Journal of Experimental Psychology:
Human Learning and Memory

VOL. 6, NO. 5 SEPTEMBER 1980

Evidence Against a Semantic–Episodic Distinction

John R. Anderson
Carnegie-Mellon University

Brian H. Ross
Yale University

The possible empirical bases for the semantic–episodic distinction are examined. Two kinds of results are particularly relevant. The first concerns whether the same functional laws can be demonstrated for semantic as for episodic memory. Failure to demonstrate such generalities argues for the distinction, whereas success at demonstrating generalities argues against the need for such a distinction. The second kind of diagnostic result concerns whether one can demonstrate transfer between semantic and episodic memory. Failure to find transfer argues for a separation, whereas evidence for transfer argues against any real separation between the memories. A review of the literature relevant to these two criteria yields considerable evidence against the semantic–episodic distinction. Three experiments are reported that provide new evidence against the distinction. The experiments test whether learning episodic material interferes with the retrieval of semantic information. The dependent measure was the time to respond true or false to a categorization statement (e.g., "A spaniel is a dog."). Before participating in that task, subjects learned sentences that provided information about the item and category. The results indicated that the time to make the semantic judgment was affected by the item–category relation in the prior study sentences. Generally, the study effects, as well as some other findings, were correctly predicted by the ACT theory, which makes no semantic–episodic distinction. However, contrary to the ACT theory, it was found that learning additional information about a concept does not interfere with making positive semantic memory judgments and possibly facilitates.

Tulving (1972) proposed a distinction between semantic and episodic memory. Episodic memory is an autobiographical memory for specific events and is temporally dated. Semantic memory is essential for language use and contains organized knowledge about symbols and concepts detached from autobiographical reference. Tulving suggested that these two memories may be governed by different principles and proposed several other possible differences between them, including the mode of reference (autobiographical vs. cognitive) and retrieval characteristics. Though in his 1972 article he stated that this distinction is not necessarily a functional one, it has often been taken to be such (Atkinson, Herrmann, & Wescourt, 1974; Kintsch, 1975; Lockhart,

This research was supported by National Science Foundation Grants BNS76-00959 and BNS78-17463.

Both authors contributed equally to this article; the order of authors was determined by a coin toss. We thank Bill Jones and Ed Smith for their comments on this manuscript. We also thank Barbara Adams for taking responsibility for the programming, execution, and analysis of Experiment 3.

Brian Ross is now at Stanford University.

Requests for reprints should be sent to John R. Anderson, Department of Psychology, Carnegie-Mellon University, Pittsburgh, Pa. 15213.

Craik, & Jacoby, 1976; Shoben, Wescourt, & Smith, 1978; Tulving, 1976; Watkins & Tulving, 1975) and has gained considerable acceptance. The semantic–episodic distinction has not been accepted as implying a functional difference by many associative and neoassociative models of memory (see Anderson & Bower, 1973). Such theories are inclined against accepting this distinction because of their empiricist leanings, their emphasis on the continuity of knowledge, their rejection of the synthetic–analytic distinction, and their belief that all knowledge derives from experience. None of these factors compel one to reject the distinction, but they do predispose one. Our purpose in this article is not to question the utility of the semantic–episodic distinction as a conceptual heuristic, but rather to question whether there is really any functional difference between the memory that stores semantic knowledge and the memory that stores episodic knowledge.

In this article, we will try to accomplish a number of objectives relevant to evaluating the semantic–episodic distinction. We review the state of the literature with respect to the issue of whether there is a functional distinction between semantic and episodic memory. We argue that the state of the literature does not support a functional distinction between semantic and episodic memory. The two types of memory appear to obey the same laws and to be so highly interdependent that they are best considered one memory system. We present a series of three experiments from a fact recognition paradigm that demonstrates the strong interdependence of semantic and episodic memory. In addition to addressing the semantic–episodic distinction directly, we consider how well a particular memory theory called ACT (Anderson, 1976) fares in accounting for the results of our experiments. This theory assumes that all information is stored in a single memory and so expects a high degree of interdependence between facts supposedly in "semantic memory" and facts supposedly in "episodic memory." This theory is interesting because it not only predicts a high degree of interdependence, it predicts the nature of the interdependence. As such it will serve to bring out other interesting aspects of the data. The reader should keep in mind, however, that the semantic–episodic issue and ACT provide somewhat independent perspectives on the data. For instance, it might well turn out that there is no functional distinction between semantic and episodic memory and that the ACT theory of memory is wrong, too.

The ACT Theory

Before considering the relevant literature, it is useful to review the ACT theory (Anderson, 1976). Long-term memory in ACT is represented by a network of propositions interconnecting concepts. The relations between concepts are encoded in the propositional structure connecting the concept nodes. All information, semantic and episodic, is stored together. This does not mean that there are no differences among the memory traces. The various possibilities that Tulving (1972) suggested, such as autobiographical tags and differing richness of connections, are represented. However, the ACT theory does not require a strict correlation of trace characteristics as envisioned by Tulving. So, for example, a particular trace may be temporally tagged as are episodic memories but be richly connected as are semantic memories. In this model, there is no functional distinction between semantic and episodic information in terms of storage, retention, or retrieval. It is basically all one big memory.

One extensive application of the ACT theory has been in understanding how subjects recognize facts like *A spaniel is a dog.* To verify a particular proposition, the nodes representing the concepts in the proposition are activated and a search is started from all nodes. The search process consists of the activation spreading from the nodes down all links until there is an intersection. At that time, the intersection is evaluated to determine whether it matches the proposition to be verified. Quillian (1968), Collins and Quillian (1972a), and Collins and Loftus (1975) have proposed similar search mechanisms. The spread of activation from a node is assumed to be a limited-capacity parallel process (see Townsend, 1974), so that the more links connected to a node, the slower the search along each link. In addition, links may vary in strength, with stronger links being traversed more quickly. The time to verify a proposition, then, depends on how strongly the concepts are connected relative to the other connections attached to each concept.

