INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
Brooks, John Oliver, III

AFFECTIVE JUDGMENT AND RECOGNITION MEMORY

Rice University  M.A.  1996

University Microfilms International  300 N. Zeeb Road, Ann Arbor, MI 48106
RICE UNIVERSITY

AFFECTIVE JUDGMENT AND RECOGNITION MEMORY

by

JOHN OLIVER BROOKS III

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

MASTER OF ARTS

APPROVED, THESIS COMMITTEE:

[Signatures]
Michael J. Watkins, Professor of Psychology, Chairman

[Signature]
Sarah A. Burnett, Associate Professor of Psychology

[Signature]
Craig A. Anderson, Associate Professor of Psychology

Houston, Texas

May, 1986
Affective Judgment and Recognition Memory

John Oliver Brooks III

Abstract

Previous research has demonstrated that exposure to novel stimuli makes the stimuli more likeable. Several researchers have suggested that this effect is more sensitive than, and independent of, recognition of the stimuli. Four experiments were conducted to evaluate this suggestion. The first two experiments failed to find enhanced liking for previously presented items, even though they were recognized at well above chance level. In Experiments 3 and 4, an "affect" test did distinguish between previously seen and new stimuli, but not to the same extent as did a recognition test. Moreover, contingency analyses of subject-items indicated that affective judgment and recognition were related.
Acknowledgments

I would like to express my gratitude to the members of my committee, Sarah Burnett and Craig Anderson, and to Michael Watkins, whose unending support, patience, and guidance enabled me to complete this thesis.

Ταρασσει ους Ἀνθρώπους σοι τα Πραγματα, Ἀλλα τα περι των Πραγματων Δογματα.
(It is not things that disturb men, but their judgments about things.)

- Encheiridion of Epictetus
## Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Experiment 1</td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>8</td>
</tr>
<tr>
<td>Results</td>
<td>11</td>
</tr>
<tr>
<td>Experiment 2</td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>17</td>
</tr>
<tr>
<td>Results</td>
<td>20</td>
</tr>
<tr>
<td>Experiment 3</td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>27</td>
</tr>
<tr>
<td>Results</td>
<td>29</td>
</tr>
<tr>
<td>Experiment 4</td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td>34</td>
</tr>
<tr>
<td>Results</td>
<td>35</td>
</tr>
<tr>
<td>General Discussion</td>
<td>40</td>
</tr>
<tr>
<td>References</td>
<td>47</td>
</tr>
<tr>
<td>Appendix A</td>
<td>50</td>
</tr>
<tr>
<td>Appendix B</td>
<td>53</td>
</tr>
<tr>
<td>Appendix C</td>
<td>55</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1. A sample of the stimuli used in Experiment 2  
Figure 2. Mean performance levels in Experiment 2

List of Tables

Table 1. Proportion Correct When the Recognition Test Preceded the Affective Judgment (Experiment 1)  
Table 2. Proportion Correct When the Affective Judgment Preceded the Recognition Test (Experiment 1)  
Table 3. Contingency Tables and Phi Coefficients for Subject-Item Data for the Immediate Test (Experiment 1)  
Table 4. Contingency Tables and Phi Coefficients for Subject-Item Data for the Delay Test (Experiment 1)  
Table 5. Proportion Correct by Level of Confidence for the Immediate Test (Experiment 1)  
Table 6. Proportion Correct by Level of Confidence for the Delay Test (Experiment 1)  
Table 7. Proportion Correct When the Affective Judgment Preceded the Recognition Test (A-R) (Experiment 2)  
Table 8. Proportion Correct When the Recognition Test Preceded the Affective Judgment (R-A) (Experiment 2)
Table 9. Contingency Tables for Subject-Item Data from (Experiment 2) High Illumination  

Table 10. Contingency Tables for Subject-Item Data from (Experiment 2) Medium Illumination  

Table 11. Contingency Tables for Subject-Item Data from (Experiment 2) Low Illumination  

Table 12. Product-moment Correlations for Recognition-Confidence Measure  

Table 13. Mean Ratings for the Japanese Characters (Experiment 3)  

Table 14. Mean Affect Ratings According to Judged Status and Actual Status (Experiment 4)
Often it seems that the more we encounter something, the more we like it. We may be indifferent to a concerto on first hearing it, but come to like it more and more with repeated hearings. This commonplace has been borne out by laboratory experiment (e.g., Zajonc, 1968; Harrison, 1977) and is generally known as the "mere exposure effect" (Zajonc, 1968; but see Burgess & Sales, 1971). The mere exposure effect appears to be rather robust, occurring under a variety of conditions with different types of stimuli. In his 1968 monograph, Zajonc reported that liking for stimuli resembling Chinese characters increased in a linear fashion as a function of the number of times subjects saw the stimuli. This pattern of results was replicated when Turkish adjectives were used as stimuli and subjects were asked to infer "goodness" or "badness" upon the adjectives. Becknell, Wilson, and Baird (1963) found that subjects preferred frequently presented nonsense syllables to unexposed syllables; similar effects have been reported using a variety of stimuli, such as tone sequences (Wilson, 1979), oriental characters (Moreland & Zajonc, 1977; Saegert & Jellison, 1970; Zajonc, 1968), pseudowords (Matlin, 1971), and random shapes (Hamid, 1973; Kunst-Wilson & Zajonc, 1980; Seamon et al., 1983). The presentation times of the stimuli have varied widely across studies, ranging from 1 ms (Kunst-Wilson & Zajonc, 1980) to 2 s (Moreland & Zajonc, 1977). Exposure effects have also been found when the number of exposures is varied between subjects (Moreland & Zajonc, 1976), which suggests that, contrary to an argument put forth by Stang (1974a), the effect is not a result of demand characteristics that could arise if subjects discover the experimental hypothesis.
The use of novel stimuli is crucial; repeated exposures do not appear to alter liking for stimuli that are already familiar (cf. Harrison, 1977). This is only reasonable because subjects may be assumed to have already established opinions of familiar stimuli, opinions which are not likely to be swayed by their additional occurrence in an experiment. Otherwise, exposure effects appear to be as automatic as memory: One need only to expose subjects to novel stimuli and ask for affective judgments to obtain an effect.

**Affect and Recognition**

Of course, things that have been previously exposed may well be recognized as such, for exposure induces recognition as well as liking. For example, the number of times a concerto has been heard in the past will determine the likelihood that the concerto will be recognized as having been heard before. The purpose of this thesis is to explore this relation between affect and recognition. Perhaps performance on one test drives performance on the other. If affective judgments were driven by recognition, for example, then the mere exposure effect should occur only for recognized stimuli. Contrariwise, if recognition were driven by affect, then subjects should recognize only the words they like.

Zajonc (1968) realized that exposure enhances affect as well as recognizability and dismissed the possibility of one driving the other as follows:

...it is possible that this [exposure] manipulation is now confounded with ease of recognition. This danger of confounding, however, is probably minimal because at no time were the subjects ever required to recognize or discriminate the [stimuli]. (p. 15)
Such reasoning is not completely in line with a phenomenological perspective of recognition. From everyday experiences we know that recognition can occur without prompting and without an explicit statement that recognition has occurred. Therefore, it is quite likely that subjects in Zajonc's (1968) experiments recognized at least some of the stimuli. In a later paper, Zajonc claimed that ". . . affective reactions to a stimulus may be acquired by virtue of experience with that stimulus even if not accompanied by . . . conscious recognition" (1980, p. 163).

The nature of the relation between recognition and affect was first explored experimentally by Matlin (1971). Subjects viewed pseudo-words (e.g., NANSOMA) three times each and, in a test that followed, indicated that they liked previously presented pseudo-words more than nonpresented pseudo-words. An analysis of errors led Matlin to claim that, for all words subjects thought were old, previously presented words were liked more than nonpresented words; the same pattern also held for words that subjects thought were new. Furthermore, Matlin found that for previously presented words, judged status\(^1\) was associated with liking in that subjects liked words thought to be old better than words thought to be new. In light of this somewhat inconsistent finding, Matlin went on to point out that, from her results, it was not clear whether affect depended on judged status.

