

Levels of processing in facial recognition memory

BRUCE N. STRNAD and JOHN H. MUELLER
University of Missouri, Columbia, Missouri 65201

Subjects studied pictures of faces while making judgments of honesty ("deep" processing) or sexual identification ("shallow" processing). In each case, half of the subjects were informed that there would be a subsequent test, while half were not so informed. Deep processing led to superior recognition memory, independent of test expectancy.

It has been argued (Craik & Lockhart, 1972) that memory strength is a direct function of the "depth" of processing of the stimulus material. If the stimulus is processed in terms of sensory (shallow) features, then the resultant memory trace will presumably be less durable than if the stimulus had been processed in terms of semantic (deep) features. Virtually all of the research on this question has used verbal stimuli. For example, it has been found that subjects who answer orienting questions about the pleasantness of words perform better than subjects who simply report the number of letters in the word, even in the absence of a test expectancy, that is, an incidental learning set (e.g., Hyde & Jenkins, 1969). Very little has been done involving pictorial material, and it was the purpose of the present study to replicate one of the few studies (Bower & Karlin, 1974) which used pictures as a test of the depth of processing analysis.

Bower and Karlin (1974, Experiments 1 and 2) presented pictures of faces to subjects and required judgments of honesty, likableness, or gender, under the cover of a reaction time experiment. A subsequent recognition test revealed better performance after the honesty and likableness decisions (deep processing) than after the sex decision (shallow processing), whether or not subjects were informed of the eventual memory test. This pattern of results corresponds to that found with verbal materials (e.g., Hyde & Jenkins, 1969), even though unfamiliar faces as such do not have semantic meaning to identify as the factor involved in deep processing. Bower and Karlin (1974) hypothesized that a "deep" analysis of a face likely involved the examination of more details or features for comparison to some vague prototype (e.g., an "honest face").

The purpose of the present experiment was to further examine the effect of orienting task and test expectancy in facial recognition. Whereas each subject in the Bower

and Karlin study made both shallow and deep processing judgments for different stimuli, subjects in the present experiment made only one type of judgment throughout. Bower and Karlin (1974, p. 756) noted that their subjects might even have been able to further state what type of judgment they had originally made for each old item. However, a between-subjects design provides greater homogeneity in terms of such contextual cues, so that the old-new decision should be less affected by extra features of that nature. This change from a within-subjects to a between-subjects design was the primary procedural difference, along with the elimination of the likableness decision. Other minor changes are noted below.

METHOD

Design and Subjects

Forty students participated as part of course requirements in introductory psychology. They were randomly assigned to the four groups formed by the factorial combination of two levels of test expectancy (informed, uninformed) and two types of orienting decisions (honesty, gender).

Materials

One hundred black-and-white 35-mm slides of pictures from a college yearbook were used. Half of the slides were of male faces and half were females. All pictures were Caucasians. Distinctive cues (e.g., a hat) were eliminated to assure uniformity. These were divided into two sets of 25 male and 25 female slides, Set A and Set B, each used equally often as the study set. Bower and Karlin (1974) used 144 stimuli and did not report counterbalancing halves as the study set.

Procedure

In the study phase of the experiment, all subjects were led to believe that they were participating in a reaction time experiment. They were told to decide as rapidly as possible whether the person in each photograph was honest or dishonest (male or female), and to indicate that decision by pressing the appropriate button on the panel in front of them. The left-right position of the button for each outcome, honest or dishonest (male, female), was balanced across subjects. Wires ran from the panel into an adjoining room, but these were not connected to any equipment. The cover task was emphasized by asking subjects if they had served in such an experiment before and noting the need to press the button firmly. In addition, the experimenter went into the adjoining room briefly, just prior to the first slide, "to activate the response timing apparatus," and

Requests for reprints should be sent to J. H. Mueller, Psychology Department, 210 McAlester Hall, University of Missouri, Columbia, Missouri 65201. The authors wish to acknowledge Alvin Goldstein, Donald Kausler, and Matt Marler for their comments on an earlier draft.

turned on a tape recording of various laboratory noises (e.g., relays clicking).

The slides in the study set were randomly ordered and presented one at a time at a 4-sec rate using a Carousel projector. Subjects were told to make their decisions as rapidly as possible (using whatever criteria they wished for honesty) and to continue watching the slide until it changed. Half of the subjects making each type of judgment were told that there would be a retention test later (intentional learning set), while the others were not (incidental set).