Deciding that a particular proposition was not studied or is unknown may be accomplished in several ways. One proposal (Anderson, 1976, chap. 8; King & Anderson, 1976) is that subjects decide that a proposition has not been studied after waiting for a period of time in which no intersection of activation occurs. The reasoning behind this strategy is that if no correct intersection has been found after one would normally have been expected, no such intersection probably exists. It is possible for spurious intersections to occur that will slow the rejection process, since the subject will have to evaluate the spurious path of intersection and then reinstitute the waiting process. We do not mean to imply that this waiting process is the only means by which subjects reject propositions, only that it is an important process. Alternate rejection processes have been proposed for network models (Anderson & Reder, 1974; Collins & Quillian, 1972b; Holyoak & Glass, 1975).

Evaluation of the Semantic–Episodic Distinction

Some objections to the semantic–episodic distinction have been made on logical grounds. In particular, Schank (1975) argues that the distinction is a false one, since all conceptual knowledge must be obtained through experience. He believes instead that most of memory is an experientially based conceptual memory. Ortony (1975) pointed out some flaws in Schank's arguments, such as the conceptual confusion between knowledge *of* experience and knowledge *from* experience. Even though both semantic and episodic information may be obtained from experience, as Schank claims, only episodic memory is hypothesized to contain knowledge of experience. We think Ortony convincingly refutes Schank's arguments against a semantic–episodic distinction, but he is much less persuasive when providing evidence in favor of the distinction.

There are two kinds of empirical results that are particularly diagnostic in evaluating the semantic–episodic distinction. One is whether similar memory effects are obtained in experiments on episodic memory as in experiments on semantic memory. Similar effects would be evidence that the two types of knowledge obeyed similar functional laws and would be evidence against a functional distinction. Different effects would be evidence for the distinction. The second result that would be diagnostic is whether one could obtain transfer (positive or negative) of learning between episodic and semantic knowledge. This would be evidence against the separation between the two memories. We will review the relevant evidence under these two headings.

Note that it might be possible that the two systems will display similar functional laws but no transfer. This might indicate two physically separate memory systems operating according to the same principles. It is also possible that there might be different functional laws but many transfer effects. This would indicate two structurally interconnected systems operating according to different principles. Of course, the clearest results would be to find both the same laws and transfer or neither.

Similar Memory Effects

Interference

Tulving (1972) suggested that semantic memory might be impervious to the kinds of interference effects that have been demonstrated over and over again in episodic memory. Episodic traces are remembered only through their temporal tags, so they are very vulnerable to interference. Semantic memories, however, are embedded in rich cognitive structures, allowing access by many means and hence providing protection from interference.

Of particular importance to this article is the class of interference experiments (Anderson, 1974, 1976; Hayes-Roth, 1977;

Thorndyke & Bower, 1974) showing that the more episodic facts subjects learn about a concept, the slower they are to recognize any of these episodic facts about this concept. This result is predicted by the limited-capacity activation process within the ACT theory. The rate at which a fact is activated from a concept is a function of the strength of that fact relative to the other facts. Thus, the more facts learned, the slower the activation.

Normally, relative strength and number of interfering facts will covary, but they can be decorrelated. When they are decorrelated, subjects are as fast to verify a fact about a category with more facts as a fact about a category with fewer, if the two facts are of same relative strength (see Anderson, 1976, pp. 285–290, for a demonstration). On the other hand, if one holds a number of facts constant, subjects are slower to verify the less frequently studied (hence weaker) facts. If one uses production frequency norms as a measure of relative associative strength, there is ample evidence that associative strength affects verification time in semantic memory (Glass, Holyoak, & O'Dell, 1974; Loftus, 1974; Smith, Shoben, & Rips, 1974; Wilkins, 1971). There is also evidence that controlling for relative strength, there is no effect of category size (Freedman & Loftus, 1971) —as ACT would predict. So, when we define interference effects in terms of relative strength, we get analogous effects both for semantic and episodic memory.

By this reasoning, it should be the case, but it seems not to have been directly tested, that the mean time to verify a semantic memory fact about a category should be a function of number of facts in the category. That is, the mean time to verify that one of the 12 mo. is a month should be faster than the mean time to verify that one of the hundreds of birds is a bird. This result is predicted because the larger the category, the lower the mean relative strength. There have been some studies that have shown weak but positive effects of category size (e.g., Landauer & Meyer, 1972; Wilkins, 1971), and studies that have failed to find such effects (Freedman & Loftus, 1971). However, these studies have not systematically sampled members of larger categories to get a mean time. Sampling from larger categories has been biased in favor of more frequent members.

Shoben, Wescourt, and Smith (1978) reported a pair of experiments that they interpreted as showing that these kinds of interference effects are not found in semantic memory. They had subjects verify semantic facts like *sparrows have feathers* and varied the number of questions they asked about the concept *sparrow*. They argued that this manipulation was the analogue of the episodic manipulation of number of learned facts. They found no effect of this variable. However, their negative conclusion may be at least partially due to how they operationalized the analogy. The better analogy to number of learned facts in the episodic experiment would be the total number of semantic facts known about *sparrow*, rather than the number tested in the experiment. The ACT predictions for the Shoben et al. task differ from the predictions for experiments in which the number of facts learned about a concept are varied.

ACT predicts[1] that the important variable should be total number of facts, and only with much repeated testing should the number of facts tested become important. Repeated testing should result in an increase of strength for tested propositions, relative to the ones not tested. Once there is sufficient

[1] The following is a summary of the ACT predictions for the Shoben et al. (1978) experiment (See Anderson, 1976, chap. 8 for details of the ACT analysis of RT): Let S = the baseline strength of each fact. Let k = the total number of facts known about a concept. Let n = the number of facts tested in the experiment, the variable. Let p = the number of times a fact has been tested. The strength of the fact will increase with p. The strength of each of the n facts tested p times is $S + p$. The strength of each of the $k - n$ nontested facts is S. So, the RT to a tested fact will be inversely proportional to its relative strength.

$$\mathrm{RT} = a\left[\frac{n(S + p) + (k - n)S}{S + p}\right] = a\left[n + \frac{(k - n)S}{S + p}\right].$$

For small values of p relative to S, RT will be largely determined by k. As p gets larger, differences among conditions will get larger with longer times for higher values of n.

practice that the experimental strength is large relative to the preexperimental strength, it will be important how many experimentally strengthened propositions there are competing. Thus, the ACT prediction is that there should be an interaction between the number of facts tested about the concept and the number of times they are tested, although it is unclear how much testing would be required to get the interaction. The second experiment of Shoben et al. shows some effect in this direction, but it was not significant.

