Semantic interpretation effects on memory for faces

ROBERTA L. KLATZKY, GALE L. MARTIN, and ROBBIE A. KANE
University of California, Santa Barbara, California 93106

Five experiments investigated the effects of semantically interpreting faces on their recognition. Experiment 1 demonstrated that faces could meaningfully be related to occupational categories. In Experiments 2-5, each of a set of faces was presented with either the label of a congruent occupational category or a noncongruent label. Presenting a face with a congruent occupational label was found to enhance recognition that the face had previously been seen (Experiments 2-4), but congruent labeling also impaired the detection of distractor faces that matched previously seen faces' categories (Experiments 3 and 4). In a forced-choice recognition test, in which distractors were highly similar to originally seen faces, congruent occupational labeling failed to increase recognition of previously seen faces (Experiment 5). These results indicate that interpreting faces with respect to stereotyped categories, rather than improving a physical code in memory, led to the formation of a semantic code, which increased recognition of old items at the expense of the detection of new items. Another general finding was that faces that were rated more stereotypical were more recognizable, regardless of the accompanying label.

The present experiments address the question of whether memory for faces is enhanced when the faces are interpreted in terms of occupational categories. As such, the experiments are part of a larger research effort exploring the effects of semantic interpretation on memory for visual stimuli. This research has adopted a general distinction between two kinds of memorial information about visual stimuli: information about their conceptual contents and information about their physical detail (Bartlett, Till, & Levy, 1980; Klatzky & Stoy, 1978; Mandler & Ritchey, 1977; Nelson, Reed, & Walling, 1976; Rafnel & Klatzky, 1978). The first type of information has been called semantic, conceptual, meaningful, or verbal; the second has been called physical, structural, schematic, sensory, pictorial (Klatzky, in press), or simply "visual." They will be called here the semantic and physical codes.

Put in these terms, the present studies address whether the act of semantically interpreting a face, or, more specifically, relating it to a category whose stereotype it matches, affects memory, and whether it does so by influencing a semantic code or a physical code.¹ Although these codes cannot be specified precisely, the physical code of a face is assumed to include concrete visual attributes and to be achieved with minimal influence from "top-down" processing, and the semantic code is assumed to represent the face as a meaningful concept. The semantic code might include information about visual appearance, such as "looks clean-cut" or "has

moustache," but this visual information would be more abstract than that of the physical code. The semantic code could also include information not directly related to vision, for example, regarding the probable income or the beliefs of the person whose face is being viewed. (The semantic code might also incorporate lexical information about a label for the face, although such information is of lesser interest here.)

The question of whether and how semantically interpreting a visual stimulus affects memory is of particular interest regarding faces, because there is a substantial literature suggesting that face recognition is mediated predominantly by physical codes. For example, extensive training in the verbal description of facial features fails to improve recognition (Baddeley, 1979; Malpass, Lavigneur, & Weldon, 1973). A link between face processing and physical coding is also suggested by a frequently observed advantage for processing faces presented to the right cerebral hemisphere (see Sergent & Bindra, 1981, for a review), which indicates dominant contributions are from the nonverbal domain.

Notwithstanding the importance of the physical code to face memory, semantic interpretation could still play an important role. Rafnel and Klatzky (1978) reviewed the evidence for two types of semantic interpretation effects on memory for visual patterns; these same effects could also apply to face memory. One type of effect, which is consistent with the idea that the physical code is of primary importance in remembering faces, would be for semantic interpretation of a face to enhance its physical code. Relevant to this hypothesis is research that attempts to extend the levels-of-processing approach (Craik & Lockhart, 1972) to face recognition.

This research was supported by Grant MH 25090 from the National Institute of Mental Health to the first author. Reprint requests should be sent to Roberta Klatzky, Department of Psychology, UCSB, Santa Barbara, California 93106.

Several studies (Bower & Karlin, 1974; Patterson & Baddeley, 1977) have shown that making semantic judgments about faces at encoding (e.g., judging personality traits) leads to better recognition than making physical judgments (e.g., judging thickness of lips). Investigating the basis for this effect, Winograd (1981) found evidence that semantic judgments are effective because they lead to scanning a face as a whole, which in turn increases the probability of encoding a distinctive feature. Thus, this would constitute an effect of semantic interpretation on the physical code. Other situations in which interpretation might affect physical coding are also plausible. For example, when interpretation corresponds to relating the face to a category stereotype, as is the case here, categorical knowledge might direct the viewer to particularly informative regions of the face, leading to encoding of more physical detail.

The second means by which semantic interpretation might enhance memory for faces is by providing a distinct semantic code, which, when associated with the physical code in memory, would increase the amount of information stored about the face. This would be essentially the reverse of Paivio's (1971) hypothesis that visual imagery enhances memory for verbal items by adding a physical code. Here, interpretation would enhance memory for visual items by adding a semantic code. Rafnel and Klatzky (1978) found evidence for just such an effect: When subjects interpreted nonsensical visual stimuli (interpretation was induced by providing a meaningful label), the effect was to increase their ability to detect meaning-changing distractors on a subsequent test, but not meaning-preserving ones. This suggests that the interpretation provided a semantic code but did not augment a physical code. Providing a categorical interpretation for a face could have a similar effect.

In the present studies, semantic interpretation of a face was induced by presenting, along with the face, a category label that the face had previously been judged to match. The semantic categories that were used were common occupations. These were chosen because there is evidence that the memorial representations of at least some occupations include substantial information about facial appearance. This evidence stems from psychological studies of facial stereotyping by occupation (Klatzky, Martin, & Kane, in press; Litterer, 1933; Shoemaker, South, & Lowe, 1973), and it is also indicated by everyday stereotyping of faces (e.g., in advertising). At the same time, occupations are not so obvious that faces are unambiguously classifiable without a category label, as other facially relevant categories (e.g., race, sex) would be. Thus, it should be possible to direct interpretation of a face by category labeling.

The general idea behind these experiments was that if semantic interpretation of this sort enhanced the physical code of a face, several findings would result.

First, recognition should be improved. That is, recognition should be better for faces that were seen initially in conjunction with a matching occupational label than for those seen with a nonmatching label or a neutral, control label. And, previously unseen faces should be more discriminable from faces that were initially seen with a matching, rather than nonmatching, label. Moreover, the effect of supplying an occupational label should be greater, the more relevant that label is (i.e., the better the fit is between the labeled face and the occupation).

