JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 15, 505–517 (1976)

Context Effects in Recognition Memory for Faces

MICHAEL J. WATKINS, ELAINE HO, AND ENDEL TULVING

University of Toronto

Three experiments showed that a study-to-test change in the presentation context of pictures of unfamiliar faces impairs their recognition. Experiment I showed that a face studied beside a second face was more likely to be recognized when accompanied by the same rather than a different face. In Experiment II, faces were paired with descriptive phrases and again re-presentation of the study context was shown to enhance recognition. Experiment III replicated the finding of Experiment II with a forced-choice rather than a free-choice procedure. It is argued that these findings are consistent with episodic theory but pose a problem for a tagging theory of retention.

Most psychological research on memory has dealt with the retention of experimenter-created words or episodes. Typically the experimenter presents a series of items and subsequently observes the probability, or the speed, with which these presentations are recollected. The class of item chosen for presentation has most often been that of individual words. Although this choice has several advantages at a practical level, it poses a trap of a theoretical sort.

The trap lies in the possibility of confusing an event or episode of a word's occurrence in a to-be-remembered list with the word as a linguistic unit used in communication. As an example of this confusion, researchers almost always talk of remembering a word instead of remembering the occurrence of a word-event. This lapse is transparent and therefore trifling. Less trifling, however, is the assumption that the retention of the fact that a word occurred in a list is somehow secondary to the existence of the word, or of the concept it represents, in a permanent knowledge store.

This research was supported by the National Research Council of Canada Grant Number A8632. Michael J. Watkins is now at Princeton University. Requests for reprints should be sent to Endel Tulving, Department of Psychology, University of Toronto, Toronto M5S 1A1, Ontario.

Copyright © 1976 by Academic Press, Inc.
All rights of reproduction in any form reserved.
Printed in Great Britain

The need to postulate a permanent memory for familiar words, and the concepts to which they refer, cannot be denied. But researchers disagree on the role that a word's permanent representation plays in the retention of an episode in which the word is presented as a to-be-remembered item. According to traditional theory, a person's general knowledge of the world consists in a network of idea nodes, many of which are labeled and correspond to words in the person's vocabulary. Associations between concepts are embodied in the links between the nodes. Within this general framework the assumption is made that memory for episodes is carried in the form of tags attached to the appropriate nodes (e.g., Anderson & Bower, 1972, 1973, 1974; Bahrick, 1969, 1970; Kintsch, 1970). This account of episodic memory may be called *tagging theory*.

TAGGING THEORY AND CONTEXT EFFECTS IN WORD RECOGNITION

The popularity that tagging theory enjoys owes much to its straightforward accounts of the recall and recognition processes. These can be summarized in the following way. When a subject learns a word in a list, a tag indicating its list occurrence is attached to the appropriate node and perhaps certain associations

between that node and others are marked. If the subject is subsequently asked to recall the word a search is made through the associative network for appropriately tagged nodes. On the other hand, if a recognition test is given then no search is necessary, since locating the relevant node is automatically triggered by the "old" word on the test sheet; under these conditions the subject's task reduces to a decision of whether the node is appropriately tagged. Thus, recall involves a *search* (or "generation") process followed by a *decision* (or "recognition") process, whereas in recognition the first of these processes is bypassed. From these assumptions it follows that recognition failure must reflect the absence of an acceptable tag and cannot be attributed to a failure of access. (For a fuller discussion of this point see Watkins & Tulving, 1975a,b.)

A critical feature of this simple version of tagging theory is its assumption that an encounter with an item automatically triggers access to the appropriate node in the associative network. The assumption applies not only when the item is encountered as a member of a to-be-remembered list, but also when it is encountered on a recognition test sheet. This assumption of perfect access rules out the possibility that access is affected by context. The theory predicts, therefore, that if the tag state of the target nodes is held constant, then recognition should be independent of whether the study context is preserved or is changed at test.

This prediction of no context effects in recognition has been tested in many studies. Most of these studies have failed to support the prediction. For instance, Light and Carter-Sobell (1970) presented homographic words in one context (JAM in the context of *strawberry*) and tested for recognition in a context that was either similar (*raspberry* JAM) or different (*traffic* JAM). Recognition proved more probable when the study and test contexts were similar. Subsequent studies have demonstrated recognition context effects with words that are not obviously homographic. Thus, Tulving and Thomson (1971) and Thomson (1972) showed subjects familiar words, either singly or in pairs, and subsequently presented the words as test items in a context either identical with, or different from, that at study. Changes in context were obtained by testing a singly presented word in the context of another word and by testing a word studied as a member of a pair either alone or in a new pair. These experiments showed recognition performance to be reduced by study-to-test changes in context.