Relatedness Effects

In the semantic memory literature, there have been numerous demonstrations of the effect of semantic relatedness (e.g., Collins & Loftus, 1975; Collins & Qullian, 1972a; Glass et al., 1974; Meyer, 1970; Schaeffer & Wallace, 1970; Smith et al., 1974). This factor refers to the extent to which the subject and predicate terms are related and has been measured by direct ratings or production norms. The general result is that high relatedness speeds the confirmation of true sentences and slows the disconfirmation of false ones. There is no consensus on how to interpret the positive effect for true sentences. High relatedness correlates with production frequency and other variables that might be better thought of as measures of associative strength (see Anderson & Bower, 1973, chap. 12), and in fact, production frequency can be a better predictor of verification time than direct ratings of semantic relatedness (see Smith et al., 1974). The semantic relatedness effect has been interpreted in terms of associative strength (Anderson, 1976; Anderson & Bower, 1973; Anderson & Reder, 1974; Glass et al., 1974), which has been shown to have strong effects in episodic memory (see Anderson, 1976, chap. 8 for a review).

The relatedness effect with falses has been more problematic for many associative analyses. However, such an effect is expected by the spreading activation process in ACT because related falses should result in more spurious intersections. For instance, the reason that a *bat is a bird* is slower than a *bat is a rock* is because irrelevant intersections of *bat* and *bird* (e.g., through *wings*, *animal*) slow down the rejection process. The subject must consider these sources of intersection before rejecting the assertion. King and Anderson (1976) showed analogous effects of spurious intersections in episodic memory on rejecting statements about an episodic data base. McCloskey (Note 1) has shown that semantic relatedness will also slow down an episodic judgment. In McCloskey's experiment, subjects, after studying *Fred is a rabbi*, were slower to reject *Fred is a priest* than *Fred is a banker*.

Shoben et al. (1978) reported a result that they interpreted as indicating that episodic memory does not show a relatedness effect. They had subjects study true, highly related facts like *tigers have stripes*; true, less related facts like *tigers have ears*; false, related facts like *tigers have fingers*; and false, unrelated facts like *tigers have cars*. Subjects were to commit all such facts, true or false, to memory. In a later recognition test, they were asked to discriminate such facts from unstudied, true statements like *tigers have teeth* and false statements like *tigers have wings*. Shoben et al. were concerned with the studied items. They claimed that if episodic memory were like semantic memory, there would be a positive effect of relatedness on trues and a negative effect of relatedness on falses.

However, this does not correspond to the ACT predictions for studied items. In all studied cases, subjects have to associate subject and predicate together in committing the items to memory. Therefore, relatedness should help in establishing this association and retrieving it. Lack of relatedness should only help in the case in which the subject can reject a probe on the basis of no connection between subject and predicate as in the case where the subject is judging the probe *tigers have cars* for the first time. However, in their experiment, a studied probe like *tigers have cars* was neither novel nor was the subject supposed to reject it. Seven of the eight comparisons that they reported ($p < .05$, by sign test) support ACT's prediction that subjects should be uniformly faster with related studied probes.

Encoding Specificity

The research on encoding specificity is frequently presented as relevant to the semantic–episodic distinction (see Flexser & Tulving, 1978; Kintsch, 1974; Postman, 1975; Tulving & Thomson, 1973; Watkins & Tulving, 1975). This class of results involves demonstrations of the context dependency of memory. For instance, a word that can be remembered in one context will not be remembered in another context. This context dependence is interpreted as showing that one does not have access from a semantic trace (the word in general) to an episodic trace (the specific word). A major problem with these findings as evidence for the semantic–episodic distinction is that the results can be interpreted in other ways that do not require a semantic–episodic distinction (see Anderson, 1976, chap. 10). For instance, Anderson argued that a word has many memory structures (senses, elaborations) associated with it and that context determines which structures can be accessed. Recall depends on accessing at test the same structures that were accessed at study. Under this explanation, context dependency arises from retrieval difficulties within memory in general, not from a semantic–episodic distinction. Consistent with this other interpretation is the result of Muter (1978), who demonstrated a similar context dependency in memory for the names of historical figures. His results seemed to indicate that such context dependency can be found within semantic memory and does not depend on a "gap" between semantic and episodic memory.

Transfer Between Semantic and Episodic Memory

There are two directions for transfer effects between semantic and episodic memory. Demonstrations of a transfer of semantic knowledge to episodic memories need not be problematical for the semantic–episodic distinction. Tulving (1972) recognized that if episodic traces were to contain meaningful information, they would have to be affected by semantic memory, since it contains the information essential to comprehension. There have long been findings that the retrieval of episodic information is affected by semantic factors, such as meaningfulness (see Hall, 1966, chap. 10, for a partial review). These results do not argue against Tulving's distinction because they require a dependence of episodic memory on semantic memory only in the creation of the episodic trace. Once established, this trace may not be affected by semantic memory.

A recent experiment indicates that semantic and episodic memory may not be independent after the initial encoding. Perlmutter, Harsip, and Myers (1976) measured reaction time to recall the response of a paired associate. They found that the cued recall time was affected by the word frequency (negatively) of the stimulus and the strength of its primary preexperimental associate. This finding shows the effect of semantic variables (frequency and associative strength) on the speed of retrieving an episodic memory. It is difficult to see why such semantic information would be encoded in the episodic trace.