The other hypothesis to be considered is that supplying a matching occupational label for a face, rather than enhancing its physical code, will lead to the formation of an ancillary semantic code, representing the occupational category in general. If two codes are better than one, the ability to recognize a labeled face should be facilitated by the addition of the semantic code (as the physical code hypothesis also predicts). But because the semantic code is sufficiently abstract to represent the category in general, as a stereotype, new faces that bear a categorical resemblance to the given old face will tend to match that code. And this could lead to errors in detecting that the faces are new. Such category-induced errors have previously been found in memory for text (Dooling & Lachman, 1971) and visual patterns (Daniel, 1972; Franks & Bransford, 1971).

These hypotheses were investigated in five experiments. Experiment 1 was a preliminary study that verified that faces could be matched with occupational categories and that occupational categorization was not entirely unambiguous. Experiment 2 addressed the general questions of whether occupational labeling of faces would facilitate their recognizability: the more so, the better the face-category match. Experiments 3-5 probed more deeply for the source of labeling effects, to determine whether the interpretation of a face with respect to an occupation affected a physical or semantic code.

EXPERIMENT 1

Method

The purpose of Experiment 1 was to attempt to collect face photographs that could reliably be assigned to a set of occupational categories. An initial set of categories was obtained from the *Dictionary of Occupational Titles* (1977) and the authors' intuitions. From occupations listed in these sources, 30 were selected, and members of the psychology department at the University of California at Santa Barbara (UCSB) were asked to rate each category according to the goodness of their mental prototypes of category members' faces. They were also asked to describe features of their prototypes. From the initial set, 13 male occupations that had been given the highest ratings were selected for further study. The categories were: accountant, athlete, farmer, hairdresser, judge, model, pilot, policeman, rock musician, salesman, truck driver, undertaker, and watchmaker. (Females were eliminated because it was feared that mixing genders within a category would reduce its prototypicality, and because it was impossible to find as many female-prototype categories as male.) A collection of 83 male-face photographs (with the clothing cut away so that only the face

was exposed) was then culled from magazines and yearbooks, following the guidelines suggested by the Departmental responses. Each was tentatively assigned to 1 of the 13 occupational categories so that there were four to seven potential prototypical faces for each category. The faces were mounted on sheets and photographed in a common size (approximately 3.5 cm in diameter), and individual photographs were mounted on 15 x 23 cm cards for subsequent portions of the study.

In the next phase, 19 students at UCSB took part in a rating task, in order to satisfy an introductory psychology course requirement. They were shown the faces and a list of 13 target occupations and were asked, for each face, to select the occupation that the depicted individual seemed most likely, second most likely, and third most likely to hold. The same was done for the first, second, and third least likely occupations.

Results and Discussion

For each face, a score was computed that corresponded to the number of subjects who assigned it, within any of their top three choices, to the category that the authors had initially designated as its stereotypical occupation. This score will be called the individual-face stereotypy score. Each occupational category was also assigned a face stereotypy (FS) score, corresponding to the sum, over the highest scoring four faces in the category, of the number of subjects who assigned a face to its appropriate category as their first choice. (The first choice was used instead of the top three, because the former resulted in fewer ties between categories.) For the faces ultimately used, the part-whole correlation between the individual-face stereotypy scores for a category and its FS score was $r(42) = .85$ ($p < .01$), which indicates that the stereotypy of faces within a category tended to be quite uniform. Thus the FS score is used in correlations to be reported below.

The subjects' assignments also yielded a cooccurrence matrix (an off-diagonal half matrix) that showed, for each pair of categories, the number of times that both categories were used (among the top three choices) when a single face was assigned by a single subject. The entry in the matrix was divided by a normalization factor, namely, the total number of times the two category names were used across all ratings. In essence, the resulting matrix indicated intercategory similarity. The negative of a matrix value was used to measure the incongruence of a face with a category label other than its best matching one.

A second stereotypy measure was acquired in the absence of showing the faces, for the categories used later. Graduate students and faculty of the UCSB psychology department were asked to indicate, on a 1-11 scale, how good a stereotype they thought existed for each category. These ratings were averaged over the 18 respondents to provide the category stereotype (CS) score. The correlation between the CS score for a category and its FS score, as defined above, was $r(9) = .55$ ($p < .05$). The two measures are given by category in Table 1, along with similar measures from a later experiment.

The question to be addressed here was whether faces could be reliably assigned to categories, that is,

Table 1
Face and Category Stereotype Scores by Occupational Category

Occupation	FS Score Experiment 1	CS Score	FS x 100 Experiment 5
Accountant	21	5.36	4.7
Athlete	54	7.10	5.4
Farmer	26	5.97	10.0
Hairdresser	29	6.70	11.7
Judge	36	7.47	6.1
Model	24	8.50	10.2
Rock Musician	61	7.96	7.3
Salesman	21	4.29	5.0
Truck Driver	38	6.31	6.8
Undertaker	22	6.02	6.9
Watchmaker	23	3.79	6.7
Cleaning Woman			5.9
Librarian			10.9
Movie Actress			15.6
Stewardess			11.1
Police Detective			4.3

whether subjects' assignments were nonrandom. An individual-face stereotypy score indicates the number of subjects (of 19) who assigned the given face to its a priori designated category, within their top three choices. The smallest number that could do so, within the constraint that the chance probability for that number or greater be less than .01, was 10. Of the 83 faces, 60 were matched with the initially designated category by more than 10 subjects, and 11 of the 13 categories had at least 4 faces that were so matched. Note that the 83 faces were chosen only to be potentially classifiable, so that the presence of 23 faces in the pool that were classified less often than the criterial value of 10 is no surprise. The agreement on 60 faces is to be compared with the single face that would be expected to be appropriately categorized by 10 or more raters by chance. Thus, it is apparent that faces could be related to categories.

EXPERIMENT 2

Experiment 1 provided a set of faces that could be meaningfully related to occupational categories, as indicated by the above-chance agreement on which category a face fit. The departure from total agreement among subjects on faces' categories (compare the FS scores of Table 1 with the total-agreement score of 76) further suggests that faces are not entirely obvious category members.

Experiment 2 addressed whether semantically interpreting these faces, that is, relating them to matching categorical information, would affect memory. Each face was presented along with a category label, which was either congruent (i.e., it was the face's assigned category) or incongruent. It was proposed that a congruent label would lead to semantic interpretation. (While an incongruent label might induce some semantic processing, its incongruence would preclude relating

the face and category in a very meaningful way. The term "interpretation" will therefore be reserved for the congruent labeling condition.) The principal question was whether congruent category labeling would improve performance, and whether recognition would be higher for congruently labeled faces, the better the face-to-category match.

Method

Subjects. The subjects were 28 students from the introductory psychology class at UCSB who participated as one means of fulfilling a course requirement.

