

# Typicality and familiarity of faces

JAMES C. BARTLETT, SUSAN HURRY, and WARREN THORLEY  
*University of Texas, Dallas, Texas*

**We examined context-free familiarity information as a source of the effects of face typicality upon face recognition. Experiment 1 tested memory for typical and unusual faces by (1) subjects who received an input list followed immediately by a recognition test (standard condition), (2) subjects who viewed all test faces (targets and lures) prior to the input list (prefamiliarization condition), and (3) subjects who viewed all test faces after the input list but prior to recognition (postfamiliarization condition). Although false-alarm errors in the standard condition were lower for unusual than for typical faces, this effect was reduced by postfamiliarization and was eliminated entirely by prefamiliarization. The prefamiliarization and typicality effects were replicated in Experiment 2, which showed that patterns of old judgments were compatible with the hypothesis that, although familiarity of new faces is greater if these faces are typical, the increment in familiarity that results from presentation is greater if these faces are unusual.**

Intuition suggests that "distinctive" or "unusual" faces are easier to recognize than are those that are "average" or "typical." Indeed, an effect of typicality upon face recognition has been documented in recent research (Cohen & Carr, 1975; Going & Read, 1974; Light, Kayra-Stuart, & Hollander, 1979). The effect has considerable practical importance, because typicality might influence the accuracy of identifications by eyewitnesses to crimes (Loftus, 1979; Yarmey, 1979). Moreover, the effects of typicality obviously pertain to theoretical questions of whether or in what way distinctiveness of encoding influences memory (Eysenck, 1979; Jacoby & Craik, 1979; Lockhart, Craik, & Jacoby, 1976). Unfortunately, just how the phenomenon is relevant, either to eyewitness identification or to distinctiveness effects on memory, currently is rather obscure. The problem is that there are several plausible mechanisms through which typicality might affect face recognition, and we presently lack evidence to decide which of these mechanisms are critical.

Light et al. (1979) tested the possibility that typicality effects reflect "deeper" or more "elaborate" encoding for unusual than for typical faces. In support of this depth/elaboration mechanism, two of their experiments showed a hit-rate advantage for unusual targets over typical targets in recognition, but only when the input task encouraged deep or elaborate processing (i.e., when likableness-rating or intentional-learning tasks were used). Another of their experiments showed that relatively lengthy (8-sec) exposures for input pictures can in some

cases be necessary for a hit-rate advantage to occur. These findings suggest that more unusual faces receive deeper or more elaborate processing, so long as such processing is compatible with input instructions and timing. The findings also are consistent with Winograd's (1981) idea of a feature-scanning mechanism. The hit-rate advantage for unusual faces might reflect their possession of highly distinctive features. Elaborate processing tasks and/or lengthy exposure times might increase the probability that such highly distinctive features are scanned.

It is plausible that a depth/elaboration or a feature-scanning mechanism contributes to the typicality effect on hits. However, the effects of typicality are not restricted to hits, but also are found in analyses of false-alarm errors (which are lower for unusual faces), as well as in analyses of target/lure discrimination (which is greater for unusual faces). Furthermore, Light et al. (1979) found no convincing evidence that elaborate encoding or lengthy exposures at input were necessary prerequisites for these two effects to occur.

Since effects of typicality upon false alarms and target/lure discrimination are large and highly consistent, it is critically important that they be understood. The idea that motivated the present research was that context-free "familiarity information" (Mandler, 1980) contributes to these effects. The aim of the experiments was to test a simple "familiarity hypothesis," according to which (1) entirely new faces have nonzero levels of perceived familiarity, which are greater if the faces are typical than if they are unusual (producing the effect of typicality upon false alarms), (2) previously presented faces are perceived as more familiar than entirely new faces (supporting above-chance discrimination between "targets" and "lures"), and (3) the increment in familiarity that results from a single prior presentation is greater for unusual than for typical faces (producing the effect of typicality upon target/lure discrimination).

The authors thank W. J. Dowling and R. Hastie for fruitful discussions concerning this article. Some of the data from Experiment 1 were the basis for a master's thesis by Susan Hurry, and Experiment 2 was conducted as an undergraduate honors project by Warren Thorley. Correspondence concerning this article should be sent to James Bartlett, Program in Psychology and Human Development, University of Texas at Dallas, Box 688, Richardson, TX 75080.

Although the familiarity hypothesis initially might seem nothing more than a description of the data, it gains substance from the assumption (e.g., Mandler, 1980) that whereas perceived familiarity is sensitive to the frequency of prior experience with a stimulus, it is largely independent of the context(s) of that experience. Because of this assumption, the familiarity hypothesis has an interesting and somewhat counterintuitive implication: Typicality should influence the increment in familiarity that results from a presentation of a face, whether or not this presentation is within the context of an input list. It follows that the phenomenon of reduced false-alarm rates to unusual lures should be attenuated, if not eliminated or even reversed, if subjects are exposed to recognition-test lures (i.e., if lures are familiarized) prior to recognition, but outside an input list (cf. Kinsbourne & George, 1974).

In order to evaluate the familiarity hypothesis, the experiments that follow examined recognition memory for typical and unusual faces under standard conditions, and also under conditions in which all experimental faces, targets as well as lures, were presented to subjects (familiarized) prior to recognition. The prediction was simply that (1) false alarms should be fewer to unusual lures than to typical lures under standard conditions, but that (2) this effect should be reduced, if not eliminated or even reversed, under familiarization conditions.

Our intent was the modest one of testing the plausibility of a familiarity hypothesis for typicality effects, rather than that of testing this hypothesis against alternatives. It should be noted, however, that other possible views of typicality effects would not make identical predictions. Consider in particular the hypothesis that more unusual faces are associated more easily with presentation context, that is, that memory for "circumstances of encounter" is better with unusual faces. Although knowledge of context in recognition of faces is known to be limited (Brown, Deffenbacher, & Sturgill, 1977), it obviously might be less limited if the faces employed are unusual. If so, false-alarm errors to unusual faces should be less frequent than those to typical faces, not only under standard conditions, but also under familiarization conditions.