The other transfer possibility involves the effect of episodic information on the retrieval of semantic knowledge. According to the proposed distinction, access to information in semantic memory may be affected by the organization of semantic knowledge, but it should not be influenced by what is stored in episodic memory. (Of course, semantic memory must be capable of being changed in response to experience; however, it never has been spelled out how these changes take place.) If learning episodic material facilitates or interferes with the retrieval of semantic information, the distinction is weakened. If the transfer effects are similar to those found with solely episodic information (e.g., negative transfer in interference paradigms), there is good evidence against a functional basis for the distinction. The experiments to be reported test this possibility.

A few studies have examined whether episodic information facilitates or interferes with semantic memory. Unfortunately, the results have been mixed. Slamecka (1966) found that subjects' ability to recall free associates to a stimulus was not affected by learning other experimental associates.

A pilot experiment reported by Collins and Quillian (1972b, pp. 134–136) also failed to find an effect of episodic information on semantic memory retrieval time.

A recent study, however, found effects of episodic memory on semantic retrieval time. Lewis and Anderson (1976) tested whether experimentally learned information would interfere with preexperimental historical knowledge. They had subjects study artificial facts about famous people and measured the time to verify a well-known fact about a famous person (e.g., "Washington crossed the Delaware."), verify artificial facts studied, or falsify re-paired foils. The pertinent results are that the time to verify a well-known fact increased with the number of artificial facts learned about the individual, and this was true in both pure test blocks (only real facts are true items) and mixed test blocks (both real and artificial facts as true items).

The semantic–episodic boundary is not always clear, however, and historical knowledge may be considered by some to be in the gray area. It may not possess the rich interconnections posited of semantic knowledge nor be free of temporal dating.In the present experiment, this possible objection was eliminated. The semantic–episodic distinction was tested by examining whether the transfer effects found with episodic information apply to the kind of knowledge that might underlie language use. The semantic material was categorical, or set-inclusion, knowledge in which judgments are made as to whether an item is a member of a category, such as "A spaniel is a dog." The issue is whether studying episodic facts about these concepts affect later semantic verification times. This type of knowledge was chosen because in addition to its use in comprehension, it has been the focus of many experiments (Collins & Quillian, 1969; Holyoak & Glass, 1975; Meyer, 1970; Smith et al., 1974) and has been important in semantic memory theorizing.

Priming

There has been considerable recent research on priming effects (e.g., Fischler, 1977; Meyer & Schvaneveldt, 1976; Neely, 1977), and it has often been assumed that these priming effects are confined to semantic memory. Processing of one item is facilitated if that item is preceded by a semantically related item. Although most of the research on priming used connections supposedly in semantic memory, there is now evidence for semantic priming using episodic connections. The experiment of King and Anderson (1976) can be seen as displaying such an effect. The clearest demonstration of priming with episodic connections comes from the research of McKoon and Ratcliff (1979), which directly compared priming via episodic and semantic connections both for lexical decision and item recognition tasks. In both tasks, equal effects were obtained from priming by semantic connections and by episodic connections.

Summary of the Literature

This survey of the literature is somewhat mixed, but on the whole it is against a functional distinction between semantic and episodic memory. There are numerous demonstrations of similar effects in episodic and semantic memory. The experiments of Shoben et al. (1978) claim to show the opposite, but we have argued against some of their conclusions. There are a number of demonstrations of transfer between semantic and episodic memory that are difficult to reconcile with the distinction. There are some failures to find transfer, but, since these are null results, they should be weighted less than the positive results. The encoding specificity effects have been advanced as a demonstration of lack of transfer. Though these results are not based on accepting the null hypothesis, the interpretation of encoding specificity is in dispute and similar effects have been obtained entirely within semantic memory. The verbal-learning literature in transfer (Kjeldergaard, 1968) is full of failures to transfer within episodic memory. There seems no reason to suppose that the semantic–episodic distinction provides any special barrier to transfer.

Of the two ways to address the semantic–episodic distinction, to look for similar memory effects or to look for transfer from one memory to the others, we feel that the transfer results are more persuasive in arguing against an episodic distinction. The research in this article is directed at showing how episodic experience affects semantic judgments. We feel that this is the most powerful sort of demonstration given general assumptions that semantic memory is impervious to influence from episodic memory.

We were also interested in whether we would get transfer effects as predicted by the ACT theory (Anderson, 1976), which does not make a distinction between semantic and episodic memory. The ACT theory provides a concrete realization of a theory that makes no distinction. The analysis of Shoben et al. (1978) comes closest to a concrete theory for these experiments, which instantiates the opposite point of view and we will frequently refer to it. Although the issue is more general than these specific theories, these theories will serve to bring some focus to a somewhat vague but still important general issue.

Experiment 1

The general purpose of the first experiment was to see whether the time to retrieve semantic information could be affected by the prior learning of episodic information. The episodic study sentences involved the concepts that the subject would later have to verify in semantic probes. These study sentences varied with respect to their relation to the semantic probes. Following the memorization of these study sentences, subjects made categorization judgments of the test pairs. The dependent measure of interest is the reaction time (RT) to these item–category pairs. If semantic and episodic memories are functionally separate, learning episodic material should not affect later retrieval of semantic information. In addition to this study manipulation, two other independent variables, test speed and test block, were orthogonally varied to check some further predictions of ACT relevant to the semantic–episodic distinction.

The study condition variable is central to testing the semantic–episodic distinction. Examples of the five conditions for true and false test pairs are shown in Table 1. For a given subject, each test pair was assigned to one of these five conditions. In the control condition, no facts were studied about the item or category. The other four study manipulations involved learning information (adding links) about the item and category. In the practice condition, subjects studied a sentence that gave information relevant to the categorization judgment. For the true pairs, subjects studied the exact item–category information to be tested later. This manipulation should facilitate the judgment by strengthening the relevant proposition's links. For false pairs, subjects learned a sentence in which the item–category subset relation was explicitly

Table 1
Examples of Materials Used in Experiment 1

Study condition	Test pair: True: Spaniel–Dog	Test pair: False: Rose–Insect
	Sentences studied	
1. Control		
2. Practice	A spaniel is a dog.	A rose is not an insect.
3. Interference action verb	A spaniel retrieves a ball. A plumber pets a dog.	A worker smells a rose. An insect buzzes noisily.
4. Interference copula verb	A spaniel is not an elephant. A collie is a dog.	A rose is not a drawer. A bee is an insect.
5. Spurious connection	A spaniel sniffs a dog.	An insect flies around a rose.