The simple version of tagging theory makes the further prediction that if a subject is capable of recalling the presentation of a word then he should also be able to recognize the word. The reasoning here is straightforward: Since the recognition process is included in the recall process, success in recall necessarily implies success in recognition. This prediction also appears to be at variance with experimental data. There are now a large number of experiments showing that, under certain conditions, target words that are not recognized as such can be subsequently recalled with a substantial probability (e.g., Tulving & Thomson, 1973; Tulving & Wiseman, 1975; Watkins & Tulving, 1975a).

The findings of context effects in recognition and of recognition failure of recallable words have led to a revision of tagging theory. The data have been reconciled with the theory by way of the assumption that words can be represented in the associative network by more than one node. This assumption was initially invoked (e.g., Bobrow, 1970) to explain context effects in homograph recognition, and it was subsequently extended (Anderson & Bower, 1974; Martin, 1975; Reder, Anderson, & Bjork, 1974) to account for recognition failure of words that are not obviously homographic. When the to-be-remembered words are homographs the revised version of the theory makes good sense: If the node corresponding to *edible* JAM has been tagged, then recognition will fail if access is gained only to the node for *traffic* JAM. Applied to words that are not

transparently homographic, the revised theory is less attractive, assuming as it does that a word as a linguistic unit typically comprises many distinct meanings or senses. Nevertheless, this assumption has been made and recognition context effects and recognition failure thus brought within the fold of tagging theory: "... the multiple senses of the 'nonhomograph' are distinct, and if one sense is tagged during study, that sense must be retrieved later for successful recognition" (Anderson & Bower, 1974, p. 410).

Words and Faces

Tagging theory has grown out of research in which the to-be-remembered items have been verbal, and hence the question arises of how the theory fares with other kinds of items. For instance, can the tagging account of word recognition be extended to face recognition? More particularly, suppose that on a specific occasion an individual meets someone he has never met or known about before, and that he is subsequently shown, say, a head-and-shoulders photograph and asked whether this is of the same person. His response may be characterized in one of four ways: The test picture is of the person he met and he makes a (correct) decision to this effect—a *hit*. The picture is not of the person met and he makes a (correct) decision to this effect—a *correct rejection*. The picture is not of the person met but he (incorrectly) decides that it is—a *false positive*. The picture is of the person met but he (incorrectly) decides that it is not—a *miss*.

These four classes of response in our hypothetical situation are, of course, also those of a typical recognition test in a word-memory experiment; and with words as the to-be-remembered items each of these outcomes can be readily described within the framework of tagging theory. The interpretation that tagging theory offers when the to-be-remembered items are words, however, does not seem well suited for when the to-be-remembered items are strange faces. We will examine some of the difficulties in extending tagging theory in this way, taking the four response classes in turn.

Hits

The correct recognition of a word as having been encountered on some specified occasion is interpreted in terms of tagging theory by assuming that the node in the knowledge store corresponding to the word (or to a sense of the word) was tagged at the time of encounter, that the tag was retained over time, and that the node was consulted at test and judged to carry an appropriate tag. Now, how can the same interpretation be applied to the correct recognition of a face seen on a single occasion? If a face is encountered for the first time, what can be tagged? It makes little sense to assume that the individual carries in his permanent memory network representations of all the possible faces that might exist in the world, and that the appropriate node is found and tagged for the very first time when a strange face is encountered. More useful, perhaps, is the assumption that encountering a new face results in the creation of a new node in the associative network, complete with appropriate tag. Correct recognition would then involve precisely the same process as the correct recognition of a word, with automatic access to the appropriate node plus a decision that the node carries the right tag. Although tagging theorists have so far said very little about novel, nonverbal information, this approach does seem to be the one advocated by Anderson and Bower (1973) in their *Human Associative Memory*, a sophisticated and detailed version of tagging theory. The weakness of this approach, it seems to us, is that it is somewhat unparsimonious. There is a basic distinction between items that are familiar to the rememberer (e.g., words) and items that are unfamiliar (e.g., strange faces). When a word is presented as a to-be-remembered item a node is *found* in the existing associative network; when a strange face is presented as a to-be-remembered item a node is *created* and

added to the network. The lack of economy has been noted by Kintsch: "Once the necessity of constructing structures *de novo* is admitted, the notion of tagging appears superfluous. Each construction occurs in a certain context and is context specific. Thus, these context specific aspects of a memory code play the same role as a list tag" (1974, p. 79). A more parsimonious view is that the memory trace of a unique event—be it the appearance of a word in a particular place at a particular time, a face, or something else—is always created in the form of a new gestalt.