### EXPERIMENT 1

Experiment 1 of this report used a between-subjects manipulation of familiarization to examine the interactive effects of typicality and familiarization. Free-choice recognition memory for typical and unusual faces was assessed under standard conditions, and also under conditions in which faces had been "familiarized" immediately prior to the input list (prefamiliarization) or immediately following the input list but prior to recognition (postfamiliarization). The comparison of effects of pre- versus postfamiliarization was for purposes of establishing generality, because there seemed no reason to

predict that these effects would differ. Our thinking was that if a familiarity mechanism is operative, familiarization—pre- or post—should reduce the false-alarm-rate advantage for unusual lure faces in recognition. Indeed, it might eliminate or even reverse this advantage.

Although the familiarization procedures were expected to reduce the false-alarm-rate advantage for unusual lures, they were not expected to reduce the hit-rate advantage for unusual targets. Indeed, given that a familiarized target is presented two times prior to test, and given that the familiarity increment on each presentation plausibly is greater for an unusual target than for a typical target, the opposite prediction might be made. That is, familiarization should increase the effect of typicality on hits.

Reaction times for recognition decisions were measured, because these were expected to converge with the data of hits and false alarms. For exploratory purposes, a manipulation of verbal coding at input also was employed. Although prior research (e.g., Chance & Goldstein, 1976) offered little support for effects of verbal coding upon face recognition, such effects with typical and unusual faces have not, to our knowledge, been compared. Since we did not want to confound verbal coding effects with depth-of-processing effects, we compared a verbalization input task with a "traits" input task, friendliness rating, known to produce high face-recognition performance (Winograd, 1981).

### Method

#### Materials

The 96 to-be-remembered faces and 8 filler faces were selected from the Texas A & M yearbook of 1965. All faces were head-and-shoulders photographs of young Caucasian men dressed in dark suits, white shirts, and ties. None had beards, mustaches, or glasses, and all had close-cropped hair. An attempt was made to find an equal number of typical and unusual faces, although this was based only on a single investigator's intuitions. The faces were rephotographed to produce (35-mm) black-and-white transparencies.

**Typicality ratings.** Five judges rated each face slide on a 7-point typicality scale (1 = highly typical). The averaged ratings from two of these raters correlated at  $r = .80$  with those from the remaining three raters. The faces were divided at the median (3.6) into typical (mean rating = 2.45,  $SD = .62$ ) and unusual (mean rating = 4.79,  $SD = .86$ ) categories, each including 48 items.

**Familiarity ratings.** As a preliminary test of the familiarity hypothesis, the faces were presented to a second group of raters ( $n = 7$ ), who rated each face with regard to familiarity. For the purposes of this task, a familiar face was one that a subject might have seen before somewhere (e.g., on a city street). A 6-point scale was used (1 = sure unfamiliar), and the average familiarity rating was 2.38 for typical faces and 2.06 for unusual faces. The difference was reliable with items taken as the random factor [ $t(94) = 3.01$ ,  $p < .01$ ] and also with subjects taken as random [ $t(6) = 2.94$ ,  $p < .05$ ]. The effect suggests that entirely new faces are perceived as more familiar if they are typical than if they are unusual.

**Presentation and test lists.** The 96 to-be-remembered faces were assigned randomly to two 48-item lists, each containing 24 typical and 24 unusual faces (each list was presented to one-half of the experimental subjects). The ordering of faces was

random, with the constraint that each block of 24 faces included 12 typical and 12 unusual items. Each of the lists began and ended with three filler faces.

The recognition test included all 96 faces. The ordering was random, with the constraint that each 48-item block included 6 faces in each combination of typicality, input list, and input block (first vs. second half of an input list). Which test items were targets and which were lures were determined by which input list a particular subject received. The test began with 8 "warm-up" items, 6 of which had been presented previously as input-list fillers ("warm-up targets") and 2 of which were new ("warm-up lures").

**Design**

In addition to the within-list variables of typicality and item type (targets vs. lures), the design included the between-subjects variables of familiarization condition and input task. With respect to the former, the standard-presentation group received the usual treatment of an input list followed by a recognition test. The prefamiliarization group was treated identically, except that the entire set of target and lure faces was presented immediately prior to the input list. The postfamiliarization group also was treated identically, except that the entire set of faces was presented immediately after the input list, but prior to recognition.

With respect to input task, one-half of the subjects within each familiarization group wrote verbal descriptions for each input-list face. The remaining subjects made ratings of perceived friendliness for each input-list face. There were 16 subjects in each factorial combination of familiarization condition and input task.

For counterbalancing purposes, each of the two input lists was presented to 8 of the 16 subjects within each experimental group. The consequence was that each of the 96 to-be-remembered faces served as a target for one-half of the subjects and as a lure for the remainder. Within each of the subgroups of 8 subjects, the assignment of judgment (target vs. lure) to response key (right vs. left) was varied, with 4 subjects receiving each assignment.

**Procedure**

Each subject in the standard condition was seated approximately 3 ft away from a white wall and was told that a list of faces would be presented. Instructions were given for the appropriate input task (verbalization vs. friendliness rating), but no mention was made of a forthcoming recognition test. The input-list faces were presented via slide projector for 2 sec each (visual angle = 20 deg), with an interstimulus interval of approximately 15 sec.

In the interstimulus interval after each input-list face, each verbalization-group subject wrote a verbal description, which was to be "as complete as possible, so that another person, seeing only your description, could get as accurate an idea as possible of what the face is like." Each friendliness-rating subject simply indicated, on a 6-point scale, the perceived "friendliness" of each input-list face.