Stimuli. The stimuli included a "primary set" of 44 faces, the ones that were included in the 11 most reliably assigned categories of Experiment 1. These were the four most stereotypical faces from each of the 11 categories; each face had been categorized appropriately (within the top three choices) by at least 10 subjects in Experiment 1. The faces were photographed and made into slides, as were the 11 category labels. In addition to the primary set, a secondary set of faces from the original pool was used for filler items, which occupied the first and last three positions in the to-be-remembered list and occurred within the list as necessary to complete the design (see below). There were 10 filler faces and five accompanying occupational titles: pilot, policeman, carpenter, professor, and dentist.

Procedure. Subjects participated in groups of one to four. The task had three phases. During Phase 1, the subjects viewed a series of 32 face slides, each accompanied by an occupational label. A face and label occurred simultaneously, face on the right and label on the left, for 2.0 sec. The interstimulus interval, during which the next slide advanced into position, was approximately 300 msec. Timing was controlled by an electronic timer and projector-control apparatus. The subjects were told to try to associate the face and label, because it would help them to remember, and that they would receive a memory test.

The 32 faces that were presented in Phase 1 consisted of half the 44 primary faces, 2 from each category, and 10 fillers. Three fillers occurred in the first three positions of the list and three in the last positions; the remaining four were randomly placed within the list. (These four were necessary so that each of the five filler labels could be used twice, as the primary-set labels were.) Of the 22 primary faces, half were accompanied by the congruent category label and the other half by an incongruent label. Each label was used once at each level of congruence, within a list. The assignment of incongruent labels to faces was random, except for the constraint that the three pairs of categories judged most similar (having the highest entries in the cooccurrence matrix of Experiment 1: farmer/truck driver, model/hairdresser, judge/accountant) were never paired (i.e., a face judged as typical of one was never accompanied by the label of the other). The order of the 22 primary faces in the list was random.

The presentation phase was followed by a retention interval of approximately 10 min, during which subjects read a passage on statistics in psychology from an introductory text and then had a comprehension test, consisting of filling in the blanks in the same passage with words deleted.

Finally a yes/no recognition test was given. It consisted of 54 faces, half of which had been seen in Phase 1 and half of which were new. All 44 primary-set stimuli were included, five of the old filler faces from Phase 1 were included, and there were five new filler faces. The 54 faces were randomly ordered, and no category labels appeared. Each test face appeared in isolation. While it was present, subjects were given 10 sec to mark a 10-point scale that read from "very sure" it was new (1) to "very sure" it was old (10), with a mark at midpoint to separate "old" responses from "new." No subject failed to respond within the 10-sec period, and when all were ready, the next face was exposed.

Counterbalancing procedures ensured that across subjects, each primary-set face was accompanied equally often at presentation by a congruent and an incongruent label and that each face occurred equally often as an old and a new item.

Results and Discussion

The principal data were accuracy scores, weighted by confidence. For responses to old items, the weighted accuracy was simply the mean rating (across items in a condition) on the 1-10 response scale. For responses to new items, the weighted accuracy was the mean rating subtracted from 11, which inverts the scale. Thus for both old and new items, weighted accuracy can vary from 1 to 10, with a high score meaning strong confidence in the correct response and a low score corresponding to strong confidence in the incorrect response.

In addition, the response scale could be split at the midpoint to yield a partition of responses into correct and incorrect responses to old items (rating of ≥ 6 or ≤ 5 , respectively) and correct and incorrect responses to new items (ratings of ≤ 5 or ≥ 6). In general, the patterns of results for percent correct scores, computed in this fashion, and weighted accuracy scores were very similar, and only analyses of the latter will be reported. However, the percent correct scores can be used to compute d' s, which will be of importance in further discussion.

The weighted accuracy for the three types of test item (old, presented with a congruent label; old, presented with an incongruent label; and new) were 7.96, 7.18, and 7.03, respectively. The corresponding percents correct were 79.3, 69.7, and 72.0. An analysis of variance on the test-item factor, with subjects a random factor, revealed significant differences among the three test-item conditions [$F(2,54) = 8.72, p < .01$]. The two types of old items differed significantly [$F(1,27) = 14.70, p < .01$], with congruently labeled items having higher weighted accuracy scores. This is as predicted by the hypothesis that congruent labeling would lead to semantic interpretation and would thus facilitate performance.

A comparison of weighted accuracy on old items that were labeled congruently vs. incongruently was performed, this time with occupation rather than subjects treated as randomly sampled, and it was also significant [$t(10) = 1.84, p < .05$]. (Here, the occupation corresponding to an incongruently labeled item was the face's stereotyped category of assignment in Experiment 1, rather than the category of its label at presentation.) This indicates that the superiority of the congruent-label condition generalized over the occupations.

The question of whether incongruent labeling affected performance is addressed by the correlation, for incongruently labeled faces, between a label's incongruence (as measured from Experiment 1) and weighted accuracy. The correlation was small but significant [$r(42) = -.27, p < .05$], indicating that the more inappropriate a given label, the worse was performance on the face. (Entered into this correlation were mean accuracy scores for 44 unique face-label pairs.)

Also of interest was the correlation between weighted accuracy of recognition to faces in a category and the stereotypy of that category. Recall that there are two stereotypy measures for each occupational category. One is from the classification data of Experiment 1, indicating the assignability of faces to the category. It has been called the FS score. The second is the rating of category stereotypy in the absence of particular faces; it is the CS score. Weighted accuracy was found to correlate positively with both stereotypy measures, for both congruently labeled faces [for FS, $r(9) = .49$, $p < .10$; for CS, $r(9) = .68$, $p < .05$] and incongruently labeled faces [for FS, $r(9) = .68$, $p < .05$; for CS, $r(9) = .39$, $p < .15$]. The latter correlations suggest that the positive relationship between stereotypy and recognition was not dependent on how the face was labeled, but on properties of the face per se. In support of this hypothesis, there was also a positive correlation between stereotypy and weighted accuracy for distractor items (never presented before with any label) on the recognition test [$r(9) = .24$ for FS, $.61$ for CS].

To summarize, Experiment 2 found that interpreting a face with respect to an occupational category aided performance on a recognition test. Congruent occupational labeling led to higher accuracy than did incongruent labeling. Also, the degree of incongruence of a label correlated negatively with performance. It was also found that accuracy was positively related to facial stereotypy. However, the latter finding appears to be independent from labeling effects.

EXPERIMENT 3

Experiment 2 indicates that labeling faces with congruent occupations improves their recognizability. We now turn to the question of whether the labeling effect reflects an enhancement of the physical code of the face, due to interpreting it in relation to the occupational stereotype. One finding from Experiment 2 is somewhat at odds with this hypothesis: Whereas the physical code hypothesis suggests that face-to-category fit (as measured by FS and CS scores) would correlate with recognition when faces are given congruent category labels, it was found that this correlation held even when faces were given incongruent labels. Apparently, the correlation reflects a relationship between stereotypability of faces and their physical properties, such that stereotyped faces are more memorable regardless of how they are labeled.

