueller et al. (Note 2), and virtually identical to those used by Strnad and Mueller (1977).]

Procedure. For the study phase, subjects were read instructions that indicated the experiment was concerned with how fast they could make decisions about people based only on facial photographs. They were provided with a sheet of paper describing their particular decision and a computer answer sheet on which to mark their judgments (true or false).

There were three basic decisions, with four specific versions of each. A subject made only one type of decision (and one version of it) throughout the session. In each case, the subject had to decide whether the person in the photograph was above or below average on some feature. The "deep" task directed attention to an abstract trait, namely, intelligence, friendliness, generosity, and confidence. There were two types of physical-feature judgments, involving either facial features per se or the body as a whole. The specific attributes for the "face" decision group were size of nose, thickness of lips, distance between eyes, and height of forehead. The attributes for the "body" group were weight, height, muscularity, and posture.

Subjects were run in small groups. The 50 study slides were randomly ordered and presented one at a time at either a 5- or 15-sec rate. Subjects were told to mark their decisions as rapidly as possible and continue watching the face until the next slide appeared. There was no mention of a subsequent recognition test.

The test phase followed immediately. The 50 study slides were randomly ordered with 50 new slides, and these were shown one at a time at a 5-sec rate. Subjects were told to mark on their answer sheet whether they thought each slide had been used in the study set. They were not restricted to "50" old responses, and no mention was made for or against guessing. Following the test phase, subjects completed the State-Trait Anxiety Inventory (Spielberger, Gorsuch, & Lushene, 1970) and a questionnaire that asked them to rate their confidence at various phases in the experiment, their anxiety, their expectancy of a retention test, the perceived difficulty of the task, and the perceived benefit provided by the orienting task, all on 5-point scales.

Results[1]

Table 1 summarizes recognition performance. The four different versions of each orienting task were examined in three separate one-way analyses of variance (pooled over rate). None of the measures shown in Table 1 revealed a significant difference [Fs(3,20) < 2.41], so the versions were pooled in subsequent analyses.

The Face group had a significantly lower hit rate than the Body and Deep groups [F(2,66) = 3.55, MSe = .018], and the latter two groups were not signif-

Table 1
Mean Recognition Scores in Experiment 1 by Orienting Task and Presentation Rate During Study

	5-Sec Rate			15-Sec Rate		
	Deep	Face	Body	Deep	Face	Body
Hit Rate	.60	.57	.66	.72	.58	.68
False-Alarm Rate	.20	.24	.26	.17	.15	.16
Old Responses	39.75	40.67	45.83	44.58	36.08	41.92
d'	1.13	.93	1.08	1.61	1.31	1.55
β	1.46	1.48	1.23	1.39	1.82	1.69

icantly different. The slower rate during study led to slightly more hits, especially for the Deep group, consistent with the notion that deep processing takes longer (Craik & Tulving, 1975). However, there was no rate main effect nor Task by Rate interaction (Fs < 2.51).

For the false alarms, there was no difference due to type of orienting task, nor was there a Task by Rate interaction (Fs < 1.42). However, false alarms were reduced with the slow rate [F(1,66) = 12.60, MSe = .008].

The results were also analyzed in signal detection terms.[2] While d' was consistently lower in the Face condition, there was no main effect of task [F(2,66) = 2.26, MSe = .184, p < .12]. Comparing individual tasks, collapsing over rate, the Face group was significantly worse than the Deep group (p < .05) but not the Body group, the latter two groups not differing. There was no Task by Rate interaction (F < 1), but performance was better for subjects who had the slow rate during study [F(1,66) = 19.05, MSe = .184]. The criterion measure (β) revealed no main effects due to task (F < 1) or rate [F(1,66) = 2.66], nor their interaction (F = 1.14).

There were no significant effects due to task or rate or their interaction for the total number of "old" responses, hits plus false alarms (Fs < 2.29). Similarly, there were no differences in terms of the number of failures to respond (Fs < 2.25).

There were no differences in state, trait, or rated anxiety within the subgroups of the experiment (Fs < 2.92). Over all subjects, self-rated anxiety correlated significantly with hit rate (r = −.35, p < .004), false alarms (r = −.35, p < .004), and β (r = .42, p < .001), but not with d' (r = .0). However, median-split analyses involving state and trait anxiety scores failed to show performance differences.

In other analyses, mean rated test expectancy was 2.67, 2.63, and 3.08 for the Deep, Face, and Body conditions, respectively, which was not a significant difference [F(2,66) = 2.07]. Rated benefit from the orienting task was significantly correlated with d' (r = .27, p < .03) but not with hits (r = .07), false alarms (r = −.23, p < .07), or β (r = .22, p < .07). The mean ratings were 3.81, 3.00, and 3.45 for the Deep, Face, and Body groups, respectively [F(2,66) = 2.40, MSe = 1.55, p < .10].

EXPERIMENT 2

Method

A second experiment was conducted to replicate Experiment 1, especially the nonsignificant advantage for the Deep group over the Body group. All procedures were identical to Experiment 1, except that only the 5-sec rate was used, and the State-Trait Anxiety Inventory was replaced by the Test Anxiety Inventory (Spielberger, Gonzalez, Taylor, Algaze, & Anton, 1978). Forty-five students in an introductory learning class were tested en masse on the first day of class.

Results

Table 2 summarizes recognition performance in Experiment 2. The hit rate was again lower for the Face group than for the other two $[F(2,42) = 10.13,$ MSe = .013], with the Body and Deep groups not significantly different. The same pattern was also present for old responses $[F(2,42) = 4.41,$ MSe = 63.91] but not for false alarms (F < 1).