Moreland and Zajonc (1977) extended Matlin's (1971) research. They used Japanese characters (ideographs) to examine the relation between affective judgment and recognition at different exposure frequencies using Japanese characters. Their central concern was the removal of demand

---

\(^1\)From this point on, the term "judged status" will be used to denote the subject's opinion of an item's status. This does not necessarily coincide with the actual status of the item (i.e., presented or nonpresented).
characteristics that could have existed in Matlin's study. To this end, one group of subjects was asked only to recognize the Japanese characters that had been previously presented, whereas another group was asked only to rate how much they liked the characters. For the purposes of their analyses, Moreland and Zajonc randomly designated yoked pairs of subjects, combined the affect and recognition ratings on an item basis, and then treated the pairs of numbers as single sources of data. Their explanation for this unusual procedure was that the measurements were taken "under objectively the same conditions" (p. 194), and thus the recognition judgments from one subject could be matched with the affective judgments of another subject. Finally, steps were taken to eliminate the variance due to the yoked pairs so that each observation could be considered independent of the others.

The above method of matching ratings is undesirable for a variety of reasons, the most salient of which is that Moreland and Zajonc's (1977) method of combining data would decrease the magnitude of any correlation that might have been present. Because each of the ratings in a pair was provided by a different subject, more variance is introduced in the pairs of scores and hence a lower correlation is not surprising. More importantly, Moreland and Zajonc concluded that (perhaps only as a result of their analytical methods!) recognition was not a necessary condition for enhanced liking of a previously exposed stimulus. A correlation between affect and recognition may have existed in their data, but may not have been large enough to detect using their analytical methods.

Birnbaum and Mellers (1979a; 1979b) objected to Moreland and Zajonc's (1977) conclusion of independence, arguing instead that, on the basis
of Moreland and Zajonc's data, one could not reject the hypothesis that the exposure effects are mediated by recognition. In support of their claim, Birnbaum and Mellers used Moreland and Zajonc's correlation matrix and demonstrated that it fit a model in which judged status mediated affect ratings.

Wilson (1979) explored the relation between liking and recognition by attempting to lower recognition performance while maintaining affective performance at a relatively constant level. A dichotic listening task was used in which subjects were instructed to proofread a literary passage while listening to an auditory version of the passage in one ear, and to ignore anything presented in the other ear, in which tone sequences were played five times each. In two experiments, Wilson reported that, in a subsequent test, subjects liked the tone sequences that had been played more than the ones that had not been played, in spite of a rather low level of performance on a recognition test for the tone sequences. Thus, Wilson concluded that enhanced liking does not depend on recognition and, in effect, that affect was more sensitive to prior exposure than recognition.

Kunst-Wilson and Zajonc (1980) reasoned that, if affective judgment were indeed more sensitive to exposure than recognition, subjects might be able to discriminate affectively between new and old stimuli when overall recognition performance had been brought down to the chance level. To obtain chance recognition performance, Kunst-Wilson and Zajonc presented subjects with ten random shapes, five times each, with a presentation time of 1 ms and a small interstimulus interval under conditions of low illumination. The subjects could tell only that a flash had occurred, but not what had been presented, and on a subsequent forced-choice recognition test
they were unable to tell which shapes had been flashed earlier and which had not. Nevertheless, subjects preferred old shapes over new shapes about 60% of the time in a forced-choice affective task, which was greater than the chance level of 50%. From this finding, Kunst-Wilson and Zajonc concluded that enhanced liking may occur in the absence of recognition. Thus, Kunst-Wilson and Zajonc's data tell us that there were some old shapes that were liked but not recognized, but they provide little information on the relation between liking and recognition, which must be derived from contingency analyses.

Seamon et al. (1983) undertook to replicate the findings of Kunst-Wilson and Zajonc (1980), but used a 3 ms presentation of a stimulus in conjunction with a shadowing task in order to lower recognition performance. Even when recognition performance was lowered because of the shadowing task (from 66% to 59%), affective discrimination of the shapes was above chance (about 65%), although it too was affected by the shadowing task. In a second experiment, the presentation time was reduced to 2 ms in an attempt to lower recognition performance even more. Recognition in the shadowing condition dropped to the chance level, whereas affective judgment remained above chance levels. In light of the finding that affective judgment could exceed recognition, Seamon et al. concluded that affective judgment and recognition were independent and that "... questions about which stimuli are liked is a better way to access familiarity than questions about which stimuli were seen before" (1983, p. 554).

The last two experiments conducted by Seamon et al. (1983, Experiments 3 & 4) investigated the effect of presenting shapes in different
visual fields. When a 2 ms presentation of a shape was either pattern
masked or not masked at all, they found that affective judgments were
better for stimuli presented in the right visual field and recognition
judgments were better for stimuli presented in the left visual field. Thus, by
manipulating which visual field a stimulus is presented in, one may
differentially affect liking and recognition.

There are certain aspects of Seamon et al.'s (1983) data, however, that
appear to contradict the notion of independence. First, there was no
interaction between the presence or absence of the shadowing task with the
type of test. That is, the shadowing activity did not differentially affect
recognition and liking. Therefore, the two tests were not dissociated with
respect to the activity of shadowing. More evidence of an association
between the tests may be found in the contingency analyses. In Seamon et
al.'s second experiment, all phi coefficients were about .20, which the
researchers characterized as "slight, positive association[s]" (p. 548). The
consistency of the positive associations, although they were not large, could
be construed as indicative of a true underlying association between the tests.

Thus, some researchers have argued that affective judgment tasks are
more sensitive to an item's previous occurrence than recognition tests. This
claim is in line with Zajonc's (1980) characterization of affect as a basic and
omnipresent quality of perception. Such a conceptualization led Zajonc to
argue that not only could affective judgment and recognition be
independent, but also that affect could in some way be the basis for
recognition. The four experiments that follow were designed to investigate
these claims.
Experiment 1

In the first experiment, subjects viewed single presentations of Turkish words and were tested on the words both immediately following the presentation list and after a delay of 30 min. Turkish words have been used in previous exposure research (e.g., Zajonc, 1968), in which the effect of a single exposure could not be determined. The delay before testing was included to address experimentally Stang's (1974b) argument that exposure effects are more likely to occur after a delay. For each word presented in Experiment 1, both liking and recognition judgments were made. Liking was assessed indirectly by asking the subject to guess which of a pair of words had an evaluatively positive, or "good" meaning. Previous findings have indicated that subjects choose previously presented words over completely new ones as having a good meaning (cf. Moreland & Zajonc, 1977, Experiment 2; Zajonc, 1968). Because previous research has suggested that exposure effects vary with respect to confidence subjects have in their answers (Kunst-Wilson & Zajonc, 1980), confidence ratings were taken for both the affective task and the recognition test.

Method

Subjects. The subjects were 16 Rice University undergraduates who received course credit for their participation. They were tested individually.

Stimuli and Apparatus. The stimuli, which are listed in Appendix A, consisted of 240 Turkish words drawn randomly from a Turkish psychology text. All words were 5 to 7 letters in length and those resembling English words were avoided; otherwise, selection was random. All of the words were presented on an Apple Macintosh computer.
**Design.** The design was a $2 \times 2 \times 2 \times 2$ factorial with order of testing at the immediate test (recognition followed by affect, R-A, or affect followed by recognition, A-R) and order of testing at the delayed test (R-A or A-R) manipulated as between-subjects variables and testing interval (immediate or delay) and type of test (affective judgment or recognition) manipulated as within-subjects variables.

The 240 words were randomly assigned to two sets of 120 words. Half of the subjects viewed and were tested on one set of words; henceforth, the "old" words. The other set of words served as lures in the recognition tests and as comparison words in the affective judgment tests, and may be referred to as the "new" words. For the other half of the subjects, the old and new word sets were interchanged. In this way, each word served as both an old word and a new word equally often.

Within each testing interval, the contents and procedures of the recognition and affective judgment tests were identical for a given subject; they differed only in the instructions given to the subjects. At each testing interval, 60 of a given subject's old words (i.e., half of the previously presented set) were randomly paired with 60 of the subject's new words.

**Procedure.** After being asked whether they had any knowledge of the Turkish language (none did), subjects were told that they were about to see a series of Turkish words. They were instructed to watch each word as it appeared on the computer screen because they would later be given various tests concerning the words. The nature of the tests was not specified. The words were presented, back to back, at the rate of one word per second. Before the presentation of the actual list began, each subject saw a practice
set of ten Turkish words in order to become familiar with the presentation rate; this set was the same for all subjects.