In Experiment 3, an alternative to the physical code hypothesis was addressed, namely, the semantic code hypothesis. This states that the effect of congruent occupational labeling is to create a relatively abstract, semantic code for a face. The use of more abstract codes might lead to false recognitions of new faces that matched the old faces semantically. This possibility could not be assessed in Experiment 2, since it requires

that separate false alarm rates be measured in congruent- and incongruent-label conditions. Experiment 3 did so: One group of subjects was shown faces accompanied by congruent occupational labels; another group saw faces accompanied by incongruent occupational labels.

In addition to the faces shown with occupational labels, subjects in each group were shown some faces accompanied by personal names. This manipulation was meant to provide a baseline recognition rate with which the effects of occupational labels could be compared. Being arbitrary, personal names offer only minimally meaningful categories into which faces can be assigned. If even incongruent occupational labels provide a more useful mechanism for encoding, performance on faces so labeled should be above the personal-name baseline. This would indicate that the difference between the congruent and incongruent occupational labeling conditions does not reflect a disadvantage for incongruently labeled faces, but an advantage for congruently labeled faces.

Method

The subjects were 42 students at UCSB who either received introductory psychology course credit or were paid \$3 for their participation. They were assigned to one of two groups according to their order of appearance. An additional subject was eliminated because, despite instructions to use the entire rating scale, she called every test item "new."

The stimuli were those used in Experiment 2, with the addition of 11 personal names, which replaced either the congruent or incongruent occupational labels of Experiment 2, and filler personal names, which replaced half the filler occupations from Experiment 2. The personal names were common informal names of two to five letters in length, such as Bill, Bob, and Ed.

The general procedure was identical to that of Experiment 2, except that personal names replaced the congruent occupational labels for the incongruent-label group and the incongruent labels for the congruent-label group. Thus the former saw a randomly ordered mixture of 11 faces with incongruent occupational labels and 11 with personal names, preceded and followed by 3 faces with filler labels or names, and with 4 filler items within the list (half personally named, half occupationally labeled). The latter group had congruent instead of incongruent occupational labels but was otherwise identical. To eliminate potential effects of differential reading times for the personal names and occupations, each label was presented for 2 sec before the corresponding face, which appeared for 2 sec. The interstimulus time, during which the slide changed, was again about 300 msec. The retention interval and test were identical to those of Experiment 2.

Results

The weighted accuracy and percent correct, averaged over subjects, are shown in Table 2 for each group (subjects given congruent occupational labels and personal names, and subjects given incongruent occupational labels and personal names—called the congruent and incongruent groups, respectively) and test-item type (old, initially presented with occupational label; old, presented with personal name; new).

An analysis of variance on weighted accuracy with group and test type as fixed factors and subjects as a

Table 2
Recognition Scores by Group and Test-Item Type
in Experiment 3

Test-Item Type	Recognition	
	Weighted Accuracy	Percent Correct
Congruent Group		
Old, Occupational Label	7.70	76.0
Old, Personal Name	6.94	69.0
New	6.97	70.4
Incongruent Group		
Old, Occupational Label	7.05	67.6
Old, Personal Name	6.21	57.6
New	7.52	76.5

random factor showed significant effects of test type [$F(2,80) = 5.48$, $p < .01$] and a Test Type by Group interaction [$F(2,80) = 3.84$, $p < .05$]. These reflect the fact that the two groups differed on the relative ordering of test types. Although in both groups personal names led to poorer performance than did occupational labels [$F(1,40) = 13.51$, $p < .01$], the new test items were relatively poorly recognized by the congruent group and the old test items were poorly recognized by the incongruent group.

Given the apparent tradeoff between the groups on performance with old and new test items, a signal detection analysis was indicated. The hit rates for occupationally labeled and personally named old items were averaged for each group and combined with the false alarm rate to yield measures for memory sensitivity (d') and response criterion (β) for each group.² The average d' s of 1.20 and 1.21 for the congruent and incongruent group, respectively, essentially did not differ. However, the β values of 1.01 and 1.78 did differ [$t(40) = 2.75$, $p < .05$, two-tailed], indicating that subjects in the congruent group had a lower criterion for calling an item old.

An analysis of variance with occupations treated as a random factor and test type (including old items only, for which the occupational category of a face was its true categorical label, regardless of the label given it) and group as fixed factors was performed, to determine whether the old-item effects generalized across occupations. The differences between occupational labels and personal names were confirmed [$F(1,10) = 15.91$, $p < .01$], and the group effect was significant [$F(1,10) = 9.15$, $p < .05$].

A correlation was conducted for the old items in the incongruent group that were presented with occupational labels, to assess the relation between the incongruence of the label, given a face (see Experiment 2 for definition) and confidence that the face was old. Although this correlation was not significant [$r(42) = -.11$], its direction was as observed before.

Also as before, correlations were conducted between recognition performance and the measures of facial

stereotypy. Given the evidence for differences between groups in their use of the response scale (see the signal detection analysis above), the measure of recognition performance that was used was d' , rather than weighted accuracy. (In general, both weighted accuracy and d' were positively correlated with stereotypy.) The correlation across categories between d' and FS, and between d' and CS, are shown in Table 3, which shows corresponding correlations for all experiments reported here. As in Experiment 2, the positive relationship between stereotypy and recognition appeared to hold regardless of the type of label subjects were given.

Discussion

Experiment 3 addressed the locus of the initially observed advantage for recognizing faces that had been presented with a congruent occupational category label. The results indicate that the answer to this question is not provided by the physical code hypothesis. When congruence effects on old and new items were both ascertained, it was apparent that the group given congruent labels did not have an overall memory advantage, as measured by the d' measure of "true" memory sensitivity.

Instead, groups given congruent and incongruent occupational labels for faces at presentation differed on the criterion measure of the signal detection analysis. In theory, there is a criterion value of familiarity, which subjects compare with the apparent familiarity of test items, responding "old" to items that have an above-criterion familiarity value. A difference between groups in criterion is often referred to as a difference in "response bias." However, in the present case, a difference in criterion could result from the congruent group's having stronger semantic codes.