Correct Rejections

With words as the to-be-remembered items, tagging theory assumes that the correct rejection of a lure results from a decision that the consulted node in the memory network does not bear an appropriate tag. But the rejection of a strange face presented as a lure must involve a different process, since it makes no sense to postulate a decision concerning the tag state of a node that does not exist. Within the tagging framework the correct rejection of a lure face can only be interpreted as a result of a failure of access to any representation of the face in memory. Thus, in terms of the node and network model the rejection of a lure face reflects the absence of an appropriate node, rather than of an appropriate tag.

False Positives

When the to-be-remembered items consist of discrete words, tagging theory can account for a lure being erroneously designated a target item by assuming either that the lure node was inadvertently tagged during the study sequence, or that the decision mechanism accepted an inappropriate tag. But neither of these possibilities can be applied to memory tests involving strange faces, since if the associative network contains no nodes for lures then none could have been tagged and none can carry tags to be incorrectly read. A more promising approach would be to relinquish the assumption that access in recognition is infallible, and to admit the possibility of "mistaken identity," inappropriate access to a node corresponding to one of the faces of the study list.

Misses

With words as the to-be-remembered items tagging theory can account for a failure of recognition simply by assuming that the appropriate tag has become lost or is for some reason unacceptable. At first glance it seems that this account can be extended to cover recognition failure for strange faces. But unfortunately, this approach would leave us with basically different interpretations for the rejection of targets (no tags) and the rejection of lures (no nodes). More important, it makes the curious prediction that a person would claim that he "knows" those target faces he rejects, but believes he recognizes them from a context other than the study sequence.

It thus seems that tagging theory has several problems in handling the retention of encounters with strange faces. Although the theory can explain the correct recognition of these encounters by assuming that nodes can be instantaneously constructed and added to the associative network, complete with encounter tags, such an interpretation is different from that for recognizing encounters with words. Problems of parsimony and consistency become even more acute when tagging theory is applied to the case of an encounter with a strange face being forgotten, to the case of a strange face being falsely recognized, and even to the case of a face never encountered being correctly judged as unfamiliar.

Context Effects in Recognition of Faces

The problems that remembering encounters with unfamiliar items poses for tagging theory are not restricted to the various outcomes of a simple recognition situation; they also extend to context effects. It was noted earlier that context effects in word experiments can be explained within the framework of tagging

theory by assuming that words are typically represented by more than one node in the memory network and that recognition may fail because of access to the wrong node. But the same explanation cannot be applied when the remembered item was previously unfamilar and thus presumably without representation in the memory network. The study encounter may result in the construction of a single new node but at test there can be no possibility of access to the wrong node. Context cannot bias the selection of nodes if there is only one node that is relevant. A second encounter with the target item should therefore trigger access to the appropriate node regardless of whether the original study context is preserved or changed. It would seem, therefore, that whether or not context effects occur in the recognition of previously unfamiliar faces is an important question in evaluating tagging theory. The principal purpose of this article is to describe three experiments that demonstrate such context effects.

Before describing these experiments it is perhaps appropriate to say something about an alternative way of thinking of the retention of an event in memory. This approach, which may be called *episodic theory*, has been proposed and developed in a number of recent articles (Tulving, 1972, 1976a; Tulving & Thomson, 1973; Watkins & Tulving, 1975a) and is similar to a theory proposed by Kintsch (1974). The starting point of episodic theory is the distinction between a person's memory for autobiographical episodes and his memory for facts about the world, the distinction between personal history as recollected and general knowledge. Episodic theory holds that every experience that a person has is unique in the sense that he has not had an exactly identical experience before nor will he have one again. When a person experiences an event, a trace of the episode is formed. This is true regardless of the nature of the event; the event may be the occurrence of a familiar and meaningful word in a particular list in a particular situation, or a glimpse of a strange face in a crowd. Although information in semantic memory plays an important role in the construction of an episodic trace (e.g., Tulving, 1976b), retention of the trace is thought of as being separate from semantic memory. In this sense episodic and semantic memory are independent. Recollection of an event will occur only when the information extracted from the retrieval cue matches or complements the information in the trace (Kintsch, 1974). The information extracted from the retrieval cue depends, among other things, on semantic properties of the cues. Thus, both trace construction and retrieval are affected by semantic memory, again regardless of the nature of the event that the trace represents. Moreover, the principles underlying the recollection of an event are the same whether memory is tested under conditions conventionally referred to as free recall, cued recall, recognition, or whatever. In free recall, the rememberer has to rely on general retrieval information present in his cognitive environment. In cued recall and recognition the person is given additional retrieval information, directed at specific traces (Tulving & Watkins, 1973). In each case the event will be recollected to the extent that the retrieval information overlaps with that in the trace.