After the words had been presented, subjects received either an affective judgment test followed by a recognition test, or vice versa. A practice test preceded the actual test for both affective judgment and recognition in the immediate testing condition. After a 30 min delay, during which subjects engaged in an unrelated memory span task, they received an affective judgment test and a recognition test.

For the recognition test, subjects were told that they would be presented with pairs of Turkish words, and that one word of each pair was a member of the list that they had just viewed whereas the other word was new. They were asked to choose the word they thought had been presented and give a confidence rating of their selection (1 - Guess; 2 - Possible; or 3 - Probable). The confidence rating numbers and their labels were displayed on the screen. Subjects made all of their responses by typing a key on the computer keyboard.

The affective judgment test was identical to the recognition test except for the instructions. Subjects were told that the Turkish words were in fact adjectives paired in such a way that, in the Turkish language, one word had a good connotation and the other a bad connotation. The subjects were told that people could sometimes pick up on characteristics of words that might lead them to believe that the word has a good or bad meaning in another language. The subjects' task was to guess which word had the "good" meaning. In reality, the pairing of the words was the same as in the recognition test for the particular testing interval and determined randomly by the computer for each subject.
A confidence rating of each choice, again on a 3-point scale, was taken after each judgment. Because subjects were of necessity guessing which word was "good", they were told to adjust the meaning of the labels in such a way that they used all three levels about equally often throughout the test. For example, they were told that "probable" could be interpreted to mean "less of a guess than some other ones."

During the immediate and delayed tests, subjects engaged in an unrelated memory span task that lasted approximately 30 min. After the delay, subjects received either an affective judgment test followed by a recognition test, or vice versa.

Results

The results are summarized in Tables 1 and 2. For both tests, an item was scored as "correct" if the subject selected the old item (i.e., one that, for that subject, had been included in the presentation list). Thus, in the affective judgment task, a response was classified as correct if the old word was chosen as the word with the good meaning. The data were submitted to a four-way analysis of variance. From the cell means, it is apparent that the delay manipulation affected only recognition performance and then only for subjects in the R-A condition (F (1, 12) = 5.19, p < .05, MS_e = 4.578). Further, there were statistically significant main effects of testing interval (F (1, 12) = 11.11, p < .01, MS_e = 10.161) and type of test (F (1, 12) = 30.39, p < .001, MS_e = 11.724). These main effects reflect a decline in performance over the delay and the superiority of recognition performance in comparison to affective judgment. Pairwise comparisons revealed that performance on the affective judgment task did not exceed the chance level in any condition.
**Table 1**

Proportion Correct When the Recognition Test Preceded the Affective Judgment (Experiment 1)

<table>
<thead>
<tr>
<th>Delay interval</th>
<th>Recognition</th>
<th>Affect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate</td>
<td>.673</td>
<td>.522</td>
</tr>
<tr>
<td>30 min.</td>
<td>.565</td>
<td>.538</td>
</tr>
</tbody>
</table>

**Table 2**

Proportion Correct When the Affective Judgment Preceded the Recognition Test (Experiment 1)

<table>
<thead>
<tr>
<th>Delay interval</th>
<th>Recognition</th>
<th>Affect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate</td>
<td>.550</td>
<td>.522</td>
</tr>
<tr>
<td>30 min.</td>
<td>.550</td>
<td>.440</td>
</tr>
</tbody>
</table>

That the test x immediate order x testing interval interaction term achieved statistical significance ($F(1, 12) = 11.88, p < .01, MS_e = 4.578$) is not too surprising given that the content of the two tests was identical. When the affective judgment task preceded the recognition test, subjects had seen all of the lures before being asked to make recognition judgments, which
made it more difficult for them to discriminate between items that were members of the presentation list and those that were not.

The contingency data are given in Tables 3 and 4. In all conditions, the fourfold point correlation coefficients were rather small, between .05 to .12. If the affective judgments were being made at chance with respect to the actual status of the words (i.e., old or new), then no correlation would be expected. Because there was no exposure effect, these coefficients are not especially interesting.

The data from Experiment 1 were also broken down according to level of confidence, shown in Tables 5 and 6. In only one cell did performance on the affective judgment task exceed recognition performance, and then only slightly so. This pattern of results contrasts sharply with that obtained by Kunst-Wilson and Zajonc.

### Table 3

**Contingency Tables and Phi Coefficients for Subject-Item Data for the Immediate Test (Experiment 1)**

<table>
<thead>
<tr>
<th></th>
<th>Affect—Recognition</th>
<th>Recognition—Affect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rn</td>
<td>Rn</td>
</tr>
<tr>
<td>Affect</td>
<td>Fail Pass</td>
<td>Fail Pass</td>
</tr>
<tr>
<td>Pass</td>
<td>.219 .304</td>
<td>.144 .379</td>
</tr>
<tr>
<td>Fail</td>
<td>.231 .246</td>
<td>.183 .294</td>
</tr>
<tr>
<td>ø</td>
<td>.067</td>
<td>.116</td>
</tr>
</tbody>
</table>
### Table 4

**Contingency Tables and Phi Coefficients for Subject-Item Data for the Delay Test (Experiment 1)**

<table>
<thead>
<tr>
<th></th>
<th>Affect—Recognition</th>
<th></th>
<th>Recognition—Affect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rn</td>
<td></td>
<td>Rn</td>
</tr>
<tr>
<td>Affect</td>
<td>Fail</td>
<td>Pass</td>
<td>Fail</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>Pass</td>
<td>.185</td>
<td>.254</td>
<td>.221</td>
</tr>
<tr>
<td>Fail</td>
<td>.265</td>
<td>.296</td>
<td>.215</td>
</tr>
<tr>
<td></td>
<td>.050</td>
<td></td>
<td>.050</td>
</tr>
</tbody>
</table>
### Table 5

**Proportion Correct by Level of Confidence for the Immediate Test (Experiment 1)**

<table>
<thead>
<tr>
<th>Confidence</th>
<th>Affect—Recognition</th>
<th>Recognition—Affect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rn</td>
<td>Affect</td>
</tr>
<tr>
<td>Guess</td>
<td>.56_{104}</td>
<td>.60_{129}</td>
</tr>
<tr>
<td>Half-sure</td>
<td>.54_{186}</td>
<td>.49_{208}</td>
</tr>
<tr>
<td>Sure</td>
<td>.55_{190}</td>
<td>.50_{143}</td>
</tr>
<tr>
<td><strong>Means_{480}</strong></td>
<td>.55</td>
<td>.52</td>
</tr>
</tbody>
</table>

*aThe subscripts indicate the number of subject-items in the particular cell.*
Table 6
Proportion Correct by Level of Confidence for the
Delay Test (Experiment 1)\textsuperscript{a}

<table>
<thead>
<tr>
<th>Confidence</th>
<th>Affect—Recognition</th>
<th>Recognition—Affect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rn</td>
<td>Affect</td>
</tr>
<tr>
<td>Guess</td>
<td>0.51161</td>
<td>0.37144</td>
</tr>
<tr>
<td>Half-sure</td>
<td>0.58228</td>
<td>0.51237</td>
</tr>
<tr>
<td>Sure</td>
<td>0.6191</td>
<td>0.3799</td>
</tr>
<tr>
<td>Means\textsubscript{480}</td>
<td>0.55</td>
<td>0.44</td>
</tr>
</tbody>
</table>

\textsuperscript{a}The subscripts indicate the number of subject-items in the particular cell.

It is interesting that there was no exposure effect: Performance on the affective judgment task hovered around the chance level in all conditions (see Tables 1 and 2). For this reason, a meaningful test of the relation between affective judgment and recognition was not possible. This experiment points up the lack of pervasiveness of the mere exposure effect in that one exposure is not necessarily sufficient to induce enhanced liking. Previous research has focussed on the effects of many exposures and, as a result, there is no data pertaining to single exposures of Turkish words. Perhaps the enhancement of liking through repeated exposures would be more accurately described as an effect of mere exposures.
Experiment 2

Kunst-Wilson and Zajonc (1980) demonstrated that subjects could fairly accurately discriminate between previously presented random shapes and new random shapes on the basis of affective judgments, while performing at chance on a forced-choice recognition test. Kunst-Wilson and Zajonc used this finding as support for the pervasiveness and high sensitivity of the mere exposure effect. Because Experiment 1 did not allow an investigation of the relation between recognition and affect ratings, it seemed appropriate to attempt to replicate Kunst-Wilson and Zajonc's finding.