The reasoning is as follows. First, the semantic code hypothesis assumes that the presentation of a congruent occupational label with a face leads to the formation of a corresponding semantic code in memory. The

Table 3
Correlations Between d' and Category Stereotypy,
by Experiment and Condition

Experiment	Condition	Stereotypy Measure	
		FS	CS
2		.62	.56
3	Congruent Group	.70	.61
3	Incongruent Group	.52	.55
4	No-Reinstatement Group	.75	.53
4	Reinstatement Group, Congruent Labels	.17	.64
4	Reinstatement Group, Incongruent Labels	.31	.25

Note—In Experiments 2 and 4, no-reinstatement group, d' was based on the average of the congruently and incongruently labeled old items; in Experiment 3, d' was based on the average of the occupationally labeled and personally named old items. For all r s, $df = 9$; the critical value for .05 significance = .52 (one-tailed test).

presentation of an incongruent occupational label may also lead to the formation of a semantic code for the corresponding face. (The modest, but consistent, negative correlation between the incongruence of an occupational label and performance would suggest that this semantic code is less facilitative, the more incongruent the label.) However, the semantic code of a congruently labeled face would have greater strength than the semantic code of an incongruently labeled face, where strength corresponds to the quantity or quality of memorial information that a code conveys.

Next, we assume that when a face appears on a recognition test, it activates memorial information, with the decision reflecting how much is activated and how well it matches the test face. These factors determine the face's apparent familiarity. Usually, the activated information may be a physical code, but if an occupational label has led to a semantic code's being formed, this too may be activated by the test face (either directly or through an association with a similar physical code), and the decision will then reflect the combined familiarities of the physical and semantic codes. In the present task, new items as well as old items will match the semantic codes. If subjects in the congruent group have generally stronger semantic codes, all test items will match more information in memory and thus will have greater familiarity than test items have for incongruent subjects. The result is a response-criterion effect: Even if both groups require the same amount of familiarity for an "old" response, this amount will be lower, relative to the distributions of test-item familiarities, in the congruent group. Subjects in the congruent group will call more items old, and, assuming that their ratings on the response scale directly reflect item familiarity, they will also give test items generally higher ratings, as was observed.

The one rather surprising aspect of Experiment 3's results is that the congruent group's advantage at detecting old items held not only for those given occupational labels, but also for those given personal names. This result suggests that the foregoing analysis applies not only to new items, but also to the old items encoded with personal names. It is reasonable, under the semantic code hypothesis, to suppose that new items will activate semantic codes formed to old items, on the basis of their resemblance to the old items. But old items encoded with personal names might be expected, *a priori*, to activate the semantic code for the personal name. In this case, the congruent group and incongruent group should have given equivalent responses to the personally named items, but this did not occur. Apparently, personally named old items acted like new items on the recognition test, activating semantic codes for occupations. And since the congruent group had stronger semantic codes for old test items, they gave higher oldness ratings even to the personally named items than the incongruent group did.

In short, the present results seem to be best explained by the hypothesis that interpreting faces with respect to occupational categories enhanced the formation of semantic codes, rather than physical codes. A semantic code, being more abstract, would tend to match test faces that had not been seen initially but that were congruent with the occupational categories of initially seen faces. This would lead to false recognition errors, as well as higher recognition of old items.

This semantic code hypothesis accounts for differences between groups. Differences in response accuracy among the three types of items within each group—occupationally labeled, personally named, and new—can be attributed to differences in the strengths of the physical codes. If we speculate that occupationally labeled items have stronger physical codes than personally named items have (which seems consistent with the levels-of-processing studies described previously) and that new items on the test have essentially no physical code in memory, the first type of item would get the maximum oldness ratings and the last type, the minimum.

EXPERIMENT 4

According to the semantic code hypothesis, the decision as to whether a face has been seen before relies in part on the match between semantic codes retrieved from memory and semantic attributes of the test item. Experiment 4 further tested this hypothesis by manipulating the decision process. Subjects who had initially encoded some faces along with congruent labels and others with incongruent labels were tested with faces that were accompanied by a congruent label, an incongruent label, or no label. Subjects who saw labels at test were told that if a test face was old, its label was the same as the original. The label at test was intended to explicitly provide semantic information about the test face, which could be compared with semantic codes retrieved from memory.

Consider first the effect of this manipulation on old items. These are assumed to be stored in memory as associated physical and semantic codes, and the semantic codes of congruently labeled faces are assumed to be stronger than those of incongruently labeled faces. To the extent that reinstating the label of an old item at test provides semantic information that can be matched with the face's previously formed semantic code, reinstatement should increase the face's apparent familiarity, thus improving performance. This will occur regardless of the label's congruence, but because the semantic code is stronger for congruently labeled items, the congruence effect will not be eliminated when the label is reinstated. (If, instead, the congruence effect were due to poorer retrievability of semantic codes for incongruently labeled items, it should be eliminated when the previous label is reinstated on the test.)

Next, consider the effect of labeling on new items. As was proposed to account for the results of Experiment 3, a new item is assumed to activate the physical codes of physically similar (i.e., same-category) old items in memory, which leads in turn to the activation of associated semantic codes. These codes will be compared with semantic information provided by the test face's label (if any). The semantic codes of congruently labeled old items, being stronger than those of incongruently labeled old items, will predominate in this comparison.

An unlabeled new face will match the semantic code for a congruently labeled old item to some extent, as Experiment 3 indicates. But if a new face is explicitly labeled with a congruent occupation, the match with the semantic code should be even greater. This will increase the apparent familiarity of the test face and will thus deter recognition that it is new. If a new face on the test is incongruently labeled, in contrast, the semantic information in the label will mismatch the strong semantic code for a congruently labeled, physically similar old item. And since incongruent labels for new items were randomly assigned, the label will also mismatch the weaker semantic code associated with a physically similar, incongruently labeled old item. Thus, the apparent familiarity of an incongruently labeled new item will not be increased by matches between the label and previously stored semantic codes, and detection of the item's newness will not be impaired. As a result of these effects, comparisons between unlabeled and labeled new items should reveal a disadvantage for detecting that congruently labeled faces are new and an advantage for incongruently labeled new faces.

Method

The subjects were 32 students who participated to fulfill an introductory psychology course requirement or were paid \$3 for their participation.

The materials and general method were identical to those of Experiment 2, with the following changes. Subjects were assigned to one of two groups, in alternation according to order of appearance. One group, to be called the no-reinstatement group, was a replication of Experiment 2. The other, the reinstatement group, had a procedure identical to that of Experiment 2, except that at the time of test, each face was presented along with an occupational label. If the face was old, the label was identical to that presented at the initial viewing. For new faces used as distractors on the test, half were given their congruent label and half, the label of another category (an incongruent label), selected at random. All labels were used once for a congruent new face and once for an incongruent new face. Subjects in this group were informed of the fact that old test items would always be accompanied by the original label.