Prior to the recognition test, the subjects were told that a new list of faces would be presented and that they were to judge whether each of these faces was or was not from the input list. Judgments were made by pressing one of two telegraph keys, which were wired to a Lafayette reaction time control apparatus. The subjects also stated their confidence, on a 3-point scale, in each recognition response. Each recognition-test slide was preceded by a "ready" signal, a 1-sec blank interval, and then the slide itself, shown for 5 sec. After presentation of each test slide, the experimenter recorded the keypress response, response latency, and confidence rating.

The procedure for the pre- and postfamiliarization conditions was identical to that for the standard condition, except for the presentation of the familiarization set (immediately before or immediately after the input list). The instructions for familiarization were simply to examine each familiarization slide (the experimenter was able to watch each subject, to ensure that each slide actually was viewed). Each familiarization-set picture was presented for 2 sec, with a minimal interstimulus interval (the change time of the Kodak Carousel projector).

**Subjects**

The subjects were 96 undergraduates at the University of Texas at Dallas; they participated as one alternative means of fulfilling a course requirement. The average age in this population was approximately 30 years.

**Results**

Table 1 shows probabilities of recognition (old) judgments to typical and unusual targets (hits) and to typical and unusual lures (false alarms), in each familiarization condition. The data were subjected to analysis of variance, with the between-subjects variables of familiarization condition and input task and the within-subjects variables of typicality and item type (targets vs. lures). The analysis produced reliable main effects for both familiarization condition [ $F(2,90) = 4.53, p = .01, MSe = .042$ ] and item type [ $F(1,90) = 568.0, p < .0001, MSe = .028$ ], and also supported a familiarization x item type interaction [ $F(2,90) = 7.19, p = .001, MSe = .028$ ]. It can be seen in Table 1 that familiarization had generally negative effects on recognition accuracy, clearly increasing false-alarm rates while reducing (prefamiliarization), or only slightly increasing (postfamiliarization), hit rates.

The main effect for typicality was not significant ( $F < 1$ ), because the overall probability of old judgments (hits and false alarms) was virtually identical for typical

**Table 1**  
Probabilities of "Old" Judgments to Typical and Unusual Lures (False Alarms) and to Typical and Unusual Targets (Hits) for the Three Familiarization Conditions of Experiment 1

Typicality	Item-Type and Familiarization Condition					
	Lures			Targets		
	Standard	Pre-	Post-	Standard	Pre-	Post-
Typical	.284	.324	.388	.687	.641	.710
Unusual	.158	.328	.319	.753	.703	.768
Difference	.126†	-.004	.069**	-.066**	-.062**	-.058*

\* $p < .02$  by  $t$  test. \*\* $p < .01$  by  $t$  test. † $p < .001$  by  $t$  test.

items (mean = .506) and for unusual items (mean = .505). There was, however, a two-way interaction between typicality and item type [ $F(1,90) = 62.0, p < .0001, MSe = .006$ ], supporting the pattern that, whereas the hit rate was greater for unusual targets than for typical targets (means = .74 and .68, respectively), the false alarm rate was less for unusual lures than for typical lures (means = .27 and .33). Thus, recognition accuracy was greater for unusual items.

There also was a two-way interaction between typicality and familiarization condition [ $F(2,90) = 3.39, p < .05, MSe = .010$ ]. However, this interaction, as well as that between typicality and item type, was qualified by a three-way interaction among typicality, item type, and familiarization condition [ $F(2,90) = 5.85, p < .005, MSe = .006$ ]. As shown in Table 1, the typicality effect with targets averaged approximately .06, regardless of familiarization condition. However, the typicality effect with lures was substantially greater in the standard condition (.13) than in the prefamiliarization condition (.00), with the postfamiliarization condition falling between those values (.07). A set of six *t* tests supported a reliable ( $p < .02$ ) effect of typicality in every combination of item type and familiarization condition, except that of prefamiliarization and lures.

In order to test trends for the familiarization procedures to reduce the typicality effect with lures, we conducted two additional analyses of variance, one on false alarms in the standard and prefamiliarization conditions and the other on false alarms in the standard and postfamiliarization conditions. Each of these analyses produced a typicality  $\times$  familiarization condition interaction [ $F(1,60) = 14.5, p < .001, MSe = .009$ , for the standard vs. prefamiliarization comparison, and  $F(1,60) = 4.02, p < .05, MSe = .007$ , for the standard vs. postfamiliarization comparison. Thus, both pre- and postfamiliarization reduced the typicality effect with lures. The effect of prefamiliarization, however, was more convincing.

The original analysis of variance produced no significant effects involving input task. Importantly, the typicality  $\times$  item type  $\times$  familiarization interaction was not qualified by input task ( $F < 1$ ).

### Analysis With Items as Random

A second large analysis of variance was performed on hits and false alarms, this time with items, rather than subjects, treated as the random factor. All previously reported effects were once again reliable, at comparable confidence levels. Importantly, the accuracy advantage for unusual faces was again supported by a typicality  $\times$  item type interaction [ $F(1,94) = 18.9, p < .0001, MSe = .063$ ]. Furthermore, the critical three-way interaction among typicality, item type, and familiarity condition yielded  $F(2,188) = 5.19, p < .01, MSe = .024$ .

### Reaction Time Data

The reaction time data were not highly informative. The average latencies of each subject's correct responses to typical and unusual targets, and to typical and unusual lures, were subjected to an analysis of variance (times over three standard deviations beyond a subject's mean were discarded). The analysis supported three reliable effects: a main effect for item type [ $F(1,90) = 23.9, p < .0001, MSe = .160$ ], a main effect for typicality [ $F(1,90) = 16.0, p < .0001, MSe = .038$ ], and a four-way interaction among item type, typicality, familiarization condition, and input task [ $F(2,90) = 6.11, p < .005, MSe = .021$ ]. A separate analysis of the friendliness-condition data supported no reliable effects involving typicality. In contrast, an analysis of the verbalization-condition data supported a main effect for typicality [ $F(1,45) = 13.3, p < .001, MSe = .042$ ], and also for the interaction among typicality, item type, and familiarization condition [ $F(2,45) = 4.11, p < .05, MSe = .023$ ]. As shown in Table 2, the form of this interaction was similar to that of the typicality  $\times$  item type  $\times$  familiarization condition interaction found with hits and false alarms (Table 1).