In particular, the goals of this experiment were to determine whether Kunst-Wilson and Zajonc's original finding could be replicated and, in addition, whether affective judgments are independent of recognition judgments, as Moreland and Zajonc (1977), Kunst-Wilson and Zajonc (1980), and Seamon et al. (1983) have argued. Different levels of illumination were used so that an examination of the relation between affect rating and recognition could be made across different levels of performance -- or at least recognition performance. The issue of independence was addressed through the computation of association coefficients on a subject-item basis.

Method

Subjects. Twenty-four Rice University undergraduates participated, some for pay some for course credit. Subjects were tested either individually or in groups of up to three. For the subjects tested in groups, each subject used a separate computer and was not able to see the stimuli or responses of any of the other subjects.
Apparatus and Items. All stimuli and test questions were presented on an Apple Macintosh computer. The stimuli consisted of 60 eight-sided random shapes provided by Vanderplas, Sanderson, and Vanderplas (1965).

![Figure 1](image)

Figure 1. A sample of the stimuli used in Experiment 2.

Design. The design was a $2 \times 2 \times 3$ factorial with order of test (recognition followed by affect, A-R, or affect followed by recognition, R-A) manipulated as a between-subjects variable, and type of test (affect or recognition) and level of illumination (low, medium, or high) manipulated as within-subjects variables. Half of the subjects were assigned to the R-A condition and half to the A-R condition.

The stimuli were randomly divided into three groups of 20. These three groups formed the pools of shapes for each of three respective trials. For half of the subjects, ten of the shapes in a group served as target items and the remaining ten shapes served as lures. For the other half of the subjects, the target and lure statuses were reversed.
Each trial was presented under a different level of illumination. Under the low illumination condition, subjects were scarcely able to see the shape, and could tell only that an object had been flashed on the screen. This condition was intended to be analogous to the illumination condition used by Kunst-Wilson and Zajonc (1980). For second, or medium illumination, condition the screen was bright enough so that subjects could see the shapes, but not so bright as to allow for easy discrimination. High illumination was obtained by using the screen's maximum brightness. Subjects had no trouble discerning the shapes in the high illumination condition. Across subjects, each illumination condition accompanied the first, second, and third lists an equal number of times. Within the A-R and R-A conditions, three orderings of illumination conditions (low-medium-high; medium-high-low; high-low-medium) were present an equal number of times with each of the shapes.

Procedure. Subjects were told that a series of shapes would be presented on the screen at a fairly rapid rate. They were asked to watch the screen closely, even if they could not easily see the shapes, because they would later receive several tests concerned with the shapes. The nature of the tests was not specified.

The shapes were presented as solid black figures for 16.67 ms, followed by an interstimulus interval of 750 ms, for an onset to onset time of about 768 ms. Each of the ten shapes was presented five times in a random order, hence each sequence consisted of 50 presentations. A tone from the computer signalled the end of a list.

After the stimuli had been presented, subjects made either affective judgments followed by recognition judgments or vice versa. One of the shapes was a member of the previously presented list and the other shape
was not. For the affective judgment, subjects were asked "Which one do you like better?" Next, subjects were asked to rate their judgment on a 3-point confidence scale: 1 — guess; 2 — half-sure; or 3 — sure. Before proceeding to the next pair, the recognition judgment required the subject to indicate which shape was believed to be a member of the previous list. Confidence in the recognition judgment was assessed using the same three-point scale.

For the test, a pair of shapes was presented on the screen under high illumination. In all, each subject saw three lists, for which the order of the judgments was the same. Thus, a pair of shapes appeared on the screen and the subject made both affective and recognition judgments pertaining to the pair before proceeding to the next pair. The reason for this was that, though it is not entirely clear from their description, Kunst-Wilson and Zajonc (1980) seem to imply that they used such a procedure.

The above procedure was repeated for each of the three illumination conditions for each subject.

Results

The recognition test was one of forced-choice and was scored in a straightforward manner. For the affective judgment test, the response was correct if the subject chose an old shape as the one better liked. The means for each condition, broken down according to the order in which the test questions were asked, are presented in Tables 7 and 8.
**Table 7**

*Proportion Correct When the Affective Judgment Preceded the Recognition Test (A-R) (Experiment 2)*

<table>
<thead>
<tr>
<th>Type of Test</th>
<th>Level of illumination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Recognition</td>
<td>.633</td>
</tr>
<tr>
<td>Affect</td>
<td>.558</td>
</tr>
</tbody>
</table>

**Table 8**

*Proportion Correct When the Recognition Test Preceded the Affective Judgment (R-A) (Experiment 2)*

<table>
<thead>
<tr>
<th>Type of Test</th>
<th>Level of illumination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Recognition</td>
<td>.567</td>
</tr>
<tr>
<td>Affect</td>
<td>.508</td>
</tr>
</tbody>
</table>

From an inspection of the means, it appears that the order of the test questions had a small effect upon the level of performance on both tests. In
order to examine these and other observations, a three-way analysis of variance (order of test x type of test x level of illumination) was carried out. The effect of test order was negligible ($F < 1$), as were its interactions with any other effects ($p$'s all greater than .20). Thus, for ease of presentation, the data were collapsed across order of testing and are presented in Figure 2.

![Graph showing the relationship between level of illumination and per cent old items selected for affective judgment and recognition.](image)

**Figure 2.** Mean performance levels in Experiment 2.

As is apparent in Figure 2, there was a main effect of Type of test ($F (1, 22) = 24.183, p < .001, MS_e = 2.646$), reflecting the superior performance on the recognition test. Though it appears that the manipulation of illumination affected recognition performance more than affective judgment, the type of test x level of illumination interaction term failed to reach statistical significance ($F (2, 22) = 2.260, p > .10, MS_e = 3.097$). No other effects or interactions were statistically significant (all $p$'s > .20).
### Table 9
**Contingency Tables for Subject-item Data (Experiment 2)**

**High Illumination**

<table>
<thead>
<tr>
<th>Affect—Recognition Rn</th>
<th>Recognition—Affect Rn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affect</td>
<td>Incorrect</td>
</tr>
<tr>
<td>Correct</td>
<td>.133</td>
</tr>
<tr>
<td>Incorrect</td>
<td>.183</td>
</tr>
<tr>
<td>( \phi )</td>
<td>-.18</td>
</tr>
</tbody>
</table>

### Table 10
**Contingency Tables for Subject-item Data (Experiment 2)**

**Medium Illumination**

<table>
<thead>
<tr>
<th>Affect—Recognition Rn</th>
<th>Recognition—Affect Rn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affect</td>
<td>Incorrect</td>
</tr>
<tr>
<td>Correct</td>
<td>.167</td>
</tr>
<tr>
<td>Incorrect</td>
<td>.142</td>
</tr>
<tr>
<td>( \phi )</td>
<td>-.01</td>
</tr>
<tr>
<td></td>
<td>Affect—Recognition</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------</td>
</tr>
<tr>
<td></td>
<td>Rn</td>
</tr>
<tr>
<td>Affect</td>
<td>Incorrect  Correct</td>
</tr>
<tr>
<td>Correct</td>
<td>.200  .358</td>
</tr>
<tr>
<td>Incorrect</td>
<td>.167  .275</td>
</tr>
<tr>
<td>$\phi$</td>
<td>.02</td>
</tr>
</tbody>
</table>

Thus, while recognition performance was substantially above chance (overall mean performance was 66.9%), the affective judgment task did not prove to be a more accurate discriminator of new and old items (overall mean performance was 53.6%) as Seamon et al. (1983) have suggested. A $t$ test was conducted at each level of illumination to determine whether affective judgment performance was ever above chance level. All tests failed to reach statistical significance ($p's > .10$).

We next turn to an analysis carried out at the subject-item level to obtain a finer grain look at the relation between recognition and affective judgment. The contingency tables and the values of phi for each condition are given in Tables 9, 10, and 11. All values of phi are rather close to zero, indicating a lack of association between an item's statuses on the two tests.
Such independence between the two tests is not very surprising because the level of performance on the affective judgment task was near chance level.