Of particular relevance to the present concern is the idea that recognition is governed by the same principles as recall. The only significant difference is that the recognition test includes a nominal copy of the study item, which formed the focal element of the target episode. The study item was, however, not the only component of the target episode. A person's experience when presented with an item will necessarily be of an item in a particular context. It follows that the probability of a test item being recognized as nominally identical to a study item should vary directly with the extent to which there is a reinstatement of the item's original presentation context, since normally the effect of this reinstatement will be to increase the informational overlap between the cue and the stored trace. The

question of whether there is an effect of context in the recognition of previously unfamiliar items, such as strange faces, is thus of considerable theoretical importance. On the one hand, tagging theory would seem to be incompatible with such context effects, while on the other hand, episodic theory assumes that recollection is always susceptible to context effects.

Strong context effects in the recognition of novel materials could presumably be demonstrated by the simple expedient of using materials that are grossly ambiguous. Thus, for instance, if that well-known ambiguous picture *the wife and the mother-in-law* were presented in a context likely to induce, say, a "wife" interpretation, then subsequent recognition is probably more likely when the test context induces the same bias (cf. Wiseman & Neisser, 1974). In like vein, Bower and Holyoak (1973) showed that ambiguous sounds are better recognized if given the same verbal label at presentation and at test. Unfortunately such findings are not particularly critical to the present discussion. To demonstrate context effects in the recognition of grossly ambiguous materials does not necessarily mean that context effects occur in the recognition of unambiguous materials. More important, such demonstrations pose no real problems for tagging theory. For instance, context effects in recognizing the *wife and the mother-in-law* could be mediated by nodes corresponding to "young" or "beautiful," or "old" or "ugly."

There is little evidence for context effects in the recognition of unambiguous, unfamiliar items such as strange faces. We are aware of only a single relevant experiment, one by Bower and Karlin (1974, Experiment 3). These authors concluded against context effects in face recognition. This conclusion is compatible with a tagging theory of episodic memory and appears to be inconsistent with episodic theory. However, Bower and Karlin accepted the null hypothesis on the basis of a rather small number of observations. Moreover, their results were not entirely inconsistent with the conclusion that context effects were present in their experiment, even though they did not compel it. Thus, since there is sufficient uncertainty about Bower and Karlin's findings, and since their conclusion is of some theoretical importance, another look at context effects in face recognition seemed called for.

Experiment I

The first experiment sought to demonstrate that recognition of a face is susceptible to a study-to-test change in context, where context takes the form of another face. The design was essentially similar to that of Bower and Karlin's experiment, and indirectly similar to Tulving and Thomson's (1971) and Thomson's (1972) word experiments, except that in the present experiment only one method was used for manipulating context.

Method

All subjects studied a sequence of 80 pairs of pictures of unfamiliar faces. They were induced to study the faces as pairs, rather than individually. The study session was followed by a test session which also involved the presentation of 80 face pairs. The left-hand, or context, members of these pairs were the left-hand members of the study pairs, though arranged in a new order. The right-hand face of a test pair (*a*) had appeared in the study list paired with the same context face, the *Same Context* condition, (*b*) had appeared in the study list paired with a different context face, the *Changed Context* condition, or (*c*) had not appeared at all in the study list and was thus a lure. The question of interest was whether face recognition would be better in the *Same Context* condition than in the *Changed Context* condition.

Materials and design. A total of 200 black-and-white, head-and-shoulder portrait photographs were selected from a recent college year-book. Those selected were of people considered difficult to distinguish verbally.

Thus all were white; none wore spectacles, earrings, beards, or the like; and apart from a modest difference between the sexes, all were clad alike. There were an equal number of males and females. Duplicate sets of slides were made from the selected photographs—one for the study sequence, the other for the test sequence.

Of the 200 faces, 24 (12 male and 12 female) were paired to form two 6-pair practice lists. Of the remaining 176 faces, 16 (eight male and eight female) were randomly selected as lures in the recognition test for the critical list. This left 160 faces for presentation in the critical list. These faces were formed into 80 pairs. One member of each pair was designated the *context* face to be shown on the left half of the display field, the other the *target* face to be shown on the right. Pairing was random, save the restriction that male–male, female–female, male–female, and female–male pairs were equally represented.