### Discussion

Two different findings from Experiment 1 support a familiarity mechanism for typicality effects. First, the familiarity ratings (see Materials section) show that perceived familiarity of entirely new faces is greater if the

Table 2  
Mean Reaction Times (in Seconds) for Correct Recognition Responses to Typical and Unusual Lures (Correct Rejections) and to Typical and Unusual Targets (Hits) for the Three Familiarization Conditions and Each Input Task of Experiment 1

Input Task & Typicality	Item-Type and Familiarization Condition					
	Lures			Targets		
	Standard	Pre-	Post-	Standard	Pre-	Post-
Friendliness						
Typical	1.80	2.09	2.15	1.55	2.09	2.03
Unusual	1.75	2.06	2.16	1.50	2.08	1.85
Difference	.05	.03	-.01	.05	.01	.18*
Verbalization						
Typical	2.00	2.31	2.08	1.88	2.16	1.64
Unusual	1.86	2.33	1.90	1.78	1.99	1.56
Difference	.14*	-.02	.18*	.10*	.17*	.08

\* $P < .05$  by *t* test.

faces are typical than if they are unusual. Second, the false-alarm data—specifically, the interactive effects of typicality and familiarization upon false alarms—add support to this first conclusion and also suggest that presentations of faces cause increments in their perceived familiarity and that the sizes of these increments are greater if the faces are unusual.

A minor puzzle posed by the results was that the effects of pre- versus postfamiliarization were not entirely consistent. Whereas prefamiliarization appeared to eliminate entirely the false-alarm-rate advantage for unusual lures, postfamiliarization appeared merely to reduce this effect. Although we do not here attempt to explain this pattern (but see General Discussion), it should be stressed that both types of familiarization effect matched qualitative predictions of a familiarity hypothesis. The goal of Experiment 2 was to evaluate more thoroughly the familiarity hypothesis, as applied specifically to pre-familiarization effects.

**EXPERIMENT 2**

In order to explain the purpose of Experiment 2, it is helpful to present the familiarity hypothesis somewhat more formally than we have heretofore. Consider “Version 1” of a familiarity hypothesis as outlined in Table 3. The top row of Table 3 shows hypothetical values of perceived familiarity (arbitrary units) for typical and unusual faces that have not been experienced previously. The numbers in the second row show the sizes of increments in perceived familiarity that result from presentations of typical or unusual faces. The numbers in rows 3 through 6 represent the resulting totals of perceived familiarity for typical and unusual faces serving as non-familiarized lures, familiarized lures, nonfamiliarized targets, and familiarized targets in an experiment. Note that a typicality effect is predicted for nonfamiliarized lures, but not for familiarized lures. This is in accordance with the standard and prefamiliarization conditions of Experiment 1.

Despite its success with the pattern of typicality ef-

fects upon false alarms, Version 1 of the familiarity hypothesis makes two clear predictions that were not supported by the results of Experiment 1. The first of these predictions is that the probabilities of old judgments to both typical and unusual faces should vary directly with the number of times these faces have been presented. That is, probabilities of old judgments should be lowest for items that have not been presented prior to recognition (nonfamiliarized lures), higher for items presented once prior to recognition (nonfamiliarized targets and familiarized lures), and highest for items presented twice prior to recognition (familiarized targets). Indeed, without further development, the hypothesis predicts no detectable difference in old-judgment probabilities to familiarized lures versus nonfamiliarized targets. In fact, the probabilities of old judgments obtained in Experiment 1 were quite different for these two item types (means = .34 and .72, respectively). Furthermore, in contrast to predictions, probabilities of old judgments were no lower for nonfamiliarized targets than for familiarized targets (means = .72 and .71, respectively). An obvious explanation for the failure of predictions is the between-subjects design of Experiment 1. This design makes it possible that nonfamiliarized and familiarized items are recognized on the basis of different response criteria.

The second unsupported prediction of Version 1 of the familiarity hypothesis concerns the typicality effect on hits. According to the hypothesis, there should be a stronger trend toward an unusual-face advantage given familiarization than given no familiarization. This prediction is not restricted to Version 1 of the familiarity hypothesis, but generalizes to any version with parameters set to produce (1) an unusual-face advantage for false alarms in a standard (nonfamiliarization) condition and (2) a reduction, elimination, or reversal of this advantage in a familiarization condition (see Table 3). Obviously, this prediction regarding hit rates simply was not supported by the data in Table 1, in which typicality effects on hit rates were impressively equivalent in the standard and prefamiliarization (as well as postfamiliarization) conditions.

**Table 3**  
**Three Versions of a Familiarity Hypothesis Which Holds That Perceived Familiarity of New Faces is Greater if Faces are Typical, and That the Familiarity Increment That Results From Presentation is Greater if Faces are Unusual**

Familiarity	Version of Hypothesis and Type of Face					
	Version 1		Version 2		Version 3	
	Typical	Unusual	Typical	Unusual	Typical	Unusual
Of New Faces	2	1	2	1	3	1
Increment from Presentation	1	2	1	3	1	2
Of Lures in Recognition						
Nonfamiliarized	2	1	2	1	3	1
Familiarized	3	3	3	4	4	3
Of Targets in Recognition						
Nonfamiliarized	3	3	3	4	4	3
Familiarized	4	5	4	7	5	5

*Note*—Numbers represent purely hypothetical values of familiarity of faces (Rows 1, 3, 4, 5, and 6), or of the increment in familiarity of faces resulting from a single presentation (Row 2).