One could argue that combining the data across confidence levels could mask possible associations. To examine this possibility, the data were re-scored for both tests, taking into account the subject's level of confidence. If an item was incorrectly identified then the score for that item was simply the inverse of the level of confidence (i.e., 1, 2, or 3 corresponding to confidence levels 3, 2, and 1, respectively). If an item was correctly identified on a test, then its score was three plus the level of confidence (i.e., 4, 5, 6). In this way, the endpoints of the scale ranged from an incorrect sure answer (1) to a correct sure answer (6). The correlations between the recognition/confidence ratings from two tests are presented in Table 12. Although the product-moment correlations involving confidence are generally greater than the fourfold point correlations, they are by no means large. It is, however, somewhat intriguing that they exceed zero by much at all considering the levels of performance on the affective judgment task.

Table 12

Product-moment Correlations for Recognition-Confidence Measure

<table>
<thead>
<tr>
<th>Order of Judgment</th>
<th>Level of Illumination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>A-R</td>
<td>-.04</td>
</tr>
<tr>
<td>R-A</td>
<td>+.12</td>
</tr>
</tbody>
</table>
In this experiment, then, the results of both Kunst-Wilson and Zajonc (1980) and of Seamon et al. (1983) were not replicated. A failure of replication in the low illumination could be explained in a variety of ways, such as the possibility that the screen was so dim that subjects could not see the shapes at all. One argument against such a suggestion is that recognition performance was not at chance level, but at 60%. Although this experiment used a slower presentation rate than earlier studies (16.67 ms as opposed to 1 ms in Kunst-Wilson and Zajonc (1980) and 2 ms in Seamon et al. (1983)), it is unlikely that this could account for the discrepancies. The results of the high illumination condition are clearly counter to the mere exposure effect. At least with respect to recognition performance, the high illumination condition may be considered approximately equivalent to the exposure conditions used by Seamon et al. Whereas Seamon et al. (1983, Experiment 1) obtained affective performance of 64.5%, in the present experiment affective performance was only 51.7%. To the extent that repeated exposures enhance liking of a stimulus, a greater liking for the presented stimuli as opposed to the nonpresented stimuli would be expected. Yet, such was not the case.
Experiment 3

One variable that might shed some light on the relation between affect and recognition is the number of presentations. Using Japanese characters, Moreland and Zajonc (1977) reported a mere exposure effect that was linearly related to the logarithmic transform of exposure frequency. In the following experiment, Japanese characters were presented a variable number of times after which subjects were asked to make liking ratings of the stimuli. Following the affect rating session, subjects made recognition judgments with confidence ratings. Contingency analyses were again conducted to determine the nature of the relation between the two judgments.

Method

Subjects. The subjects were 20 Rice University undergraduates who participated for either pay or course credit.

Stimuli and Apparatus. The stimuli were 20 seven-stroke Japanese characters randomly selected from Vaccari and Vaccari (1952). All of the stimuli were presented on an Apple Macintosh computer.

Design. The experiment conformed to a 2 x 5 factorial design with type of test (affective judgment or recognition) and exposure frequency (0, 1, 3, 9, or 27 exposures) manipulated as within-subjects variables. For counterbalancing purposes, the 20 characters were arbitrarily divided into two sets of ten. Half of the subjects received one set as stimuli and the other set as lures whereas for the other half of the subjects these statuses were reversed.

The stimuli were exposed at one of five frequencies: 0, 1, 3, 9, or 27 repetitions. The characters were counterbalanced in such a way that each
character was presented at each frequency level equally often across subjects. Of the set of eight to-be-presented characters, two appeared at each frequency level for each subject yielding a sequence of 80 randomly ordered presentations. Note that the items in the zero exposure frequency group were formally equivalent to the lures in that neither were exposed during the presentation phase. Thus, only eight different characters were actually presented. Such an arrangement generates a target to lure ratio of 2:3, the same used by Moreland and Zajonc (1977).

**Procedure.** Before the presentation sequence began, subjects were instructed that they were to watch a series of Japanese characters presented on the screen. No other instructions were given, nor was any mention made of the ensuing tests. Each character was presented for 2 s, followed by a blank screen for 5 s. Thus, the onset to onset time was 7 s. The presentation sequence lasted about 9.3 min.

Next the subjects received an affective judgment test followed by a recognition test. Unlike Experiment 2, subjects made all the affective judgments before recognition judgments.

For the affective judgments, subjects were given a sheet of paper with 20 scales on it and asked to rate each of the 20 characters on a scale of 1 to 7, according to how much they liked each character, with 1 labelled "dislike" and 7 labelled "like". In this phase of the experiment, each character appeared on the screen for 2 s and was followed by a 30 s interval during which subjects circled a number on the appropriate scale. One second before the next character was to appear, a tone alerted the subject.

For the recognition test phase, presentation of the test stimulus proceeded in exactly the same fashion. Subjects were informed that all of
the items they were about to see had been presented in the affective judgment test. However, they were asked to judge whether an item was old with respect to the presentation list, using a 7-point scale in which "1" meant certain new and "7" meant certain old. Therefore, the extent to which a response deviated from the midpoint of the scale (4) indicated the degree of confidence in the old/new choice.

It is important to realize that, for a given subject, the recognition test and the affective judgment task were identical except for the instructions and the response sheet; the same stimuli were presented at the same rate in the same order for both tasks. As an aside, it might be noted that Moreland and Zajonc (1977) reversed the endpoints on their rating scales on alternate items. Because such alternation introduces method-variance into the scores (see Harvey, Billings, & Nilan, 1985), this testing method was not used in the present experiment.

Results

Once again, the analyses concerning subjects' means will be presented first, followed by the subject-item contingency analyses.

The mean level of performance for each combination of conditions is presented in Table 13. For the nonpresented items (i.e., items with a zero frequency of exposure), the means reflect the ratings of twelve characters per subject. For the presented items the means reflect the ratings of two characters per subject.
Table 13
Mean Ratings for the Japanese Characters (Experiment 3)

<table>
<thead>
<tr>
<th>Type of Judgment</th>
<th>Frequency of Exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Affect</td>
<td>4.175</td>
</tr>
<tr>
<td>Recognition</td>
<td>2.375</td>
</tr>
</tbody>
</table>

A two-way analysis of variance was performed on the data. Recognition ratings increased across the different exposures, whereas liking ratings increased for up to three exposures and began to drop off; the interaction between type of test and exposure frequency was statistically significant ($F(4, 76) = 22.34,$ $p < .001,$ $MS_e = 1.050$) in addition to the main effects of both type of test ($F(1, 19) = 9.37,$ $p < .01,$ $MS_e = 2.15$) and exposure frequency ($F(4, 76) = 31.66,$ $p < .001,$ $MS_e = 1.555$). At this gross level of analysis one could conclude, because of the presence of the interaction, that affective judgment and recognition as measured by confidence ratings are not highly related.

If we consider only the data from the affective judgment task, then we may check for the presence of a mere exposure effect. A one-way analysis of variance was performed (exposure frequency was a within-subjects variable) which revealed an effect of exposure frequency ($F(4, 76) = 2.77,$ $p < .03,$ $MS_e = 1.474$).
Trend analyses were conducted separately for the mean affect ratings and recognition ratings. Exposure frequencies were treated as equal interval measures, which is formally equivalent to the use of the logarithmic transform of frequency, as suggested by Moreland and Zajonc (1976, 1977) and Zajonc (1968). Among the affect rating means, there were marginally statistically significant linear \( t(19) = 1.70, p > .10 \), quadratic \( t(19) = -1.75, p > .09 \), and quartic trends \( t(19) = 2.00, p > .06 \); the cubic trend did not approach statistical significance \( t < 1 \). These results contrast with those reported by Zajonc et al. (1974) and Moreland and Zajonc (1976, 1977) in which there were only strong linear components. It is not clear why a strong linear component did not emerge in the present data.

The recognition ratings were subjected to a trend analysis in the same manner. As might be expected, there was a strong linear component \( t(19) = 22.56, p < .0001 \). Also, a quadratic trend was present \( t(19) = -3.85, p < .01 \), probably because all of the items exposed 27 times were selected (i.e., given a rating of 7). No other trends approached statistical significance (for the cubic and quartic trends, \( p > .13 \) & \( p > .56 \), respectively).

To determine at just which exposure frequencies liking for the exposed shapes was enhanced relative to the liking for the non-exposed shapes, pairwise comparisons were conducted in which the mean affect rating for each frequency was compared to the zero frequency of exposure. There was no effect of prior exposure for characters exposed once \( t < 1 \) or 27 times \( t (19) = 1.44, p > .10 \). Compared with no exposures, three exposures of a character enhanced liking substantially \( t (19) = 3.34, p < .01 \) while nine exposures had a marginal effect upon liking \( t (19) = 1.77, .05 < p < .10 \).
These data do not allow one to explain the lack of an exposure effect in Experiment 2.