The recognition test of the critical list, like the study sequence, involved the presentation of 80 face-pairs. In fact the context (left-hand) faces were merely a rearrangement of those shown in the study list. The right-hand faces included 16 not seen before, and 64 that had appeared as right-hand members in the study list. Of these 64 target faces, 32 were paired with the same context faces as in the study list to form the *Same Context* condition; the other 32 were randomly re-paired to form the *Changed Context* condition. The 32 targets that served in the *Same Context* condition for half of the subjects served in the *Changed Context* condition for the other half, and vice versa.

Subjects. Thirty-two University of Toronto undergraduates of both sexes participated for pay.

Procedure. The subjects were tested individually. They each saw and were tested on two set-establishing lists and a critical list. The study sequence for all three lists involved the presentation of pairs of faces; the faces were shown side by side via two projectors at a rate of 6 sec per pair, with an exposure time of just over 5 sec. The subjects were told about the form of the tests for the set-establishing lists. In these tests all faces retained their original left–right position; half of the pairs were precisely as seen in the study list, the other half were formed by re-pairing. The subjects' task was to indicate which pairs had been studied as a pair and which had been re-paired.

For the critical list which followed, the subjects were told that the number of pairs presented would be greatly increased, but they were led to believe that another pair-recognition task would follow. After presentation of the critical list, the subjects were told that their primary task was to try to recognize which right-hand members of the test pairs had been presented in the study list irrespective of whether its present pairing was repeated or new. It was explained that while the left-hand faces consisted of a re-ordering of those seen in the study list, some of the right-hand faces had not been presented before. Thus, for each right-hand face the subjects indicated on their response sheet whether or not they recognized it from the study list; they then added a confidence rating of their decision, using a 3-point scale. Three measures were taken to reduce the possibility of the context faces being ignored during the recognition decision. (*a*) The subjects were advised that consideration of the left-hand faces would improve their recognition for the right-hand faces. (*b*) In the test sequence, presentation of the left-hand faces began one second ahead of the right-hand faces. (*c*) The subjects followed each recognition decision with an indication of whether or not the test pair had appeared as a pair in the study list. The subjects proceeded through the test sequence at their own pace, taking on average about 20 min.

Results and Discussion

The results were straightforward: Recognition in the *Same Context* condition was superior to that in the *Changed Context* condition. The *Same Context* superiority

was manifested in (*a*) a higher hit rate (.73 vs. .68), $t(31) = 2.71$, $p < .01$; (*b*) a higher mean recognition confidence rating, obtained by scoring the subjects' responses according to a 6-point scale from 1 ("definitely new") to 6 ("definitely old"), (4.46 vs. 4.28), $t(31) = 2.79$, $p < .005$; and (*c*) a higher proportion of items given a maximum ("definitely old") rating (.35 vs. .31), $t(31) = 2.07$, $p < .025$. Thus, although the superiority was not large, it was statistically reliable. The false positive rate was .25, and the mean confidence rating for lure items was 2.29.

In short, Experiment I shows that the probability of recognizing a photograph of an unfamiliar face can depend on whether the study context is present at the time of recognition. This conclusion was reached with a physical manipulation of context that was quite modest. It therefore seems that, at least when mode of study is appropriate, context effects in recognition can be shown with previously unfamiliar as well as familiar materials.

Experiment II

The results of Experiment I complement findings of context effects in word recognition (e.g., Tulving & Thomson, 1971; Thomson, 1972). Experiment II sought to extend both sets of findings by demonstrating transmodality context effects. Specifically, the procedure was similar to that of Experiment I, but with verbal rather than pictorial contexts. Thus, the context faces of Experiment I were replaced with short descriptive phrases. The finding of an influence of a verbal context on the recognition of an unambiguous nonverbal item would further extend the generality of the phenomenon of context effects in recognition memory.

Method

The method was based on that of Experiment I. Short descriptive phrases were randomly paired with photographs of faces. Subjects studied a long series of phrase–face pairs in anticipation of a pair recognition test. In fact, the subjects were tested for their recognition of just the faces. Some test faces were presented in the context of the respective phrases with which they had been studied, the *Same Context* condition, and others were re-paired with phrases from the study list, the *Changed Context* condition. Again the question of interest was whether face recognition would be better in the *Same* than in the *Changed Context* condition.

Materials and design. Ninety-six photographs of male faces were chosen from the set used in Experiment I. Eighty phrases were composed, each a description of a personality, hobby, profession, or the like (e.g., keeps tropical fish; works in cancer research; drives an Italian sports car; a Civil Rights activist; cheats in examinations). These phrases were randomly paired with 80 of the faces, to give a study list of 80 phrase–face pairs. In the test session the 80 phrases were presented in a new order, and again each was paired with a face. For each subject, 32 phrases were paired with the same faces as in the study list, the *Same Context* condition; 32 were re-paired with faces from the study list, the *Changed Context* condition; and 16 appeared with new faces. As in Experiment I, the set of 16 study faces replaced at test was the same for all subjects, but there was between subject balancing in the allocation of pairs to the *Same* and *Changed Context* conditions. Twelve further phrase–face pairs were formed for two 6-item set-establishing lists.