The failure of the familiarity hypothesis to handle the typicality effects on hits is unlikely to be attributable to criterion effects. However, it appeared to us that the between-subjects design used in Experiment 1 might still be responsible for this failure. With this design, it obviously is possible that subjects in different experimental groups adopt different strategies for encoding faces in the input list and/or in the recognition test. Encoding strategies might influence the nature of the typicality effect on hits (cf. Light et al., 1979) and might obscure the effects of familiarization per se.

In order to clarify the nature of familiarization effects, Experiment 2 employed a within-subjects design. Although all subjects received a prefamiliarization set, only half of the target faces and half of the lure faces were included in this set. An important advantage of this within-subject technique is that it eliminates the possibility of differing response criteria for familiarized versus nonfamiliarized items. It also appears to reduce the probability of differential encoding of prefamiliarized faces versus nonfamiliarized faces.

### Method

#### Subjects and Design

The 16 subjects were from the same population as those in Experiment 1. All of these subjects received a prefamiliarization set of 48 faces, followed by an input list of 54 faces (48 to-be-tested items plus 6 fillers), itself followed by a 104-item recognition test (including 96 critical items preceded by 8 "warm-up" items, as in Experiment 1). The input list included 24 items from the familiarization set and 24 nonfamiliarized items. The recognition test included 24 familiarized targets (faces presented previously in both the familiarization set and the input list), 24 nonfamiliarized targets (faces presented previously in the input list only), 24 familiarized lures (faces presented previously in the familiarization set only), and 24 nonfamiliarized lures (faces not previously presented prior to the recognition test). Of the 24 faces in each recognition-testing condition, 12 were "typical" and 12 were "unusual." Thus, the design included the variables of familiarization (prefamiliarized vs. nonfamiliarized), item type (targets vs. lures), and typicality (typical vs. unusual), manipulated factorially and within subjects.

For counterbalancing purposes, we used two different versions of the input list and two different versions of the familiarization set. Each of the four possible combinations was presented to a subgroup of 4 of the 16 subjects. The result was that, across subjects, each of the critical 96 faces served equally often as a prefamiliarized target, a nonfamiliarized target, a prefamiliarized lure, and a nonfamiliarized lure.

#### Materials and Procedure

Both materials and procedures were identical to those used in Experiment 1, except as noted below. The first and most important change was that all subjects received a prefamiliarization set of faces prior to the input list (note, however, that this familiarization set included only 24 of the 48 target items and only 24 of the 48 lures).

A second change from Experiment 1 was that an orienting task was used in the prefamiliarization phase (this was done to ensure attention to each of the familiarization-set items). The subjects were told that we were collecting information on characteristics of faces, and that their task was to rate, on a 3-point scale, each presented face for intelligence.

A third change was that we dropped the manipulation of verbal coding during input-list presentation—All subjects made

friendliness ratings (on a 3-point scale) for the input-list faces. The subjects were told that there was some overlap between the earlier—referred to as "set 1." They were told to disregard this overlap and were not forewarned that memory for either set of faces would be tested.

A fourth change involved the instructions for the recognition test, because it was necessary to modify these slightly from the prefamiliarization condition of Experiment 1. After presentation of the input list (set 2), the subjects were told that still a third set of faces would be presented and that this was to be a recognition test. The task was to judge (on a 6-point confidence scale; 6 = sure from second set, 1 = sure not from second set) whether or not each test face was from the input list (referred to as "set 2").

We note, finally, that the subjects were run in groups of one to four (instead of individually) and that response latencies during recognition were not measured. The faces were presented via slide projector and were shown on a white wall approximately 8 ft in front of the subjects. Prefamiliarization and input-list faces were presented for 5 sec each, with an interstimulus interval of 4 sec. Recognition-test faces were presented for 7 sec each, with an interstimulus interval of 4 sec.

### Results and Discussion

Our concern was with the probabilities of set 2 (input list) judgments, which were defined as responses of 4 through 6 on the 6-point recognition scale. The probabilities are displayed in Table 4, which shows that prefamiliarization increased set 2 judgments to lures (difference = .32) as well as to targets (difference = .31). A three-factor analysis of variance supported a robust main effect for familiarization [ $F(1,15) = 111.4, p < .0001, MSe = .339$ ]. This effect was expected, since (1) familiarization should increase familiarity for both targets and lures and (2) the design precluded criterion adjustments between conditions. The main effect for item type also was reliable [ $F(1,15) = 135.2, p < .0001, MSe = .364$ ], which shows that recognition accuracy was greater than chance (hits exceeded false alarms).

The familiarization and item-type effects produced a strong and direct relationship between old-judgment probabilities and frequency of presentation, precisely as predicted by a familiarity hypothesis. Collapsing over typicality, the probability of old judgments was .09 to

**Table 4**  
Probabilities of "Second-Set" Judgments in Recognition to Typical and Unusual Lures (Not from Second Set) and to Typical and Unusual Targets (From Second Set) That Were Prefamiliarized or Nonfamiliarized in Experiment 3

	Item-Type and Familiarization Condition			
	Lures		Targets	
Typicality	Non-	Pre-	Non-	Pre-
Typical	.115	.380	.427	.698
Unusual	.057	.427	.469	.818
Difference	.058*	-.047	-.042	-.120**

Note—Second-set responses were defined as judgments of 4 through 6 on a 6-point recognition scale. \* $p < .05$  by *t* test. \*\* $p < .01$  by *t* test.

nonfamiliarized lures (not presented prior to recognition), .40 to familiarized lures (presented once prior to recognition), .45 to nonfamiliarized targets (also presented once prior to recognition), and .76 to familiarized targets (presented twice prior to recognition). The near equality of response probabilities to familiarized lures versus nonfamiliarized targets is particularly important, because it supports the conclusion that old judgments were based primarily on perceived familiarity, not on knowledge of presentation context (input list vs. familiarization set). Indeed, an analysis of variance comparing just these item types failed to support a reliable difference between them (or any other effects). This is in accordance with an unelaborated version of the familiarity hypothesis (i.e., all of those versions shown in Table 3), which provides for no knowledge of context.<sup>1</sup>