Note. Each subject saw the sentences for only one study condition for each test pair.

negated. Though this condition gives information that allows the subject to make a correct categorization judgment, it is not clear that the subject uses such information in making a false judgment. For instance, none of the falsification processes mentioned earlier would make use of this information. In fact, by adding links to each concept node, it is possible that this manipulation might slow down the search process. Glucksberg and McCloskey (Note 2) reported that explicitly studying that something is not the case will slow subjects down in deciding that it is not the case.

For the two interference conditions, subjects learned two sentences, one containing the test item and one containing the test category. This manipulation should lead to an increase in RT, since the addition of an irrelevant link off each concept node slows down the rate of activation. The difference between the two interference conditions was whether the verb was copula (e.g., "is") or an action verb. The primary purpose of this manipulation was for design reasons, but it also serves to test a prediction of ACT that the speed to verify or falsify a proposition does not depend on the verb type.

The final study condition, spurious connection, connected the item and category as in the practice condition, but this time by an irrelevant fact. For false test pairs, this condition was expected to be especially interfering, since it results in an extra intersection that must be evaluated, in addition to adding a link off each node slowing search time. For true responses, the prediction is more complicated. Though here, too, the search time for the correct intersection is slowed down and a spurious intersection may occur, true responses may also be facilitated by this manipulation. It is possible that additional intersections between the two nodes will lead to more fast guesses (responses made without evaluating the activated intersection), which happen to be correct for true test items. For false items, fast guesses of this type are errors. For true items, then, the spurious connection condition has both interfering and facilitating effects, and the relative magnitudes of these counteracting influences are not certain.

In summary, the theory that makes no semantic–episodic distinction predicts that the study manipulation should have an impact on categorization times. At least some of the semantic–episodic theories predict no effect (e.g., Shoben et al., 1978). Furthermore, some non-semantic-episodic theories such as ACT predict a particular ordering on the judgments. For true responses, the ordering, from fastest to slowest, is practice, control, and interference, with the ranking of spurious connection uncertain, though slower than practice. For false responses, the ordering is control, interference, and spurious connection, with the exact ranking of the practice condition uncertain, though no slower than the interference and spurious connection conditions.

The second variable was test speed. For both true and false materials, item–category pairs were classified as fast or slow, depending on the categorization RT in an earlier experiment (Anderson & Reder, 1974). This division was made with the constraint that the lexical decision RTs were equal in the two groups, so that the speed difference was presumed to be due to memory search time. This manipulation allows a test of the ACT explanation for the results of relatedness on verification of true statements, as discussed earlier. The ACT model attributes this result to the greater strength of the connection between the item and category concepts. The ACT model assumes that the rate of spread of activation from a node along a particular link is a function of the strength of that link relative to the sum of the strengths of all links attached to that node. Since search time depends on proportional strength, ACT predicts that the strengthening and interfering effects of study will be smaller for the more strongly connected test pairs than for the weakly connected ones. We assume that our faster probes are more strongly associated and would receive higher relatedness ratings. So for true responses, there should be a Study Condition × Test Speed interaction, with larger study effects for the slow test pairs.

The test speed manipulation also allows a test of ACT's explanation for the fact that relatedness results in slower falsification times. ACT's explanation for the slow RTs is that strong but irrelevant connections between the two concepts (e.g., both bats and birds have wings) result in spurious intersections that must be evaluated. We assume that slower false items have more spurious intersections. If so, the spurious connection condition should not affect the slow pairs much (which already have spurious connections), but it should increase the occurrences of spurious intersections for the fast pairs. The interference condition should also reduce the difference between the fast and slow pairs. Since adding links to concepts slows down activation rate, this manipulation will decrease the probability of spurious intersections for the slow items before waiting time is up. For false responses, then, a Study Condition × Test Speed interaction is also predicted, but it should be the reverse of the true response case, with a larger study effect now expected for the fast pairs than the slow ones.

The final independent variable was test block. All test pairs were presented once in each of four test blocks. ACT and probably other non-semantic-episodic theories predict definite changes with blocks. For true test pairs, each categorization judgment results in a strengthening of the relevant connection. As this relevant proposition becomes stronger, it will come to largely determine the RT. The effects of study manipulation, test speed, and their interaction will decrease. Also, the RT should get faster with blocks due to this strengthening and due to general speedup of encoding and response. Thus, there should be a main effect of block, and block should interact with the study condition and the Study Condition × Test Speed interaction.

Practice should also serve to attenuate the effects for false test pairs. If a waiting process was used, the period of time for waiting should decrease with practice (as explained in Anderson, 1976, chap. 8), reducing effects of the other variables. So for both true and false test pairs, responses should get faster with blocks, and the differences between study conditions and their interactions with test speed should decrease. This interaction with block is not predicted by semantic–episodic theories like that of Shoben et al.

In summary, the principal purpose of this experiment is to test whether there is a functional basis for the semantic–episodic distinction. If the study manipulations affect categorization time, this will be evidence against the distinction. The findings will be more persuasive if they are the ones predicted by ACT, a model in which there is no distinction between semantic and episodic information. This experiment also tests some further predictions of the ACT theory concerning categorization search processes (Study Condition × Test Speed interaction) and the effects of test repetition (block main effect and interactions). Such effects and interaction would be further evidence against a semantic–episodic distinction.

Method

Materials and design. This experiment employed categorical test pairs and related study sentences. The 40 item–category pairs in the test phase were chosen from the ones used by Anderson and Reder (1974). That study obtained many measures on the stimuli, including item–category verification times (e.g., apple–fruit), item–category falsification times (for re-paired items and categories. e.g., apple–clothing), and lexical decision times for both items and categories. Ten pairs were assigned to each of four test conditions, determined by the correct response (true or false) and the test speed (fast or slow): true fast, true slow, false fast, and false slow. Thirty-nine of the 40 items and categories were paired exactly as they had been in the Anderson and Reder study. Because of an incompatability between their material and the constraints of this design, one pair (carrot–bedding) had to be used for which there was no falsification data available. Because of the lack of relatedness between subject and predicate, we assumed that it would be a fast false. All of the materials are given in Appendix A.