Subjects. Sixteen University of Toronto undergraduates participated for pay.

Procedure. Subjects were tested individually. They were first given two short set-establishing lists, intended to encourage integration of the faces with their respective phrases. Each of these lists comprised six phrase–face pairs. Testing involved presenting three of the study pairs plus three pairs formed by re-pairing the phrases and faces of the remaining three study pairs. The subject's task was one of distinguishing these two types of pairs.

For the critical list the subject received a four-page booklet containing 80 numbered phrases. He studied each of these phrases in the context of a particular face, just as he had done in the set-establishing lists. For each pair, the experimenter read out the number of the current phrase and the subject was given 3 sec to read the phrase, after which the face was presented for a further 3 sec. This procedure was repeated without a break for all 80 phrase–face pairs.

Immediately following presentation of the critical list, the subject received a new four-page booklet containing the same 80 phrases used in the study booklet, though in a new order. Presentation of the test pairs was much the same as that for the study pairs, except that this time the subject worked at his own pace. He was told that most but not all of the test faces had appeared in the study list, and that his principal task was to consider each face in turn and to decide whether or not it had been shown in the study list. To ensure that the faces were studied in the context of the phrases, the subject preceded each recognition decision with a judgment of whether or not the phrase had been paired with the test face at study. Thus the testing procedure entailed the presentation of 80 phrase–face pairs, and for each pair the subject first read the phrase and then studied the face, he decided first whether the phrase was studied with the face at study, and second whether the face had been presented at all in the study list. He recorded his decisions in his test booklet opposite the appropriate phrase, and added for each decision a confidence rating, using a 3-point scale. The test procedure took, on average, about 20 min to complete.

Results

Again the results showed that recognition was better in the *Same Context* condition than in the *Changed Context* condition. The advantage of preserving context was shown in (*a*) a higher hit rate (.84 vs. .71), $t(15) = 4.69$, $p < .001$; (*b*) a higher mean confidence rating (5.07 vs. 4.51), $t(15) = 13.35$, $p < .001$; and (*c*) a higher proportion of items given a maximum confidence rating (.57 vs. .37), $t(15) = 7.16$, $p < .001$. The false positive rate was .22 and the mean confidence rating for lures was 2.52.

Thus the results of Experiment II indicate a substantial influence of a verbally presented context on the recognition of unambiguous visual forms previously seen only once. More generally they suggest that the influence of context on recognition memory reflects a basic property of the episodic memory process.

Experiment III

The first two experiments have shown that for certain study conditions a person will be less likely to declare a test picture to be a member of a previously studied set if its context is changed from study to test. One possible implication of this finding is that an item out of context induces the person to use greater caution in his recognition response without reducing his ability to discriminate it from one not previously encountered. More particularly, it is possible that in the first two experiments the subject's recognition response was somehow directly influenced by his decision of whether the context had changed. That is, a negative response to the context question could have biased the subject to make a like response to the recognition question. Experiment III tests this possibility by using a forced-choice recognition test to ensure a response criterion that is independent of context condition. The procedure was otherwise similar to that for Experiment II, with two exceptions: The subjects knew during study that they would be tested for recognition of the faces as such, and the effect of preserving study context was contrasted with the effect of deleting rather than changing context.

Method

Materials and design. The materials consisted of 200 photographs used in Experiment

I, and a set of 110 short, descriptive phrases that included those used in Experiment II. Fifty-five male and 55 female faces were selected and paired with the 110 phrases to form the study list for all subjects. The first and last 10 pairs were not tested—they served merely to help reduce the level of recognition for the remaining 90 study items.

The test list comprised 90 pairs of faces, formed by randomly pairing each of the 90 critical faces from the study list with a new face of the same sex. Presentation order of the test pairs was random, as was the left–right display position of the study and lure faces within a pair. There were two conditions of testing. In the *Context* condition the test pair was presented along with the phrase with which the target face had been paired in the study list; in the *No Context* condition the test pair was presented alone. All subjects received 45 pairs under each of the two testing conditions; the condition of a given pair was balanced between subjects. The ordering of the test pairs was random with respect to condition.

Subjects. The subjects were 12 male and female University of Toronto undergraduates who were paid for participating.