We now turn to contingency analyses. Because the mere exposure hypothesis was formulated with respect to old items, these contingency analyses will be presented first, using the 160 old subject-items. Judgments of liking were positively correlated with judgments of recognition ($r = +.15$, $p^* < .05$). This means that the more a subject was sure that a character was old, the more he or she liked it. Quite contrary to previous findings (e.g., Zajonc, 1968; Moreland & Zajonc, 1977), the degree of liking was essentially uncorrelated with exposure frequency ($r = +.06$), and slightly correlated with the logarithmic transform of frequency ($r = +.11$, $p^* > .12$). Zajonc (1968) and Moreland and Zajonc (1977) have stated that liking increases in a linear fashion with respect to the logarithm of exposure frequency plus one. In the present study, however, the zero-order correlation between degree of liking and degree of recognition was almost nonexistent ($r = +.02$), indicating a lack of any such linear relation. A trend analysis (including the unexposed items) indicated the presence of a cubic trend of affect rating across the frequencies, $t(396) = 2.49, p^* < .01$. Thus far, the data from Experiment 4 have yielded results that are, in part, inconsistent with the findings of Zajonc and his colleagues.

It could be that the above relation between recognition and liking holds only for old items. To investigate this possibility, contingency analyses were conducted on both new and old subject-items, a total of 400 subject-items.

---

$^2$In these item analyses and the ones in the following experiment, The value of $p^*$ that is provided in the text is the probability level at which one could reject the null hypothesis if all of the observations were independent. $p^*$ should not to be interpreted as a significance level because each subject and each item contributed more than one observation.
Liking was correlated with whether or not an item was previously exposed ($r = +.17, p^* < .001$). This correlation confirms the existence of a mere exposure effect at the subject-item level in that previously presented items were liked more than non-presented items. The relation between liking and the logarithmic transform of exposure frequency is also a critical aspect of the mere exposure hypothesis. There was a positive zero-order correlation between these two variables in these data ($r = +.17, p^* < .001$), indicating the presence of a linear trend.

The relation that is central to this thesis is that between affective judgment and recognition. On an item level, the degree of liking was correlated with the degree of recognition ($r = +.24, p^* < .0001$). Further, degree of liking and degree of recognition are still correlated both when the presentation status of the items (i.e., presented or nonpresented) is partialled out ($r = +.18, p^* < .01$) and when the logarithm of exposure frequency is partialled out ($r = +.17, p^* < .01$). The relation between liking and presentation status, however, disappears when the degree of recognition is partialled out ($r = +.02$). It appears that the mere exposure effect may be mediated by judged status, reflected by degree of recognition, rather than by exposure alone.
Experiment 4

It should be informative to address the issue of association by using a different presentation rate and different stimuli. Stang (1974a) has argued that demand characteristics underlie exposure effects in that subjects may catch on to the manipulation of exposure frequency and attempt to make their answers conform to what they believe to be the experimental hypothesis. To eliminate this possibility in Experiment 4, all of the presented stimuli were exposed the same number of times (three). Through the examination of errors, it was possible to examine the effect on affect ratings of judged status when actual status is held constant. Likewise, the effect of actual status may be determined when judged status is held constant. If judged status were the important factor underlying affective judgment, as suggested by Experiment 3, then actual status would be expected to have a negligible on affect ratings whereas judged status would have a large effect.

Method

Subjects. The subjects were 18 Rice University undergraduates who received either course credit or pay for their participation.

Materials and Apparatus. The stimuli, given in Appendix C, were 36 Turkish words drawn from a Turkish psychology text, as in Experiment 1. All of the words were presented on an Apple Macintosh computer.

Design. The Turkish words were arbitrarily divided into two sets. The words in each set served as stimuli and lures an equal number of times across subjects. For the presentation phase of the experiment, each stimulus word was presented three times, yielding a total of 54 presentations. These presentations were randomly ordered for each subject.
Procedure. The only instructions given to the subjects were to watch each Turkish word as it was presented on the screen in turn. The words were shown for 1 s each with virtually no dead time between one word and the next.

After viewing all of the words, subjects were given a recognition test. Each word appeared on the screen and subjects indicated whether or not they believed it to be a member of the previous list by pressing "y" (for yes) or "n" (for no). Subjects also rated the confidence in each response using a 3-point scale as in Experiment 1.

An affective judgment task followed the recognition test. The words were re-presented in the same order as in the recognition test and subjects were asked to indicate how much they liked each word on a scale of 1 to 7, on which "1" meant that they did not like the word at all, and "7" meant that they liked the word a lot. The subjects entered their ratings into the computer; an index card with the scale and labels was provided lest they forget. No time limit was imposed on responses; the next word appeared as soon as the subject had responded.

Results

As in the previous three experiments, the global analyses will be presented first, followed by the item analyses. The mean proportion correct on the recognition test was rather high: .87 of the old words were correctly identified as old and .79 of the new words were correctly identified as new. Since subjects also provided a confidence rating for each response, it is possible to construct a scale that combines recognition and confidence. This was done such that the low end of the scale (1) indicated that subjects were relatively sure that the item was new and the high end of the scale (6)
indicated that subjects were relatively sure that the item was old. The mean recognition/confidence rating was 5.25 for old words and 2.47 for new words. This difference was statistically significant, \( t(17) = 16.36, p < .0001 \).

Let us now check for the presence of an exposure effect on affect ratings. All of the lures had been exposed in the recognition test; hence, the term "new words" will refer to those words that were not in the presentation list. The mean affect rating was 4.76 for old words and 4.02 for new words. This difference was statistically significant, \( t(17) = 5.96, p < .0001 \). To the extent that the \( t \) value may be used as an index of a test's sensitivity to the status of an item, we see that the affective judgment task is less sensitive than the recognition test. This finding contrasts with the arguments advanced by Kunst-Wilson and Zajonc (1980) and Seamon et al. (1983), who have claimed that affective judgment is the more sensitive test. In this experiment, however, there was a confound with order of testing that may have affected the results.

It is possible to conduct four pairwise comparisons to evaluate the importance of both judged status and actual status in affective judgments. In order to do this, subjects' scores were organized according to the types of errors they made. A variable number of subjects (up to four) were excluded in the following comparisons because they did not make any of the relevant errors. This may cause the mean for a particular condition to be slightly different across comparisons. Also, the mean scores for a particular type of error may be based upon different numbers of observations for each subject.

Given that a subject has judged a word as old, we may wonder whether the actual status of the word plays a role in the subject's affective rating. This question may be answered by comparing the items that subjects
thought were old and were in fact old (old/old) with words that subjects thought were old but were really new (old/new). The mean affective rating was 4.89 for old/old words and 4.62 for old/new words. This difference did not achieve statistical significance, $t(16) = 1.00, p > .30$. In other words, given that a word was believed to be old, the item's actual status did not affect the subject's liking of the word.

It is also possible to investigate the effect of actual status on items judged to be new. That is, if an item is thought to be new but is really old (new/old), is it better liked than an item that is really new and identified as new (new/new)? For new/old items, the mean affective rating was 3.99 and for new/new items it was 3.94; the difference was not statistically significant, $t(13) = .17, p > .80$. Affective judgments did not discriminate between new and old items, when both were thought to be new.

Consider now the effect of judged status with actual status held constant. Of the items that were actually old, those judged to be old (the old/old items) had a mean rating of 4.92, whereas those judged to be new (the new/new items) had a mean rating of 3.99. The difference was statistically significant, $t(13) = 2.86, p < .01$. Of the items that were actually new, those judged to be old (the old/new items) had a mean rating of 4.22, whereas those judged to be new (the new/new) items had a mean rating of 3.86. This difference, too, was statistically significant, $t(16) = 2.85, p < .01$.

$^3$In these comparisons, error classification will be denoted by "judged status"/"actual status".
Table 14
Mean Affect Ratings According to Judged Status and Actual Status (Experiment 4)

<table>
<thead>
<tr>
<th>Judged Status</th>
<th>Actual Status</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Old</td>
<td>New</td>
</tr>
<tr>
<td>Old</td>
<td>4.90</td>
<td>4.53</td>
</tr>
<tr>
<td>New</td>
<td>3.96</td>
<td>3.90</td>
</tr>
</tbody>
</table>

Note. n = 13.