Results and Discussion

The mean weighted accuracy and percent correct responses, averaged over subjects and occupations, are shown in Table 4. An initial analysis of variance was conducted on weighted accuracy, with subjects treated as random and three fixed factors: test-label presence

Table 4
Recognition Scores by Group and Test-Item Type
in Experiment 4

Test-Item Type	Recognition	
	Weighted Accuracy	Percent Correct
No-Reinstatement Group		
Old, Congruent Label	7.48	74.7
Old, Incongruent Label	7.03	70.1
New, Yoked to Congruent	7.02	72.9
New, Yoked to Incongruent	6.97	72.9
Reinstatement Group		
Old, Congruent Label	8.19	83.7
Old, Incongruent Label	7.64	76.9
New, Congruent Label	6.39	71.2
New, Incongruent Label	7.70	80.2

(whether or not labels were reinstated at test), response (the correct response to a test item: old or new), and congruence (whether the label shown at presentation, or, for the reinstatement group, at test, was congruent or incongruent with the face). For the no-reinstatement subjects, new items were assigned to a congruence level that matched the level for yoked new items (same face, same sequence) of the reinstatement subjects, although the no-reinstatement subjects actually saw no labels for new items.

The semantic information provided by a label that accompanied test items was found to affect recognition. There was a significant three-way interaction among test-label presence, response, and congruence [$F(1,30) = 5.99, p < .05$] and a two-way Response by Congruence interaction [$F(1,30) = 14.38, p < .01$]. These reflect the facts that the congruence effects on old and new items were in opposite directions and that the effect held only for those new items that were given labels (the reinstatement group).

Supplementary analyses examined each response separately. For old test items, the effect of congruence was significant [$F(1,30) = 4.28, p < .05$] and that of test-label presence was marginally so [$F(1,30) = 3.60, p < .075$]. The disadvantage for incongruently labeled items was not reduced if the label was reinstated at test (for the interaction, $F < 1$). When the same analysis was performed with occupations rather than subjects as the random factor, results were similar [for congruence, $F(1,10) = 5.62, p < .05$; for test label, $F(1,10) = 7.29, p < .05$; for the interaction, $F < 1$].

For new test items, the congruence effect was significant [$F(1,30) = 10.77, p < .01$] and interacted as expected with test-label presence [$F(1,30) = 12.47, p < .01$]. To determine the effects of giving new items congruent and incongruent labels, accuracies for these items were compared with scores for the corresponding items in the unlabeled condition. A congruent label significantly impaired new-item detection, relative to no label, and an incongruent label significantly facilitated

it [$t(30) = 1.77$ and 1.90 , $ps < .05$]. However, the percent correct data show only a minimal effect of giving new items a congruent label. The difference between unlabeled and congruently labeled new items may have been mitigated somewhat by the strength of the unlabeled items' spontaneous evocation of congruent categories.

A signal detection analysis was conducted on the reinstatement condition to separate memory and decision effects of congruence. The mean d' 's, measuring memory sensitivity, essentially did not differ for congruently and incongruently labeled items (values were 1.94 and 1.93). However, the β values for the congruent and incongruent conditions were 1.37 and 2.39 , which differed significantly (by sign test, $p < .05$). This is consistent with the idea that all items that were accompanied by a congruent label at test appeared more familiar.

The correlation between the incongruence of a label given to a face at presentation and the weighted accuracy of its recognition as an old item was $r(42) = -.25$ ($p < .075$) in the no-reinstatement group. This is nearly identical to the corresponding correlation in Experiment 2. Reinstating the initial incongruent label at test eliminated this correlation (for reinstatement subjects, the r was $-.02$).

Again, there was a correlation between facial stereotypy (as measured by FS and CS) and recognition (as measured by d') (see Table 3). These correlations were similar to those found previously for the no-reinstatement group, but for subjects who saw labels at test, they were considerably weaker. The congruence of the label again appears not to be critical to the values of these correlations.

To summarize, the results of Experiment 4 were generally consistent with the semantic code hypothesis, which assumes that semantically interpreting faces with respect to occupational categories affects recognition by promoting the formation of a semantic code. The basic finding is that providing semantic information at the time of the recognition decision affected recall. When a face was tested along with a congruent category label, which is assumed to increase the match with the strong semantic codes previously formed for congruently labeled faces, its judged familiarity seemed to increase. The increase in familiarity was indicated by the label's facilitating the recognition of old items and impairing the detection of new items, and it also emerged in the decision-criterion measure from the signal detection analysis. Accompanying a face with an incongruent label facilitated performance with both old and new items. This was predicted on the basis that incongruently labeled test faces should match stored semantic codes when they were old and minimally match stored semantic codes when they were new. It is noteworthy that reinstating the label of an old item, when it was incongruent, did not raise performance to the level of the

congruently labeled items. This is consistent with the assumption that the semantic codes of congruently labeled items are stronger.

Although they were generally as predicted, the effects of labeling at test were not terribly strong in this experiment. It may be that a great deal of occupational-category information was available from test faces even without the label, which rendered the label only marginally necessary. Nevertheless, the experiment clearly supports the idea that semantic matches and mismatches enter into decisions about semantically interpreted faces.

EXPERIMENT 5

In Experiments 3 and 4, semantic interpretation of a face with respect to a stereotyped occupational category appears to have facilitated the formation of a semantic code. No evidence has been found that labeling facilitates the encoding of physical information from the face. Conceivably, however, the physical code effect would be evident if the recognition test were more sensitive, specifically, if it relied on memory for physical details, making semantic information clearly insufficient for the recognition decision.

To investigate this possibility, Experiment 5 used a stimulus pool devised from a set of features called the *Identi-Kit*.³ The pool contained 32 distinct faces, half of which were occupationally stereotypable and half of which were not, plus a physically similar variant of each face, devised by altering a subset of its features. Subjects saw one version of each of the 32 distinct faces, accompanied by either a congruent occupational label or a personal name. They subsequently received a forced-choice recognition test, which required that each previously seen face be discriminated from its similar variant. If semantic interpretation facilitates the formation of a physical code for a face, category labels should improve performance on this type of test, but for the stereotypable faces only. The condition in which faces receive personal-name labels provides a control for any general advantage for recognizing stereotypable faces. If interpretation aids physical coding, the stereotypable faces' advantage in the occupational-label condition should be greater than in the personal-name condition.

Method

Subjects. The subjects were 36 students at UCSB.

Stimuli. The stimuli were selected in two phases. In Phase 1, 19 UCSB students (not those who served in the experiment proper) assigned each of 127 faces either to the 1 of 16 occupational categories (specified for sex) that it best fit or to a no-category response if it seemed not to fit any. The 127 faces were constructed from features of the *Identi-Kit* (augmented with a few artists' drawings). They included candidates for stereotypable (S) faces—from 5 to 8 variants (differing on one or more features) for each of the 16 categories—and nonstereotypable (N) faces—28 faces with the same number and types of

features as S faces. The occupational categories are shown in Table 1.