The original analysis of variance showed a reliable interaction between familiarization and typicality [ $F(1,15) = 5.16, p < .05, MSe = .154$ ], because the effect of familiarization was stronger for unusual items (difference = .36) than for typical items (difference = .27). There was also an interaction between item type and typicality [ $F(1,15) = 8.88, p < .01, MSe = .080$ ], because the difference between targets and lures also was stronger for unusual items (.40) than for typical items (.32). These two interactions support the claim of the familiarity hypothesis that the increment in familiarity that results from presentation is greater for unusual than for typical faces (the latter interaction also shows that recognition accuracy was greater for unusual than for typical faces). We note that the typicality main effect also was reliable [ $F(1,15) = 5.11, p < .05, MSe = .107$ ], because—averaging over familiarization condition as well as item type—the probability of old judgments was slightly greater for unusual faces than for typical faces. This difference (.04) might reflect a bias to judge unusual faces as being old.

The pattern of typicality effects shown in Table 4 is broadly consistent with the data from Experiment 1, as well as with the familiarity hypothesis. As in Experiment 1, (1) there was a false-alarm advantage for unusual lures in the nonfamiliarization condition [ $t(15) = 2.54, p < .05$ ], (2) there was not a false-alarm advantage for unusual lures in the prefamiliarization condition (indeed, the trend was for a reversal of the effect), and (3) there was a hit-rate advantage for unusual targets, although it was reliable only with prefamiliarization [ $t(15) = 3.53, p < .01$ ].<sup>2</sup>

A discrepancy between Experiment 1 and Experiment 2 was that the hit-rate advantage for unusual targets appeared somewhat larger given prefamiliarization (difference = .12) than given nonfamiliarization (difference = .04) in Experiment 2. This discrepancy is potentially important, because it renders the data from Experiment 2 highly consistent with the familiarity hypothesis (Table 3). However, it appears inappropriate to take this pattern seriously, because it was not supported statistically. An analysis of variance performed just on

the target data failed to support a reliable interaction between typicality and familiarization ( $p > .10$ ), although the main effects of both variables were highly reliable ( $ps < .001$ ). Furthermore, an experiment quite similar to the present Experiment 2 was completed recently in our laboratory, and it failed to support the pattern (instead, we found an unusual face advantage of approximately .06 for both nonfamiliarized targets and prefamiliarized targets, just as in Experiment 1). Moreover, an analysis of discrimination between target and lure faces suggested high correspondence between the results of Experiment 2 and those of Experiment 1, despite differing patterns of hit rates.

**Area-under-the-MOC scores.**<sup>3</sup> We used the nonparametric measure of area under the MOC (Swets, 1973) to assess discrimination between target- and lure-typical faces, and between target- and lure-unusual faces, for each individual subject in both experiments. The means of the area scores for all conditions are shown in Table 5. The size of these scores can be taken to reflect the familiarity increment that resulted from presentation in the input list. The data suggest that the familiarity increment was higher for unusual than for typical faces, but this effect was clearer in the nonfamiliarization (or standard) condition than in the prefamiliarization condition, in which it failed to reach conventional significance levels (the effect in the postfamiliarization condition of Experiment 1 appeared to be intermediate in size).

The area scores from both experiments converge with the hit-rate data from Experiment 1; both suggest some modification of the familiarity hypothesis. These data imply that the familiarity increment that results from presentation is greater for more unusual faces, but primarily if these faces are entirely new. We suspect that there is differential processing of previously viewed faces versus entirely new faces during the input list, especially if these faces are unusual. Such an effect could

**Table 5**  
Area-Under-the-MOC Scores for Target-Lure Discrimination From Experiments 1 and 2 for Typical and Unusual Faces That Were Prefamiliarized, Nonfamiliarized, or Postfamiliarized (Experiment 1 Only)

Typicality	Familiarization Condition		
	Non-	Pre-	Post-
Experiment 1			
Typical	.755	.707	.717
Unusual	.852	.743	.783
Difference	-.097**	-.036*	-.066**
Experiment 2			
Typical	.729	.707	
Unusual	.831	.767	
Difference	-.102**	-.060*	

Note—Area-under-MOC scores generally vary from .50 (chance discrimination) to 1.00 (perfect discrimination). \* $.05 < p < .10$  by *t* test. \*\* $p < .001$  by *t* test.

be quite independent of subjects' conscious strategies (which we attempted to control in Experiment 2). Since this clearly is a matter for future research, it will not be given additional treatment here.

### GENERAL DISCUSSION

The major empirical conclusion supported in this research is that prefamiliarization of faces can remove the false-alarm-rate advantage of unusual lures over typical lures in recognition. This conclusion was supported with a between-subjects manipulation of prefamiliarization (Experiment 1), and also with a within-subjects, within-list manipulation (Experiment 2). Furthermore, the pattern was supported when a latency measure was used, although only if the subjects had described verbally the input faces (Table 2). Postfamiliarization had effects that were similar to those of prefamiliarization, but weaker (Experiment 1). Additional empirical conclusions were that (1) entirely new faces are rated as more familiar if they are typical than if they are unusual and (2) subjects are (in some cases) poor at distinguishing in recognition between faces presented once in an input list and faces presented once in an immediately preceding familiarization set (cf. Brown et al., 1977).

At the level of theory, the most important implication of our empirical conclusions is that context-free familiarity information is operative in face recognition and is susceptible to effects of face typicality. Specifically, our data suggest that (1) the perceived familiarity of never-before-experienced faces is greater if they are typical than if they are unusual and that (2) the increment in perceived familiarity that results from presentation is greater for unusual faces than for typical faces. However, in view of the area scores shown in Table 5, this second conclusion probably should be qualified. It is supported more strongly for the first than for the second presentation of a face.