The Face group also had a lower d' than the other groups $[F(2,42) = 6.12,$ MSe = .194], with no significant difference between the Body and Deep groups. There were no differences for the criterion measure (F < 1).

There were no differences between tasks due to test expectancy or decision aid (Fs < 1.10). The subjects who had the Deep task were more confident about their rejection of new items than subjects who had a Face or Body task (means = 2.47, 3.27, and 3.33, respectively) $[F(2,42) = 3.85,$ MSe = .905], but there was no difference in confidence about their "old" responses (F < 1). Subjects who had a Deep task rated their arousal lower than those in the Face and Body conditions (means = 1.13, 1.87, and 1.93, respectively) $[F(2,42) = 3.61,$ MSe = .819]. Rated anxiety was correlated with hit rate (r = −.33) but not with false alarms, d', or β (rs = .02, −.27, and −.01, respectively). A median split revealed that high test anxiety was associated with lower d' than low anxiety [t(43) = 2.04], but there was no difference for the other recognition measures.

DISCUSSION

Both experiments found that attention to facial features during study led to poorer recognition than judgments about abstract traits or gross bodily features. In addition to the curious implications of this result for eyewitness testimony and police artist reconstructions (see also Patterson & Baddeley, 1977),[3]

Table 2
Mean Recognition Scores in Experiment 2 by Orienting Task During Study

	Deep	Face	Body
Hit Rate	.67	.48	.59
False-Alarm Rate	.23	.25	.22
Old Responses	45.27	36.60	40.87
d'	1.23	.69	1.06
β	1.24	1.45	1.47

there is the further fact that there was no significant difference between the Body and Deep conditions. It would be tempting to conclude that these are both equally "abstract" decisions in the sense outlined above, but this seems at variance with previous findings. Earlier studies have found significantly better performance for deep judgments when the shallow judgments involved some bodily features, for example, dependable and intelligent vs. height and weight in the Mueller et al. (Note 2) study, using the same stimulus set as was used here. Admittedly, the rank order difference for the Deep and Body conditions was consistent with previous results. The present results and those of Hall (Note 1) suggest that the deep-body difference may not be particularly robust for facial recognition. Even though the differences were not statistically significant between the various versions used here, there was some variation, so the deep-body difference may depend on just what specific tasks are compared. Although Hall's procedures differed in many respects from earlier studies, the results of Experiment 1 here argue that rate is not the critical confounding factor. Furthermore, many of the differences between Hall's methods and earlier studies did not hold for the present experiments vs. earlier studies, which then seems to indirectly rule out many obvious methodological differences (e.g., interslide interval).

One possibility is that the 5-sec rate is too long and that deep processing proceeds automatically even in the Body condition. However, there is no obvious reason why such surreptitious deep processing would not have occurred in the Face condition as well, though it apparently did not, and, in any event, previous studies have found the deep-shallow effect at 5-sec rates. Another possibility is that the depth advantage would be greater on a delayed test, since the levels-of-processing view assumes that deep traces are more durable than shallow (Craik & Lockhart, 1972). While this is possible, earlier studies have found this effect even with an immediate recognition test, and it has also been observed that the length of the retention interval is not a potent factor in facial recognition (e.g., Laughery, Fessler, Lenorovitz, & Yoblick, 1974). Finally, it is conceivable that less homogeneous stimulus sets yield greater deep-shallow differences (though one might hope that deep processing would help even more when the task is difficult), but in the absence of standardized materials that possibility is difficult to evaluate.

REFERENCE NOTES

1. Hall, D. F. *Obtaining eyewitness identifications in criminal investigations: Two experiments and some comments on the zeitgeist in forensic psychology.* Paper read at the meetings of the American Psychology-Law Society, Snowmass, Colorado, June 1977.

2. Mueller, J. H., Bailis, K. L., & Goldstein, A. G. *Depth of processing and anxiety in facial recognition.* Paper read at the meeting of the Midwestern Psychological Association, Chicago, 1978.

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

CRAIK, F. I. M., & TULVING, E. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 1975, **1**, 268-294.

EVANS, J. D. On the inconstant effects of study instructions. *American Journal of Psychology*, 1977, **90**, 511-516.

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

HYDE, T. S., & JENKINS, J. J. Recall for words as a function of semantic, graphic, and syntactic orienting tasks. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 471-480.

LAUGHERY, K. R., FESSLER, P. K., LENOROVITZ, D. R., & YOBLICK, D. A. Time delay and similarity effects in facial recognition. *Journal of Applied Psychology*, 1974, **59**, 490-496.

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

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). New York: 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.

STRNAD, B. R., & MUELLER, J. H. Levels of processing in facial recognition memory. *Bulletin of the Psychonomic Society*, 1977, **9**, 17-18.

WEAVER, G. E. Effects of poststimulus study time on recognition of pictures. *Journal of Experimental Psychology*, 1974, **103**, 799-801.

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

NOTES

1. Effects reported as significant involve ps < .05 or better, unless noted otherwise.

2. Analyses were also performed using nonparametric measures (Grier, 1971). Since A' yielded conclusions identical to d', the nonparametric results will not be presented.

3. It should be noted that the analogy to eyewitnesses is not perfect. Specifically, there is normally no orienting task during the "study" phase there, and any emphasis (orienting) occurs during the line-up or reconstruction (test) phase instead. However, the present results should still have some bearing on relative expectations for an emphasis on facial attributes vs. other emphases.