From the Phase 1 assignments, three faces were selected for Phase 2 from each category: the face most often assigned to that category, another highly assigned variant of that face, and a third variant with a substantially lower number of assignments. For two categories, faces from the N set had been assigned most often, so two new variants of those were constructed, one similar to the original and the other quite different. Sixteen new students ranked the three faces from each category with respect to their degree of stereotypy.

For use in the present study, the two top-ranked faces were selected from each category. They were very close in mean ranking and in number of assignments. Sixteen N faces were also selected, and a variant of each was constructed. The final pool consisted of 32 pairs of faces: two versions of a face that could be assigned to a category, for each of the 16 categories (S faces), and two versions of each of 16 minimally assignable (N) faces. Each S pair was matched with an N pair for apparent sex, number, and type of features in the two faces, and the difference between the faces with respect to number and type of differing features. Thus, for example, if S1 had glasses and S2 did not, N1 had glasses and N2 did not.

In the final stimulus pool, the top-ranked face in a category was appropriately assigned by from 5 to 15 of 19 subjects (mean = 9.2) in Phase 1. (The probability of five or more subjects' assigning a face to a category by random responding is less than .01) The top-ranked faces in the S categories were compared with the corresponding Ns on three measures from the assignment task: First, there was an equivalent to the FS score used in Experiments 2-4. This was the number of subjects who assigned the face to its category of highest assignment, divided by the total number of times any face was assigned to that category. (This normalizing was performed to correct for the asymmetry in the number of potential categories for male and female faces.) The FS scores (shown in Table 1) were significantly higher for the S faces than for the Ns [$t(30) = 4.83$, $p < .01$]. Second, the number of different categories to which faces were assigned was significantly less for Ss than for Ns [$t(30) = 1.97$, $p < .05$]. Third, the number of assignments to the no-category slot was significantly less for the S faces [$t(30) = 3.01$, $p < .01$]. (For the "farmer" category, the compared values were for the second-ranked face, since the top-ranked face was added between Phases 1 and 2.)

Finally, 18 subjects were shown the pairs of faces side by side and asked to rate them for similarity on a 1-8 scale. The mean value was 5.10 for the S pairs and 5.15 for the Ns, a nonsignificant difference. Thus, the S and N sets had equivalent similarity between pairs.

Procedure. The subjects were assigned to one of two groups, alternating from one session to another. One group saw faces accompanied by the corresponding label: the congruent occupation for S faces or that of the yoked S face for Ns. The second group saw faces accompanied by personal names, as in Experiment 3.

The experiment had three phases like those of the previous studies, except that the recognition test was a two-alternative forced-choice test rather than a yes/no test. In Phase 1, the subject saw 38 faces for 2 sec each, each preceded by a label (occupation or name) for 2 sec. The first and last three faces were fillers; the remainder were S faces and the yoked N faces, in random order. Within each subject's list, half the S faces were the top-ranked members of the category and half were the second-ranked members. Across subjects, all faces were used equally often at presentation. The retention period was identical to previous studies. At the test, the subject saw the 32 critical pairs of faces, side by side, in a random sequence. The subject indicated whether the face on the left or the right was the one originally seen and gave a confidence rating, from "very sure" to "very unsure" on a 5-point scale.

The experiment was run on a PDP-11/34 computer, capable of monitoring three subjects simultaneously. All subjects saw a common slide display from three adjacent booths, at a distance of approximately 80 in. Slides were presented from a rear-screen projector at eye level, one to the left of the center of the display (this was the label slide during the presentation phase) and one to the right. In each subject's booth was a CRT and typewriter keyboard. The top row of the keyboard was marked with the response scale, which included five degrees of confidence for each choice, left or right. The subjects had 30 sec to respond.

Results and Discussion

The data were percent correct choices of the originally presented face and weighted accuracy. The latter was the confidence (from 1 to 5) for correct responses, minus the confidence for incorrect responses, averaged across items within a condition. These data are shown in Table 5. Both measures showed the same effects, and only the weighted accuracy analyses are reported.

Preliminary analysis indicated no effect of whether the top- or second-ranked face in a category was presented initially, so data were averaged over that factor, and an analysis of variance was conducted on the fixed factors label type (occupational title vs. personal name) and face type (S vs. N). With subjects as a random factor, there was a significant main effect of face type, with the S faces being better recognized [$F(1,34) = 10.62$, $p < .01$]. The effect of label and the interaction were nonsignificant (for the latter, $F < 1$). Similar results were found when data were averaged over subjects and occupational category was treated as a random factor [for Face Type, $F(1,15) = 6.94$, $p < .05$; for the interaction, $F < 1$]. Thus, the advantage for S faces generalizes over categories.

These analyses offer little support for the physical code hypothesis, which predicted that an occupational title should improve memory for the stereotypable (S) faces. Although there was an advantage for the stereotyped faces, it occurred regardless of the label that was given. It might be argued that the weakness of the labeling effect precluded any interaction, but the small effect of the label is quite consistent with the idea that how faces are interpreted has little effect on the memorial information used in the present test, which requires close physical discriminations. (And, in contrast, pilot testing with more discriminable faces did show significant labeling effects.)

Table 5
Recognition in Experiment 5 as a Function
of Face Type and Label

	Stereotypable Faces		Nonstereotypable Faces	
	Percent Correct	Weighted Accuracy	Percent Correct	Weighted Accuracy
Personal Name	80.0	2.56	75.3	2.10
Category Label	83.7	2.87	77.4	2.42

The observed advantage for recognizing stereotypable faces, regardless of labeling, suggests that they are more recognizable than nonstereotypable faces on some physical basis. Also supporting this assumption are the correlations between stereotypy and recognition. As before, these were positive. The correlation between FS and weighted accuracy was $r(14) = .61$ in the occupational-label condition and .46 in the personal-name condition ($ps < .01$ and $.05$).

GENERAL DISCUSSION

In summary form, the principal findings of this series of studies are the following. (1) Occupational categories provide sufficient information about facial appearance to allow for reliable assignments of faces to categories. (2) Labeling a face with a congruent occupational-category name, at the time the face is presented, increases subsequent judgments that the face was seen before, although not in a forced-choice test that requires fine physical discriminations. (3) The increase in recognition of congruently labeled old items is balanced by a disadvantage in detecting that categorically similar faces are new, especially if the category is explicitly stated at the time of test. In contrast, the incongruence of a test face and a category label may increase the detectability that the face is new. (4) More stereotypable faces are better recognized, regardless of how they are labeled.

Together, the first three of these findings support the semantic code hypothesis, which states that interpreting a face with respect to an occupational category provides a code that represents category information, rather than enhancing the encoding of the face on a physical basis. The semantic code that is provided by a congruent categorical interpretation facilitates recognition that a face has been seen before, by adding to encoded information in memory. But, at the same time, it impairs the ability to detect that other, semantically matching, faces are new.