For the test speed classification, a measure of time for memory activation was desired. The pairs were divided such that the item–category verification (trues) or falsification (falses) times for the fast and slow groups were far apart with nonoverlapping ranges, whereas the lexical decision times were as close as possible. This presumably should roughly equate encoding and memory access time while distinguishing the groups on memory activation time. For the trues, the mean verification times for the fasts was 888 msec, and it was 1,063 msec for the slows; the combined item and category mean lexical decision times were 653 msec for the fasts and 668 msec for the slows. For the false test pairs, the mean falsification times were 927 msec for fasts and 1,103 msec for slows, with mean lexical decision times of 670 and 668 msec.[2] In addition to the 40 experimental stimuli shown during the test phase, 10 buffer pairs were included.

For each of the 40 critical item–category pairs, there were five possible study conditions, which are illustrated in Table 1. Sentences for each of the study conditions were constructed in the following way: The item and category from pairs in the control condition were not in any sentences during the study phase. The item and category from pairs in the practice condition were in study sentences that contained information relevant to the categorization judgment, either the exact relation to be tested (for trues) or its negation (for falses). The interference conditions consisted of two sentences, one with the item and one with the category. The two conditions differed in whether the verbs in the two sentences were copulas or action verbs. The spurious connection condition consisted of presenting the item and category in the same sentence but the sentence was irrelevant to the later categorization task. All study sentences were in the present tense and used indefinite articles. It was necessary to present 48 study sentences to establish the various study conditions for the 40 item–category pairs. In addition, there were 12 buffer sentences, made from the buffer test pairs, that were shown to all subjects.

Altogether there are 20 experimental conditions (2 values of truth × 2 values of speed × 5 study conditions). Two of the 10 pairs for each Truth × Speed combination were randomly assigned to one of the 5 study conditions. This random assignment was done independently for each subject.

Subjects. Fifty adults, 17–34 years old, from the New Haven area participated in the experiment in groups of 1–4. Most were Yale undergraduates. Six additional subjects were eliminated because of computer failure (1), low accuracy in the study phase (3), or low accuracy in the test phase (2). Subjects were paid $5 for the experiment which lasted 1½–2 hr.

Procedure. Study and test materials were presented, and response times were collected by a PDP 11/40 computer with four VT50 terminals, each in a separate room. The YEPS system (Proudfoot, 1978) was used, allowing multiple subjects to be run independently at one time.

In the initial study phase, the 60 study sentences (12 buffers followed by 48 critical sentences) were presented 1 at a time on the screen for 15 sec. The subject was told to memorize the sentences by encoding them as meaningfully as possible, imagining a situation in which someone might have said each sentence.

In the next study part of the experiment, the completion phase, the first few words of each sentence (the sentence subject) were presented, and the subject typed in the remaining part. Feedback was given, and, if the response was correct, the beginning of another sentence was presented 3 sec after the feedback. If the response was incorrect, the full sentence was displayed and could be studied until the subject pressed the return key. This procedure was used to allow the subject as much study time as desired, since mistakes may have been due to typographical errors as well as memory failures. The 48 critical sentences were presented for completion in four random orders, preceded each time by a randomized ordering of the 12 buffer sentences. This study procedure was self-paced and took about 70 min (range = 50–110 min).

On completing the study phase, the subjects were given instructions on the test phase and began after a short break. They had not previously been told anything about this part of the experiment. Item–category pairs (e.g., spaniel–dog) were presented on the screen with two blank spaces between the words. If the item on the left was a member of the category on the right, the subject was to press the *yes* button. Otherwise, the subject was to press the *no* button. The subject sat in front of the terminal keyboard and was told to respond as quickly as was consistent with being accurate. The K key on the keyboard was used as the *yes* button and the D key as the *no* button. To be sure that the subject understood, the program halted after three practice trials to allow the subject to ask questions about the procedure. Feedback was given after each trial for 2 sec, followed 1 sec later by the next trial. The 40 critical item–category pairs were presented in four random orders, with each order preceded by a random ordering of the 10 buffer pairs. Between each pass there was a minimum 30-sec rest period that the subject could extend if more rest was desired. The whole test phase lasted about 20 min.

Results

The 50 subjects were 97% correct in the final pass through the completion phase. So it seems reasonable to conclude that they had learned the study material. The real interest in the experiment is in the verification phase. For each subject, RTs for correct responses in each condition (Study Manipulation × Test Speed × Block × Truth) were averaged. To avoid effects of extremes, times longer than 2 sec (less than 1% of data) were set equal to 2 sec. In this

[2] We divided material into groups of slow and fast items rather than high and low similarity because we thought reaction time was a more theoretically neutral basis for assignment. The contrasting theories of the relatedness effect would prescribe different rating measures as the best way to divide the material. In fact, the groups of items did differ in terms of similarity. Using the ratings of Anderson and Reder (1974), the fast trues had a mean rating of 5.97 and the slow trues had a rating of 5.15 on a scale from 0 to 7, with 7 denoting greatest prototypicality of instance to category. This difference was significant, $t(79) = 3.51$, $p < .001$, using the variance estimate from the 80 items in Anderson and Reder (1974). The slow falses had a rating of 1.83, and the fast falses had a rating of 1.21. This was also a significant difference, $t(79) = 2.04$, $p < .05$.

Table 2
Reaction Time (in msec) in Experiment 1 Collapsed Over Test Blocks

Test item	Study condition: Control	Practice	Interference	Spurious connection
		Trues		
Fast	672 (.005)	639 (.018)	670 (.005)	659 (.003)
Slow	741 (.023)	664 (.005)	724 (.026)	720 (.023)
		Falses		
Fast	785 (.013)	762 (.013)	778 (.009)	776 (.025)
Slow	803 (.050)	777 (.048)	799 (.036)	813 (.068)

Note. The interference condition means are based on a maximum of 800 observations (800 minus the number of errors), and all other means are based on a maximum of 400 observations. Error rates are in parentheses.

way long times would have effect on the mean times without any single time having an overwhelming effect. The RT means and error rates, collapsed over all four blocks, are presented in Table 2. The overall standard error for each Study × Test × Block cell was 16.7 msec for the true responses and 21.4 msec for the false responses.