Procedure. Subjects were presented a long sequence of pictures of faces, and they were told that their ability to recognize the faces would later be tested. Each successive face was presented together with a unique descriptive phrase. To ensure that the faces were studied in the context of their respective phrase, the subjects were given the task of rating, on a 4-point scale, how well the phrases fitted the faces. (They had surprisingly little difficulty in following these instructions.) Presentation rate for the study list was 5 sec per face. The phrases were read aloud by the experimenter with the onset of the faces, and the subject was required to record his compatibility rating before presentation of the next phrase–face pair.

The test session was given immediately following the study session, and it involved the presentation of a sequence of pairs of faces. The subject was told that in each pair one face was from the study list and one was new; his task was to circle on his response sheet either "left" or "right" depending on the position of the target face and to add a rating of his confidence in his decision, using a 3-point scale. For half of the pairs, the experimenter read aloud the phrase with which the target face had been paired in the study list. The subjects were told to attend to the phrases, "because they might help you remember." They proceeded through the test sequence at their own pace.

Subjects were tested individually, and took on average about half an hour to complete the experiment.

Results and Discussion

Once again the results revealed that the faces were better recognized when the study context was preserved. In the *Context* condition 84.4% of the targets were circled, while in the *No Context* condition only 72.4% were circled; this effect was highly significant, $t(11) = 5.69$, $p < .001$. The confidence ratings told the same story. Thus, with ratings converted to a 6-point scale—from lure circled with maximum confidence (rating of 1) to target circled with maximum confidence (rating of 6)—the mean rating was higher in the *Context* (4.83) than in the *No Context* condition (4.25), $t(11) = 9.51$, $p < .001$. Similarly, the proportion of target faces circled with maximum confidence was greater in the *Context* condition (49% vs. 19%), $t(11) = 8.68$, $p < .001$. In fact, all subjects performed better in the *Context* condition on each of these three measures.

Since Experiment III differed from Experiment II in a number of ways, its results serve to further demonstrate the generality of the phenomenon of context effects in face recognition. In particular, Experiment III has shown that context effects in face recognition may obtain even when the recognition decision criterion is held constant.

Discussion

The experiments reported here demonstrate pervasive context effects in face recognition. Thus context effects were found regardless of whether (*a*) contexts were provided by other faces or by verbal descriptions, (*b*) the context was changed or deleted, and (*c*) the subjects were aware or unaware of the nature of the memory test. Moreover, recognition failure occurred when subjects were constrained in their recognition decision criterion, suggesting that context effects do not result merely from greater caution in declaring recognition for an item out of context. A change in context appears to lower not only the probability that an item will be declared recognized, but also the probability that it can be discriminated under conditions in which response criterion is held constant.

The findings of the three experiments described here are at variance with those reported by Bower and Karlin (1974). There are at least four possible reasons for the discrepancy. In the first place, Bower and Karlin manipulated context with the aid of other faces, and a comparison between the results of our experiments (Experiment I versus Experiments II and III) suggests that faces may be less effective than verbal material in manipulating context. Second, it is possible that Bower and Karlin's study task—a judgment of whether the faces within each pair were of people who would be friends—did not suffice to bias the encoding process in a way conducive to context effects. Although Bower and Karlin adopted their task specifically "to maximize the chances of finding a context effect on recognition memory" (p. 754), their task may not in fact have been as effective in this respect as the matching task used in our Experiment I. Third, Bower and Karlin may have failed to observe context effects because their experiment was not sufficiently sensitive. The number of observations made in their experiment was small. In particular, in the critical (compatibility judgment) condition the number of pairs of pictures studied and tested under the same context was five per subject and, since there were 12 subjects, the total number of observations was only 120. This compares with over 1,000 in Experiment I described in this report. Fourth, the statistical evidence in Bower and Karlin's study would have lent support to the hypothesis that recognition is superior in the same context condition if the authors had chosen to test for an *advantage* of preserving context. They tested for an *effect* of preserving contest, and so adopted a two-tailed rather than a one-tailed alternative hypothesis. A test of the advantage of preserving context on recognition confidence rating would have led to the rejection of the null hypothesis at a conventional level of significance.

Consider now the implications of the research reported here for tagging theory. How would tagging theory interpret context effects in the recognition of once-encountered faces? As noted earlier, the original version of the theory (e.g., Anderson & Bower, 1972; Bahrick, 1969, 1970; Kintsch, 1970) was modified to accommodate context effects in recognition memory (Anderson & Bower, 1974; Martin, 1975; Reder *et al.*, 1974), but this modification does not seem to help when the to-be-remembered items are strange faces rather than familiar words. As we noted earlier, in the revised theory a word is typically represented in the associative network not by one node but by several. The theory can therefore interpret the finding that a previously encountered word may be identified in one context but not in another by assuming that the test word provides access to only a subset of item nodes whose choice is determined in part by context. Whether the test word is recognized depends on whether the chosen subset includes a node with the appropriate tag.