Although our results support typicality effects upon the perceived familiarity of faces, the conditions were such that there was minimal memory for the presentation context of faces. There is nothing in our data to rule out the possibility that typicality of faces can affect retrieval of contextual information, under conditions in which such retrieval is possible. Indeed, the results we obtained in the postfamiliarization condition of Experiment 1 are interpretable in these terms. The false-alarm-rate advantage for unusual lures was reduced in that condition, as compared with the standard condition. However, it was not eliminated entirely (see Table 1). Furthermore, as argued previously (see the introduction), a false-alarm advantage in a familiarization condition is predicted by the notion that memory of context is better in response to unusual than to typical faces.<sup>4</sup>

An additional caveat that should be expressed is that the nature of the information underlying familiarity of faces presently is quite unclear. One particularly pressing

question is whether such information could possibly be construed as a crude, nondiscriminating type of contextual information. It is possible to argue that the advantage in recognition of unusual faces—at least the advantage in area scores and hit rates—reflects effects of typicality upon context retrieval, so long as context is defined broadly. For example, the effective context in our experiments might have been that of the entire laboratory experience, rather than that of a particular input list. If so, and if such context was more highly accessible in response to unusual faces, the effects of typicality upon hit rates and area scores might be explained. The effect of typicality upon false-alarm rates might also be accommodated, but additional assumptions would be required.

Apart from the issue of the nature of familiarity information, it is important to consider the processing mechanisms that render it sensitive to typicality effects. Mandler's (1980; Mandler, Goodman, & Wilkes-Gibbs, 1982) model of recognition memory suggests one possible view of these mechanisms: Perhaps never-before-experienced faces have nonzero levels of "baseline familiarity," which derive from their similarity to (1) other faces experienced in a person's past or (2) "prototype" faces formed through experience with many individual exemplars. In either case, given that interitem similarity among typical faces is high (Light et al., 1979), it is plausible that their baseline familiarities also are high, that is, higher than those of unusual faces. Indeed, an effect of typicality upon baseline familiarity was directly supported by the familiarity ratings collected in Experiment 1. Given this effect, and given that baseline familiarity affects false-alarm errors in recognition (see Mandler et al., 1982), an effect of typicality upon false-alarm errors would be expected to occur.

In order to account for familiarization effects, Mandler's (1980) distinction between "baseline" familiarity and "effective" familiarity is necessary. According to Mandler, a recently presented stimulus has an effective familiarity value, and this value reflects the proportional increment in its baseline familiarity caused by its recent presentation. The account implies that the effective familiarities of recently presented stimuli are inversely correlated with their baseline familiarities. Thus, if baseline familiarity is higher for typical faces, effective familiarity should be higher for unusual faces. Assuming that effective familiarity determines responsiveness to familiarized faces in recognition, false alarms to such faces should be higher if they are unusual. That is, familiarization should reverse the false-alarm-rate advantage for unusual lures.

Unfortunately, an actual reversal of the typicality effect as a result of familiarization was not convincingly supported by this research. In Experiment 1, prefamiliarization eliminated the false-alarm-rate advantage for unusual lures, but did not reverse this advantage (and, postfamiliarization merely reduced the effect). In Experiment 2 (Table 4), there was a trend for a reversal, but it



was not supported statistically—false alarms to unusual lures versus those to typical lures did not differ significantly in the prefamiliarization condition.

In order to completely accommodate our data, a hypothesis based on Mandler's (1980) model requires elaboration. For example, it might be arguable that, under prefamiliarization conditions, subjects base their recognition decisions sometimes on baseline familiarity and sometimes on effective familiarity. Since baseline familiarity would support one result and effective familiarity would support its opposite, the consequence might be that there would be no difference in false-alarm rates to typical and to unusual lures. Perhaps there are more elegant ways to modify the model.

A distinctiveness hypothesis (e.g., Eysenck, 1979; Lockhart et al., 1976; Winograd, 1981) offers an alternative way of viewing familiarity effects. It is a plausible assumption that unusual faces evoke highly distinctive encodings. Indeed, the interitem similarity data from Light et al. (1979) support this assumption. Hence, unusual lures that have not been familiarized should be minimally confusable with any and all previously experienced faces, including those experienced on the input list. A low false-alarm rate should result (see Davies, Shepherd, & Ellis, 1979, for target-lure confusability effects). Unusual lures that have been familiarized should be highly recognizable as having been previously experienced—perhaps due to high “distinctive overlap” (see Eysenck, 1979) between the test encodings and the familiarization encodings of the stimuli. This should lead to a high false-alarm rate for unusual lures. Indeed, the problem for a distinctiveness view is similar to that for a view based on Mandler's (1980) model. Both imply a reversal of the lure-typicality effect under familiarization conditions. Yet, such a reversal is not supported strongly by the data.

Within a “distinctiveness” framework, the problem of interpreting the data from prefamiliarized lures probably can be reduced to that of interpreting the data from nonfamiliarized targets. Both item classes are presented just once prior to recognition. Furthermore, subjects taking a recognition test apparently find these two item classes virtually impossible to distinguish (Table 4). Thus, it is possible that those factors that influence the typicality effect with hits also influence the typicality effect with prefamiliarized lures. The research by Light et al. (1979) suggested that presentation time and encoding task are among these factors. Indeed, a role for these factors is plausible, since both might determine the probability that the distinctive characteristics of unusual faces are initially coded or “scanned” (see Winograd, 1981, and the introduction).

Tulving's (1982) new model of “synergistic ecphory” provides a third way in which to conceptualize a familiarity mechanism. According to Tulving, retrieval involves an interaction between retrieval-cue information and stored episodic information, with the result that “ecphoric information” is produced. Furthermore,

“Familiarity reflects the similarity between the information provided by the retrieval cue, or the test item, on the one hand, and the information contained in the actualized ensemble of ecphoric information, on the other” (Tulving, 1982, p. 143). Applying this model to face recognition, it clearly is arguable that retrieval operations performed in response to nonfamiliarized lures (sometimes) produce ecphoric information, with the consequence that similarity between ecphoric information and test-item information is assessed. Such similarity should be greater for more typical faces (due to the interitem similarity effect; Light et al., 1979). Hence, more typical faces should “feel” more familiar, and this should affect the number of false alarms.