The RT means and error rates for Blocks 1 and 4 are given in Table 3. The standard errors for the true responses were 21.4 msec in Block 1 and 14.1 msec in Block 4. For the false responses, the standard errors were 25.3 and 20.5 msec. The overall error rate was .022. For false responses, the error rates were positively correlated with RT across conditions ($r = .580$). For true responses, the range of error rates was very restricted (.000–.040), and there was a small negative correlation ($r = -.099$). The times for the two interference study conditions have been collapsed in Table 2, Table 3, and in all analyses, because there

Table 3
Reaction Time (in msec) for Test Blocks 1 and 4 of Experiment 1

Test item	Study condition: Control	Practice	Interference	Spurious connection
		Trues		
Block 1				
Fast	842 (.000)	733 (.020)	768 (.005)	734 (.000)
Slow	981 (.040)	790 (.010)	914 (.040)	912 (.030)
Block 4				
Fast	587 (.010)	565 (.030)	612 (.005)	622 (.000)
Slow	609 (.010)	601 (.000)	615 (.010)	625 (.030)
		Falses		
Block 1				
Fast	1,035 (.030)	875 (.000)	916 (.005)	918 (.030)
Slow	1,096 (.120)	929 (.080)	947 (.070)	997 (.150)
Block 4				
Fast	643 (.000)	681 (.000)	676 (.005)	685 (.020)
Slow	635 (.030)	685 (.050)	692 (.015)	690 (.030)

Note. The interference condition means are based on a maximum of 200 observations (200 minus the number of errors), and all other means are based on a maximum of 100 observations. Error rates are in parentheses.

was less than a 3-msec difference overall, and in no block was the difference between the two conditions greater than 1 *SE*.

Analyses were performed separately for true and false pairs. The error terms of all contrasts to be discussed include only the variability of the cells involved in the contrast. Several of the analyses were repeated after taking log transformations of the means. The differences in the outcomes of the untransformed and transformed data analyses were small, and only the analyses of the untransformed data are reported. The randomization of materials for each subject allows the generalization for most contrasts over both subjects and items without the approximate measures advocated by Clark (1973). For the analysis of the test speed effects, however, the use of $F_{\min}$ test was necessary because the test pairs were nested within the test speed levels.

The first issue examined is whether the test speed classification had the expected effect. Over the four blocks, for the true responses, RTs were about 50 msec faster to the test pairs in the fast group, $F_{\min}(1, 25) = 10.39$, $p < .01$. Though this was a smaller difference than found with the same pairs in the Anderson and Reder (1974) study, an inspection of Table 3 shows that the difference was considerably larger in the first block (130 msec) than in the fourth block (16 msec). This decrease in the difference between fast and slow is predicted by ACT in that extra practice should serve to equalize the strength of connection. For false responses, the two test speed levels did not show a significant difference in RT, $F_{\min}(1, 23) < 1$, although the difference is in the expected direction. The large error rates for the slow false items in Block 1 suggest the possibility of a speed–accuracy tradeoff. An analysis of the error rates after an arcsine transformation showed that significantly more errors were made to the slow items, $F_{\min}(1, 23) = 5.59$, $p < .05$. This might have also been partially due to more fast guesses in the slow condition, as suggested in the introduction. Without the expected main effect in RT, the predicted Study Condition × Test Speed interaction becomes impossible to test.

The next result concerns the main effect of blocks. Overall, there was a significant block effect, with subjects speeding up considerably both for true responses, $F(3, 147) = 127.20$, $p < .001$, $MS_e = .033$, and for false responses, $F(3, 147) = 115.78$, $p < .001$, $MS_e = .058$.

The most important results concern the effects of study manipulation. For trues, the conditions ordered themselves: practice (651 msec), spurious connection (689 msec), interference (697 msec), and control (707 msec). The practice condition mean was faster than the mean of the other three conditions, $F(1, 147) = 47.08$, $p < .001$, but the differences among the other three conditions were not significant, $F(2, 147) = 1.76$. Thus, the predicted advantage for the practice condition was obtained, but the predicted deficit for the interference condition relative to the control failed to materialize. The effects of the study conditions did decrease with practice as indicated by a significant Block × Study interaction, $F(9, 441) = 6.16$, $p < .001$, and as can be confirmed in Table 3. Although it was considerably reduced, there was still a significant advantage for the practice material in Block 4, $F(1, 147) = 6.21$, $p < .02$. For true responses, the control condition in Block 1 was slower than both the practice condition, $F(1, 147) = 33.85$, $p < .001$, and the interference conditions, $F(1, 147) = 12.54$, $p < .001$, but by Block 4 there was no significant difference for either contrast, $F(1, 147) < 1$; $F(1, 147) = 1.22$, $p > .25$. This change over blocks was highly significant for both the practice-control contrast, $F(1, 147) = 22.21$, $p < .001$, and the interference-control contrast, $F(1, 147) = 12.20$, $p < .001$.

For the false items, the study conditions ordered themselves: practice (770 msec), interference (789 msec), control (794 msec), and spurious connection (795 msec). The practice condition mean was again faster than the mean of the other three, $F(1, 147) = 6.79$, $p < .01$, but the other conditions did not differ, $F(2, 147) < 1$. There was a highly significant, $F(9, 441) = 8.77$, $p < .001$, interaction between block and study condition as indicated by Table 3. Particularly noteworthy is the change in the relationship between the control and the interference conditions. In Block 1 the control condition was significantly slower, $F(1, 147) = 27.84$, $p < .001$, whereas in Block 4 it was significantly faster, $F(1, 147) = 4.82$, $p < .05$. Thus, the predicted relationship between control and interference was obtained in the final block. It was also predicted that the spurious connection condition would be worse than the interference condition. Even though the reaction time difference was in the right direction, it was clearly not significant. On the other hand, the error rate for spurious connections was .047 compared with .023 for the interference condition, $F(1, 49) = 6.76, p < .05, MS_e = .016$. Thus, a speed–accuracy trade-off may have decreased the RT difference. This accuracy difference might be due to more fast guesses in the spurious connection condition, as mentioned in the introduction.