An explanation of this sort does not seem to be readily applicable to recognition context effects for strange faces or other items which, before their single study occurrence, are not

represented in the existing memory network. To cope with such context effects, the theory requires additional assumptions. Thus, it could be assumed that seeing a strange face results in *several* nodes being added to the network, and that only a subset of these nodes would be tagged with particular occurrence information and linked with nodes corresponding to the experimentally manipulated context. The theory thus modified would be compatible with the data reported here, though it is worth noting that it makes the prediction that rejected targets, being represented by nodes in the network, should be perceived as generally familiar; and in this respect the subject should be able to distinguish these items from correctly rejected lure items which, being without any representation at all in the network, should be perceived as entirely unfamiliar.

For episodic theory, the experimental data reported here pose no particular difficulties. On this theory, the same principles apply regardless of whether the nominal to-be-remembered item is familiar or not. What is remembered is not the item, but the episode in which the item was presented in some particular context. Whether the item is a common word, a strange face, or anything else, an occurrence of the item is always unique, and so is the resulting memory trace. Thus, while a sharp distinction between an item and the context of its occurrence is convenient at the level of experimental description, there is no such distinction at the level of the memory trace: The trace will be of a unitary item-in-context. Whether the item-in-context is recollected depends on how well the information in the retrieval environment matches the information in the trace. The probability of a successful match in a recognition situation therefore varies with the extent to which the item's context at test resembles its context at study. And again, in predicting such context effects, episodic theory makes no distinctions between different sorts of items. The present findings of context effects in face recognition, like earlier findings of context effects in word recognition, are therefore quite compatible with episodic theory.

References

Anderson, J. R., & Bower, G. H. Recognition and retrieval processes in free recall. *Psychological Review*, 1972, **79**, 97–123.

Anderson, J. R., & Bower, G. H. *Human associative memory*. New York: Wiley, 1973.

Anderson, J. R., & Bower, G. H. A propositional theory of recognition memory. *Memory & Cognition*, 1974, **2**, 406–412.

Bahrick, H. P. Measurement of memory by prompted recall. *Journal of Experimental Psychology*, 1969, **79**, 213–319.

Bahrick, H. P. A two-phase model for prompted recall. *Psychological Review*, 1970, **77**, 215–222.

Bobrow, S. A. Memory for words in sentences. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 363–372.

Bower, G. H., & Holyoak, K. Encoding and recognition memory for naturalistic sounds. *Journal of Experimental Psychology*, 1973, **101**, 360–366.

Bower, G. H., & Karlin, M. B. Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology*, 1974, **103**, 751–757.

Kintsch, W. Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

Kintsch, W. *The representation of meaning in memory*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1974.

Light, L. L., & Carter-Sobell, L. Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 1–11.

Martin, E. Generation-recognition theory and the encoding specificity principle. *Psychological Review*, 1975, **82**, 150–153.

Reder, L. M., Anderson, J. R., & Bjork, R. A. A semantic interpretation of encoding specificity. *Journal of Experimental Psychology*, 1974, **102**, 648–656.

Thomson, D. M. Context effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 497–511.

Tulving, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.

TULVING, E. Ecphoric processes in recall and recognition. In J. Brown (Ed.), *Recall and recognition*. London: Wiley, 1976a.

TULVING, E. Rôle de la mémoire sémantique dans le stockage et la récupération de l'information épisodique. In S. Ehrlich & E. Tulving (Eds.), *La mémoire sémantique*. Paris: Bulletin de Psychologie, 1976b.

TULVING, E., & THOMSON, D. M. Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, 1971, **87**, 116–124.

TULVING, E., & THOMSON, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, **80**, 352–373.

TULVING, E., & WISEMAN, S. Relation between recognition and recognition failure of recallable words. *Bulletin of the Psychonomic Society*, 1975, **6**, 79–82.

WATKINS, M. J., & TULVING, E. Episodic memory: When recognition fails. *Journal of Experimental Psychology: General*, 1975a, **104**, 5–29.

WATKINS, M. J., & TULVING, E. Recall and recognition: A reply to Light, Kimble, and Pellegrino. *Journal of Experimental Psychology: General*, 1975b, **104**, 37–38.

WISEMAN, S., & NEISSER, U. Perceptual organization as a determinant of visual recognition memory. *American Journal of Psychology*, 1974, **87**, 675–681.

(Received June 4, 1976)