In the test phase, subjects were told they would view a new set of slides, including some from the earlier set. The 50 study slides were randomly mixed with 50 new slides and shown one at a time at a 5-sec rate. As each slide was shown, subjects indicated on a response sheet whether they thought that face had been in the study set, plus how confident they were of their decision, using a 6-point rating scale. Subjects were not required to make just 50 "old" responses during the test phase, nor did they know that all 50 old items were carried over. Guessing was neither encouraged nor penalized.

RESULTS¹ AND DISCUSSION

Table 1 presents the average recognition results for each group. The initial deep processing task (honesty judgment) led to more correct identifications of old faces (hit rate) [$F(1,36) = 4.03$, $MSe = .026$], but neither the test expectancy main effect nor the Task by Expectancy interaction was significant ($F_s < 1$). While the deep processing task also led to slightly higher false alarm rates, the orienting task main effect was not significant [$F(1,36) = 2.61$, $p < .12$, $MSe = .027$], nor was the test expectancy main effect or Task by Expectancy interaction ($F_s < 1$). No effects were significant for either new- or old-item confidence ratings [$F_s < 1.39$, $MSe = .435$ and $.399$, respectively]. Simply considering the number of items identified as "old" during the recognition phase, there was a significant orienting task main effect [$F(1,36) = 5.21$, $MSe = 164.247$], as shallow processing led to lower levels of responding "old" on the test, with no test expectancy main effect or Task by Expectancy interaction ($F_s < 1$).

These basic results are in substantial agreement with the findings of Bower and Karlin (1974) with pictures, as well as the results with verbal materials: Deeper processing improves retention, even in the absence of instructions about a test. In the present case, since

subjects made only one type of judgment during the orienting task, contextual cues about the type of decision made for a given old item should be more uniform in their effect when the test items are examined. However, this seems not to affect the basic results. Furthermore, the lower rate of responding "old" on the test suggests a less thorough analysis when the subjects had earlier made a shallow judgment: The old items thus truly look more like the distractors.

About the only notable differences in the results of the two experiments are the lower hit rates and new-item confidence ratings here. The confidence ratings are readily explained by noting that the 6-point rating scale used by Bower and Karlin (1974) translates to "1 = old" through "6 = new," with the binary old-new decision thus deduced indirectly. However, the present scale is "1 = sure" through "6 = guessing," subsequent to and independent of the outcome of the separate old-new judgment. Consequently, our confidence ratings for old items need not necessarily be lower than for new items. The hit rate differential is not as clearly a procedural artifact, but it does serve to underline the fact that picture recognition need not always be virtually perfect (cf. Goldstein & Chance, 1974). The lower hit rates may be due to greater homogeneity in our slides, with the test and study sets being more similar than was the case for Bower and Karlin's materials. If so, this was unintentional, and unavoidable in the absence of any normative scaling of similarity (or other attributes) for such materials. Further, the lower hit rates may be due in part to the different demands of our test procedure. That is, since our subjects had to make two decisions (old or new and certainty) in 5 sec, compared to the single decision required of Bower and Karlin's subjects, it may be that the reduction in functional decision time is somehow involved. In any event, the overall pattern of results was quite similar for the two studies.

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.
- GOLDSTEIN, A. G., & CHANCE, J. Some factors in picture recognition memory. *Journal of General Psychology*, 1974, **90**, 69-85.
- HYDE, T. S., & JENKINS, J. J. The differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology*, 1969, **82**, 472-481.

NOTE

1. Effects described as significant involve at least $p < .05$, unless noted otherwise.

Table 1
Average Recognition Performance by Orienting Task
and Test Expectancy

	Honesty Judgment		Sex Judgment	
	Inci- dental	Inten- tional	Inci- dental	Inten- tional
Hit Rate	.52	.58	.44	.47
False Alarm Rate	.36	.32	.26	.25
Confidence Rating, Old Items*	2.79	2.84	2.93	2.94
Confidence Rating, New Items*	2.99	3.13	2.83	3.16
Number of "Old" Responses	43.60	44.60	34.50	35.20

*Rating scale anchors were "1 = sure" and "6 = guessing"