This phenomenon of semantically based recognition, following semantic interpretation, has been found to emerge in a signal detection analysis as an effect on the decision-criterion measure. This measure is often viewed as indicating response bias. However, in the present situation, the criterion effect seems more appropriately viewed as reflecting a heightened familiarity for all category members, following categorical interpretation of a face. As was previously explained, the heightening of familiarity for new as well as old items from a category would lower the criterion measure, even if the absolute familiarity required for a positive recognition decision were unchanged by semantic interpretation, because the criterion is measured relative to the familiarity distributions of the interpreted items.

The fourth finding described above, better recognizability for categorizable faces, is evidenced by strong

correlations between recognition accuracy and facial stereotypy measures, whether derived for particular faces or the category in general. Just what makes stereotyped faces more memorable remains unclear. However, Experiment 5 rules out an explanation in terms of featural complexity, since stereotyped and nonstereotyped faces in that experiment were matched for the number and types of features. The features of stereotyped faces may still be more distinctive, either absolutely so or relative to others in the face (cf. Winograd, 1981). In any case, this correlation appears to be unrelated to semantic interpretation of faces at encoding.

The present effects of stereotypy contrast with those of Light, Kayra-Stuart, and Hollander (1979), who found that among a collection of faces from a single category (male high school seniors), atypical rather than stereotypical faces were remembered better. Typicality effects may have been masked by competing effects that arose because of the use of such a homogeneous stimulus pool in Light et al.'s experiment. In this case, atypical faces could gain an advantage because of discriminability, which would outweigh the advantage of stereotypicality. If a stimulus pool could be found that equated the discriminability of typical and atypical students, an advantage for stereotypicality might in fact be observed.

The finding of categorically induced errors in face recognition is of obvious interest from an applied standpoint. Misidentification of persons, by virtue of their matching remembered faces in terms of race or ethnic appearance, is not uncommon. The present results suggest that similar errors could arise from categorizing faces at less obvious levels. Whether they are explicitly labeled or not, facial stereotypes that are encoded along with the physical features of a particular face could be the basis for false recognitions on later tests. On the more positive side, the present studies indicate that there are physiognomic factors that make certain faces more memorable than others. Further research might determine the nature of these factors, with the ultimate goal of improving face identification techniques.

REFERENCES

- BADDELEY, A. Applied cognitive and cognitive applied psychology: The case of face recognition. In L.-G. Nilsson (Ed.), *Perspectives on memory research*. Hillsdale, N.J.: Erlbaum, 1979.
- BARTLETT, J. C., TILL, R. E., & LEVY, J. C. Retrieval characteristics of complex pictures: Effects of verbal encoding. *Journal of Verbal Learning and Verbal Behavior*, 1980, 19, 430-449.
- 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.
- Dictionary of Occupational Titles* (4th ed.). Washington, D.C.: U.S. Department of Labor, 1977.
- DANIEL, T. C. Nature of the effect of verbal labels on recognition memory for form. *Journal of Experimental Psychology*, 1972, 96, 152-157.

- DOOLING, D. J., & LACHMAN, R. Effects of comprehension on retention of prose. *Journal of Experimental Psychology*, 1971, **88**, 216-222.
- FRANKS, J. J., & BRANSFORD, J. D. Abstraction of visual patterns. *Journal of Experimental Psychology*, 1971, **90**, 65-74.
- FRIEDMAN, A. Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, 1979, **108**, 316-355.
- GOODMAN, G. S. Picture memory: How the action schema affects retention. *Cognitive Psychology*, 1980, **12**, 473-495.
- HOCHHAUS, L. A table for the calculation of d' and β . *Psychological Bulletin*, 1972, **77**, 375-376.
- KLATZKY, R. L. Visual memory: Definitions and functions. In R. Wyer, T. Srull, & J. Hartwick (Eds.), *Handbook of social cognition*. Hillsdale, N.J.: Erlbaum, in press.
- KLATZKY, R. L., MARTIN, G. L., & KANE, R. A. Influence of social-category activation on processing of visual information. *Social Cognition*, in press.
- KLATZKY, R. L., & STOY, A. M. Semantic information and visual information processing. In J. W. Cotton & R. L. Klatzky (Eds.), *Semantic factors in cognition*. Hillsdale, N.J.: Erlbaum, 1978.
- LIGHT, L. L., KAYRA-STUART, F., & HOLLANDER, S. Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory*, 1979, **5**, 212-228.
- LITTERER, O. Stereotypes. *Journal of Social Psychology*, 1933, **4**, 59-68.
- MALPASS, R. S., LAVIGUER, H., & WELDON, D. E. Verbal and visual training in face recognition. *Perception & Psychophysics*, 1973, **14**, 285-292.
- MANDLER, J. M., & RITCHEY, G. H. Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, **3**, 386-396.
- NELSON, D. L., REED, V. S., & WALLING, J. R. Pictorial superiority effect. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, **2**, 523-538.
- PAIVIO, A. *Imagery and verbal processes*. New York: Holt, Rinehart, & Winston, 1971.
- PATTERSON, K. E., & BADDELEY, A. D. When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, **3**, 406-417.
- RAFANEL, K. J., & KLATZKY, R. L. Meaningful-interpretation effects on codes of nonsense pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, **4**, 631-646.
- SERGEANT, J., & BINDRA, D. Differential hemispheric processing of faces: Methodological considerations and reinterpretation. *Psychological Bulletin*, 1981, **89**, 541-554.
- SHOEMAKER, D., SOUTH, D., & LOWE, J. Facial stereotypes of deviants and judgments of guilt or innocence. *Social Forces*, 1973, **51**, 427-433.
- WINOGRAD, E. Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology: Human Learning and Memory*, 1981, **7**, 181-190.

NOTES

1. The present concern is with the facilitative effects of interpreting a visual stimulus with respect to a conceptual category into which it might not otherwise be assigned. There is another type of categorization effect on visual stimuli, which should be distinguished from the effect of principal interest here. If a stimulus is readily categorized (e.g., as a "kitchen scene"), category-inconsistent components appear to receive more attention at encoding and thus are better recognized (Friedman, 1979; Goodman, 1980). Note that this effect is on features that are inconsistent with the category, rather than consistent, as in the effect of current interest.

2. To calculate d' and β , the method of Hochhaus (1972) was used, with values of .99 or .01 substituted for hit rates of 1.0 or false alarm rates of 0.

3. We thank the Identi-Kit Company for the use of their materials.

(Received for publication November 9, 1981;
revision accepted February 23, 1982.)