To test for an interaction between judged and actual status with affect rating, a two-way analysis of variance was conducted with judged status (new or old) and actual status (new or old) as within-subjects variables. Five subjects were excluded from this analysis because they did not make all types of errors. The means of the remaining thirteen subjects are given in Table 14. As is apparent from the means, judged status did not interact with actual status \((F < 1)\). There was a main effect of judged status \((F(1, 12) = 19.74, p < .001, MS_e = .4108)\), but not of actual status \((F(1, 12) = 2.00, p > .18, MS_e = .3054)\). Thus, the results of this analysis are congruent with those of the pairwise comparisons.

From the above analyses, it appears that subjects liked items that they believed to be old regardless of whether the items were really old or not. This conclusion is contrary to that of Matlin (1971), who claimed that judged status was only important with respect to the liking ratings of truly old
items. Further, she went on to argue that if judged status were held constant, degree of liking varies with true statuses. Obviously, these conclusions were not supported by the present data.

Finally, we turn to the contingency analysis for recognition and affective judgment. The following analyses are based on all 648 subject-items. Old items were preferred to new items ($r = +.22, p^* < .001$). There was a correlation between affective rating and perceived status ($r = +.29, p^* < .001$) and this correlation remained after the effect of actual status was partialled out ($r = +.28, p^* < .001$). However, the correlation between affect rating and actual status disappeared when the effect of perceived status was removed ($r = +.05, p^* > .10$).

Affect ratings were correlated with the recognition/confidence ratings ($r = +.28, p^* < .001$) and remained correlated when the effect of an item’s true status was partialled out ($r = +.26, p^* < .001$). Conversely, the correlation between affect ratings and status disappeared when the effect of recognition/confidence was partialled out ($r = +.04$).

It appears, then, that the data from Experiment 4 support and extend those of Experiment 3. Indeed, between the two experiments there are no discrepant findings and the magnitudes of the observed correlations in the item analyses are roughly the same. Both experiments point up the importance of judged status in affective ratings.
General Discussion

The experiments described in this thesis were designed to shed light on the relation between affective judgment and recognition memory. Numerous studies have demonstrated that repeated exposures to a stimulus tend to increase liking of a stimulus (e.g., Zajonc, 1968; for a review see Harrison, 1977), and some researchers have argued that this enhanced liking, or "mere exposure effect", occurs irrespective of whether the stimulus is recognized (Matlin, 1971; Moreland & Zajonc, 1977; Zajonc, 1980). In addition, some research has suggested that affective judgment may be a more sensitive test of prior exposure than recognition (e.g., Kunst-Wilson & Zajonc, 1980; Seamon et al., 1983). However, the interpretation of previous studies bearing on this issue has been clouded by inappropriate analyses and the absence of contingency analyses. The experiments reported in this paper lead to rather different conclusions. They are consistent in suggesting that recognition is a more sensitive index of prior exposure than is affective judgment, and that the two are stochastically dependent. This finding leaves open the possibility that recognition may in part underlie affective judgment.

An detailed examination of the relation between affective judgment and recognition was precluded in Experiments 1 and 2 because of a failure to obtain a mere exposure effect. For the immediate tests in Experiment 1, in which Turkish words served as stimuli, the recognition test discriminated between new and old items, whereas the affective judgment test did not, indicating that the recognition test was more sensitive.
The absence of exposure effects in Experiment 1 could be due to the fact that the stimuli were exposed only once. Whereas Zajonc (1968) analyzed Turkish words exposed 0, 1, and 2 times, the ratings were averaged across these exposure frequencies, making it impossible to determine if a single exposure generated an effect. However, Moreland and Zajonc (1977, Experiment 1) found a small and statistically nonsignificant increase resulting from a single exposure. In light of the evidence from Experiment 1 and Moreland and Zajonc's work, it is quite likely that multiple exposures are prerequisite for the effect.

In Experiment 2, random shapes were shown five times each under three different illumination levels (low, medium, & high) in an attempt to replicate the findings of Kunst-Wilson and Zajonc (1980). The affective judgment data showed no evidence of an exposure effect, although forced-choice recognition performance was above chance under all levels of illumination. It appears that, even under the low illumination condition in Experiment 2, recognition was more sensitive than affective judgment.

It is difficult to explain the absence of an exposure effect in Experiment 2, which was patterned after Kunst-Wilson and Zajonc's (1980) experiment. Indeed, the exposure effect obtained by Kunst-Wilson and Zajonc appears to be quite reliable: In a footnote, Kunst-Wilson and Zajonc reported a replication and Seamon et al. (1983) twice reported a mere exposure effect for random shapes given five brief presentations. The 16.67 ms presentation time used in Experiment 2 should not be responsible for the lack of an effect, for exposure effects with random shapes have been found with both shorter (Kunst-Wilson & Zajonc, 1980; Seamon et al., 1983) and longer (Hamid, 1973) presentation times. Perhaps the manner of asking the
recognition and affective judgment questions may play a role in enhanced liking. This possibility arises because in Experiment 2 subjects did not make all of the affective judgments before making recognition judgments. Seamon et al. blocked affective and recognition judgments, Kunst-Wilson and Zajonc were not entirely clear as to the procedure they used, and Hamid did not take recognition measurements at all.

Experiment 3, in which Japanese characters were presented a variable number of times, was similar to an experiment carried out by Moreland and Zajonc (1977, Experiment 1), and yet Moreland and Zajonc concluded that affect and recognition were unrelated. The absence of a linear trend in the affect ratings across exposure frequencies is difficult to reconcile with Moreland and Zajonc's finding of a linear trend. However, Experiment 3 and Moreland and Zajonc's experiment also differ with respect to the pattern of recognition performances: Zajonc (1980) reported that there were a number of instances in which subjects in Moreland and Zajonc's experiment failed to recognize characters presented 27 times, whereas the subjects of Experiment 3 were no less than completely sure of their identification of characters presented 27 times. Unfortunately, there is not a simple explanation for this discrepancy.

In Experiment 4, in which Turkish words were presented three times each, judged status had a greater effect on liking ratings than did a word's actual status. These results conflict with one set of results obtained in a similar experiment carried out by Matlin (1971). Matlin found that judged status was an important determinant of liking for old items, but not for new items. It is unclear why Experiment 3 yielded a pattern of results different from Matlin's. Nevertheless, both Experiments 3 and 4 indicated that the
recognition test was more sensitive to prior exposure than the affective judgment test.

The Affective Judgment Task as a Test of Implicit Memory

Is it necessary then, for all non-memory tests that may reflect the effects of a previous experience to be associated with recognition memory? The answer to this question appears to be no. There are a number of non-memory tests on which performance manifests the effects of a previous experience and yet are substantially independent of recognition test performance. Graf and Schacter (1985) have suggested that tests that do not require conscious recollection be called tests of implicit memory. This is in contrast to tests of explicit memory, which specifically ask that subjects recollect a previous experience. For the sake of convenience, the above two labels will be used to refer to various non-memory and memory tasks throughout the remainder of the paper.

In a sense, the affective judgment task may be considered a test of implicit memory in that subjects are not by any means required to consciously recollect a previous episode in order to perform the task. That is, it is entirely possible to express liking for something without having seen it before, yet previous exposure may affect performance on the affective judgment task. Other research has shown that performance on tests of implicit memory can be, and frequently is, independent with respect to performance on tests of explicit memory (e.g., Jacoby, 1983; Jacoby & Dallas, 1981; Jacoby & Witherspoon, 1982; Tulving, Schacter, & Starck, 1982).

In light of Experiments 3 and 4, it appears that affective judgment is a test of implicit memory that is related to recognition memory. The relation is most apparent with respect to judged status. Thus, as a test of implicit
memory, affective judgment tasks stand in contrast to some other measures of implicit memory that are independent of recognition test performance (Jacoby & Dallas, 1981).

Performance on tests of implicit and explicit memory has been subjected to a variety of theoretical interpretations. Some theorists, most notably Mandler (1980), have advocated a two-factor model of recognition memory in which recognition comprises two stages. Briefly, the first stage is one of familiarity, which Mandler describes as a type of context-free knowing. The second stage of the model consists of identification, in which one attempts to associate the familiar stimulus with a specific context. Mandler goes on to claim that these two stages are additive and determine whether "recognition" occurs.