The results with familiarized lure faces also are compatible with the synergistic ecphory model. Since identical copies of such faces have previously been experienced, retrieval operations should frequently produce ecphoric information, and there generally should be high similarity between this ecphoric information and information provided by the familiarized lure itself. Hence, the degree of subjectively felt familiarity should be high, both for typical and unusual faces. This clearly is consistent with the data we have obtained.

The technique of familiarization provides a promising way to assess the mnemonic effects of stimulus characteristics, not only in face memory, but in other domains as well. The technique already has been applied to the frequency effect in word recognition by Eysenck (1979) and by Kinsbourne and George (1974). Although this work has produced empirical ambiguities, these should be clarified as the use of the technique is refined. The present research demonstrates the importance of examining hits and false alarms separately, and also suggests that the effects of pre- versus postfamiliarization might differ in important ways. A particularly exciting role for the familiarization technique is to clarify the relationships between the effects of different stimulus variables—whereas familiarization apparently weakens the effects of some variables (such as typicality), it might strengthen the effects of others. Such differing patterns should be invaluable for determining just how different variables produce their effects (e.g., through perceived familiarity vs. knowledge of context). Within the domain of face memory, the effects of typicality and those of “stereotypy” (Klatzky, Martin, & Kane, 1982) might be fruitfully compared in this way.

#### REFERENCES

- BROWN, E., DEFFENBACHER, K., & STURGILL, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology*, *62*, 311-318.
- CHANCE, J., & GOLDSTEIN, A. G. (1976). Recognition of faces and verbal labels. *Bulletin of the Psychonomic Society*, *7*, 384-386.
- COHEN, M. E., & CARR, W. J. (1975). Facial recognition and the von Restorff effect. *Bulletin of the Psychonomic Society*, *6*, 383-384.
- DAVIES, G. M., SHEPHERD, J. W., & ELLIS, H. D. (1979). Sim-

- ilarity effects in face recognition. *American Journal of Psychology*, **92**, 507-523.
- EYSENCK, M. W. (1979). Depth, elaboration and distinctiveness. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory*. Hillsdale, NJ: Erlbaum.
- GOING, M., & READ, J. D. (1974). Effects of uniqueness, sex of subject, and sex of photograph on facial recognition. *Perceptual and Motor Skills*, **39**, 109-110.
- JACOBY, L. L., & CRAIK, F. I. M. (1979). Effects of elaboration of processing at encoding and retrieval: Trace distinctiveness and recovery of initial context. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory*. Hillsdale, NJ: Erlbaum.
- KINSBOURNE, M., & GEORGE, J. (1974). The mechanism of the word frequency effect on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, **13**, 63-69.
- KLATZKY, R. L., MARTIN, G. L., & KANE, R. A. (1982). Semantic interpretation effects on memory for faces. *Memory & Cognition*, **10**, 195-206.
- LIGHT, L. L., KAYRA-STUART, F., & HOLLANDER, S. (1979). Recognition memory for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory*, **5**, 212-228.
- LOCKHART, R. S., CRAIK, F. I. M., & JACOBY, L. L. (1976). Depth of processing, recognition and recall. In J. Brown (Ed.), *Recall and recognition*. London: Wiley.
- LOFTUS, E. F. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.
- MANDLER, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, **87**, 252-271.
- MANDLER, G., GOODMAN, G. O., & WILKES-GIBBS, D. L. (1982). The word frequency paradox in recognition. *Memory & Cognition*, **10**, 33-42.
- SWETS, T. A. (1973). The relative operating characteristic in psychology. *Science*, **182**, 990-1000.
- TULVING, E. (1982). Synergistic ephory in recall and recognition. *Canadian Journal of Psychology*, **36**, 130-147.
- WINOGRAD, E. (1981). Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology: Human Learning and Memory*, **7**, 181-190.
- YARMEY, A. D. (1979). *The psychology of eyewitness testimony*. New York: Free Press.

## NOTES

1. Of course, knowledge of context contributes to face recognition in some situations, and a familiarity hypothesis must be elaborated to handle such situations.

2. Compared with the effect in Experiment 1, the effect of typicality upon false alarms to nonfamiliarized lures was relatively small. This apparently was due to floor effects for unusual nonfamiliarized lures—11 of 16 subjects made 0 false alarms to such lures. When set 2 judgments were redefined as responses of 3 through 6 on the 6-point scale (instead of 4 through 6, as above), false-alarm rates for nonfamiliarized lures were .214 and .099 for typical and unusual items, respectively, a highly reliable difference ( $p < .001$ ).

3. In order to calculate area-under-the-MOC scores, a subject's confidence rating data are used to derive up to five different pairs of hit and false-alarm rates, with each pair representing a unique criterion level. The area under the resulting MOC is then computed (the MOC can be constructed by plotting hit rates against false-alarm rates, and then by connecting the resulting points to form a continuous line from the lower left corner to the upper right corner of an MOC graph). Since the MOCs are derived empirically, the technique removes the necessity of assuming that all MOCs have a particular form. Furthermore, area-under-the-MOC scores make use of all of a subject's confidence-rating data.

4. There are other interpretations of the postfamiliarization results. Note that the postfamiliarization condition involved a lengthening of the retention interval between input list and recognition test, and that subjects in this condition might have made implicit recognition decisions—during postfamiliarization—regarding oldness of input-list items. These observations provide the basis for alternative theoretical accounts.

(Manuscript received June 10, 1983;  
revision accepted for publication December 16, 1983.)