One point brought out by Mandler (1980) is that the more frequently an item is presented, the more familiar it becomes. How then, does familiarity relate to affective judgment? Like recognizability, enhanced liking increases as stimulus exposure frequency increases. It is possible that affective judgments and recognition judgments in some way rely on the same information, but that affective judgments are more intimately related with familiarity because they do not require that a stimulus be associated with a previous context. This would partially explain why items judged to be old are liked more than items judged to be new even if the items thought to be new have been previously encountered. That is, if liking judgments are in some way rooted in judged status, then subjects need only to rely on feelings of familiarity and not episodic identification.

The above explanation implies that familiarity is somehow basic, or at least more basic than stimulus identification. This notion is contrary to that
advocated by Jacoby (1984), who claimed that "... effects of prior exposure on [perceptual identification] performance are primary, and ... feelings of subjective familiarity are due to performance effects' being attributed to prior study" (p. 152). Thus, Jacoby views familiarity as an inference based on perceptual fluency. More importantly, Jacoby attributed mere exposure effects to perceptual fluency, which he defined as the relative ease with which an item is perceived. Jacoby argued that subjects may have no alternative to using perceptual fluency as a basis for affective judgments.

Is the affective judgment task the only test of implicit memory that is associated with recognition memory? Not necessarily, for there is some evidence of an analogous relation between recognition test performance and perceptual identification of nonwords. The perceptual identification task is another test of implicit memory for which researchers have suggested that performance is independent of recognition performance and, in some cases, more sensitive. The basic perceptual identification paradigm is one in which subjects view a list of stimuli and later attempt to identify visually degraded images of the stimuli. It is generally found that previous exposure to a stimulus facilitates its identification (e.g., Jacoby, 1983; Jacoby & Dallas, 1981; Scarborough, Cortese, & Scarborough, 1977).

A set of results parallel to those of Experiment 4 was obtained by Johnston, Dark, and Jacoby (1985, Experiment 2). In their experiment, subjects pronounced nonwords, one at a time, as they were presented during the study phase. They then attempted to identify visually degraded nonwords, and after each identification attempt they were presented with the nonword again for classification with respect to occurrence on the previously presented list. Johnston et al. found that items judged old were
identified more rapidly than those judged new, regardless of whether the item was in fact old or new. Johnston et al.'s pattern of results is similar to the pattern obtained for performance on an affective judgment task in Experiment 4, in that words judged to be old were liked better than words judged to be new, regardless of their actual status.

Thus, affective judgment is not the only test of implicit memory that, under some conditions, may be associated with recognition. In each of the four experiments reported here, it was clearly the case that recognition was a more sensitive index of prior exposure than was affect rating. Thus, recognition could not have been fully mediated by affect, though of course affect could be fully mediated by recognition.
References


of Experimental Psychology: Learning, Memory, and Cognition, 8, 336-342.


Appendix A
<table>
<thead>
<tr>
<th>BILIM</th>
<th>CANLI</th>
<th>OTELDE</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLABAK</td>
<td>INTIBAK</td>
<td>HAKIM</td>
</tr>
<tr>
<td>YILDAN</td>
<td>EGER</td>
<td>OLMAYA</td>
</tr>
<tr>
<td>BIRAZ</td>
<td>METOT</td>
<td>UZAGA</td>
</tr>
<tr>
<td>VERIMIN</td>
<td>BEYAZ</td>
<td>KAFESEE</td>
</tr>
<tr>
<td>BASMA</td>
<td>HAYATTA</td>
<td>BULMA</td>
</tr>
<tr>
<td>KALID</td>
<td>TAHMIN</td>
<td>SAHIP</td>
</tr>
<tr>
<td>OLSA</td>
<td>BAKIMA</td>
<td>ORTADAN</td>
</tr>
<tr>
<td>KABACA</td>
<td>TABIAT</td>
<td>SEYLER</td>
</tr>
<tr>
<td>TEMAS</td>
<td>ZAYIF</td>
<td>GELISEN</td>
</tr>
<tr>
<td>YASKIN</td>
<td>ADALE</td>
<td>KARSI</td>
</tr>
<tr>
<td>EGILIR</td>
<td>PENCESI</td>
<td>YANI</td>
</tr>
<tr>
<td>FIKRI</td>
<td>DISLERI</td>
<td>SARTLAR</td>
</tr>
<tr>
<td>YASAYAN</td>
<td>KOMASI</td>
<td>POSTU</td>
</tr>
<tr>
<td>RUYALAR</td>
<td>LIMEYI</td>
<td>GIBI</td>
</tr>
<tr>
<td>GEZEN</td>
<td>SAVUNMA</td>
<td>BUGUN</td>
</tr>
<tr>
<td>BALDAN</td>
<td>OTEKI</td>
<td>AMACA</td>
</tr>
<tr>
<td>GORULEN</td>
<td>KADAR</td>
<td>GELEN</td>
</tr>
<tr>
<td>ATALAR</td>
<td>ANCAK</td>
<td>SEKILDE</td>
</tr>
<tr>
<td>BASIT</td>
<td>ADETA</td>
<td>ELINI</td>
</tr>
<tr>
<td>KIMSE</td>
<td>BUYUK</td>
<td>MEYDANA</td>
</tr>
<tr>
<td>ANLAMDA</td>
<td>VARDIR</td>
<td>EFENDI</td>
</tr>
<tr>
<td>AMATOR</td>
<td>AKIL</td>
<td>HALINE</td>
</tr>
<tr>
<td>OLMUS</td>
<td>KENDI</td>
<td>SINDA</td>
</tr>
<tr>
<td>BILGISI</td>
<td>DEFA</td>
<td>SORUYA</td>
</tr>
<tr>
<td>PARAYI</td>
<td>VEYA</td>
<td>VARKIK</td>
</tr>
<tr>
<td>GELMEZ</td>
<td>SADECE</td>
<td>OLDUGU</td>
</tr>
<tr>
<td>INCELER</td>
<td>OLCEK</td>
<td>ONLARA</td>
</tr>
<tr>
<td>OLAYAR</td>
<td>KISIMINI</td>
<td>BUHAR</td>
</tr>
<tr>
<td>SONRA</td>
<td>YAKILAN</td>
<td>FARKINA</td>
</tr>
<tr>
<td>EDILEN</td>
<td>MASAL</td>
<td>HATIRI</td>
</tr>
<tr>
<td>BAGLARI</td>
<td>ESEKI</td>
<td>SAGILIGI</td>
</tr>
<tr>
<td>SORULAR</td>
<td>TARZI</td>
<td>BOGAZ</td>
</tr>
<tr>
<td>GIRER</td>
<td>FELSEFE</td>
<td>ONEMLI</td>
</tr>
<tr>
<td>OLAN</td>
<td>ZIHIN</td>
<td>TARIFA</td>
</tr>
<tr>
<td>NOKTAYA</td>
<td>MANEVİ</td>
<td>TOPLAMI</td>
</tr>
<tr>
<td>YENIDEN</td>
<td>EDER</td>
<td>BOYUNCA</td>
</tr>
<tr>
<td>AZDIR</td>
<td>KOKUNU</td>
<td>UZUN</td>
</tr>
<tr>
<td>MAKTA</td>
<td>ISTER</td>
<td>CALISAN</td>
</tr>
<tr>
<td>METNE</td>
<td>VERILEN</td>
<td>YAPMAK</td>
</tr>
<tr>
<td>CEVABI</td>
<td>TEORIK</td>
<td>DEVAMLI</td>
</tr>
<tr>
<td>DAHA</td>
<td>GRUBUN</td>
<td>ATILMIS</td>
</tr>
</tbody>
</table>
Appendix B
Kanji Characters for Experiment 3

系 希 羊 災 志
歩 辛 成 我 戒
兵 求 孝 赤 究
完 谷 余 含 床
Appendix C
KABACA
ISABET
VERIMEN
HATASI
YASAYAN
RUYULAR
GORULEN
ANLAMDA
BILGISI
DENEMI
BAGLARI
SORULAR
INTIBAK
NAKTAYA
SAVUNMA
KOKUNU
YAPTIGI
GAYEYE
LISELER
MUDURU
CEVABI
ORTADAN
GELISEN
SARTLAR
MEYDANA
OLDUGU
ONLARA
FARKINA
SAGLIGI
ONEYMLI
TARAF
TOPLAMI
BOYUNCA
CALISAN
DEVAMLII
YAPILAN