It is clear that a major unexpected complication is the slowness of the control condition. It may be that subjects are hurt by a general unfamiliarity with the items in this condition. Consistent with this interpretation is the fact that subjects are 370 msec faster in Block 4 than in Block 1 for the control condition but only 224 msec for the other study conditions.

A final question of interest is how this difference in study condition interacted with test speed and how this interaction changed with blocks. Let us consider as our measure of the study effect the difference between the two study conditions predicted to be most extreme: practice and interference for trues and practice and spurious connection for falses. We report a series of contrasts involving these conditions. The expectation was that the study manipulation would be larger with the less strongly connected items, but that this difference would decrease with repetition. For Block 1 trues the study condition effect was larger for the slow pairs, $F(1, 49) = 4.45, p < .05, MS_e = .022$. For Block 4, the difference was small, $F(1, 49) = 1.10, p > .25, MS_e = .013$. There was a definite change in the Study Condition × Test Speed interaction from Block 1 to Block 4, $F(1, 49) = 6.51, p < .05, MS_e = .014$. For false test pairs, since there was no significant RT difference between fast and slow pairs, the predictions are not clear. There was no overall interaction, $F(1, 49) = 1.06$, $p > .25$, $MS_e = .022$. There was no interaction in Block 1 or Block 4, nor was there any change in the interaction with blocks (all Fs < 1).

Discussion

The experiment found a number of instances of transfer between semantic and episodic memory, indicating that semantic and episodic memory are not distinct: (a) Episodic practice facilitated the making of semantic judgments. (b) Although the predicted interference effects often did not materialize, there were two examples of interference for false judgments. In Block 4 the interference condition was slower than the control. Overall, the spurious connection condition produced higher error rates. (c) Practice at decision making reduced prior differences among material and reduced the effects of the study manipulation. Note that this contradicts the results of Shoben et al., which have been used to argue for the semantic–episodic distinction. These effects were predicted by the ACT theory. There was, however, one effect predicted by the ACT theory that was not obtained—that for trues the interference condition should be significantly worse than the control. Inspection of Table 2 will confirm that if anything, the effect is in the opposite direction. Even though this outcome is damning for the ACT theory, it is less clear that it provides much support for the episodic–semantic distinction. ACT is only one species of the theories that make no episodic–semantic distinction. Clearly, no such theory is committed to the claim that every episodic manipulation will have a semantic effect, and certainly many species would not predict an interference effect in this circumstance.

The three positive instances of transfer listed earlier argue against a semantic–episodic distinction. This is not to say that no semantic–episodic theory could predict these results. However, many including the one articulated by Shoben et al. for this task would not.

We were suspicious that the poor performance in the control condition may have been due to the lack of familiarity of the items in this condition. The material in the other conditions had been seen five times during the study phase. This may have facilitated the lexical encoding and access of the words making up these items. Evidence consistent with this familiarization explanation is that after the first test block, the first time the control items were seen, the control condition sped up more rapidly than those in the other conditions.

Experiment 2

The second experiment was an attempt to get further evidence about the study condition manipulations. We wanted to eliminate the potential explanation that the control condition was slow simply because of lack of familiarity with the lexical items. We had the subjects rate the control categories before the verification phase of the experiment to practice encoding these terms. Lewis and Anderson (1976), who had found that subjects learn control material slower than interference material in a similar design, had used a prior rating task for the items that formed the control material.

Another result we wanted to investigate further was why subjects got faster and the differences among conditions disappeared with practice. Practice eliminated both effects due to episodic factors (the study manipulation) and effects due to semantic factors (the test speed manipulation). One change in the design of Experiment 2 was to run the experiment for six blocks to trace further the change in reaction time. A related manipulation was to introduce new test pairs in the third and fifth blocks. This would enable us to determine if the speedup and decrease in size of effects were due to specific practice with the items or to more general experimental practice.

Yet another manipulation occurred in the fifth block. Half of the items and categories used in negative probes were recombined to yield true probes. Similarly, half of the true probes were switched to be false. This manipulation was to get at the issue of whether subjects' speedup depended on something as simple as remembering the response of an item. If it did, subjects should suffer interference and perform worse on these switched probes than on new probes. In contrast, the ACT analysis of the speedup, which attributes it to the strengthening of the relevant semantic information, would not necessarily predict that subjects would perform worse on the switched probes than on new probes.

Because of the extra test variations to study the speedup effect, the variation of fast versus slow test items was eliminated to make it easier to assign materials to conditions. Also, two of the study conditions were deleted—the spurious connection condition for the trues and the practice condition for the falses. These were the two conditions for which the ACT predictions were uncertain. Finally, only one interference manipulation was used, since the two types of interference conditions had not differed in the previous experiment.

Method

Materials and design. The test materials were again taken from Anderson and Reder (1974). The stimuli were 36 item–category pairs. All of the categories and many of the items had been used in Experiment 1. The item–category pairs were grouped into triplets, constructed so as to maximize the relatedness of categories within a triplet, as judged by the authors. The item–category pairs for true and false responses are given in Appendix B. Twelve buffer pairs, grouped into four triplets, were also included.

The study sentences were constructed from the test pairs in the following way, as illustrated in Table 4: For test pairs assigned to either control condition, the items and categories were not included in any sentence shown during the study phase. For test pairs assigned to the true practice condition, the sentences studied included the relevant categorization information in the form "An instance of a *Category* is an *Item*."[3] For the other three study conditions, the manipulations were made over the three test pairs in each triplet. For test pairs in the true interference condition, each item appeared in two sentences, once with each category in the triplet other than its own category. Each category appeared in two sentences, once with each other item in the triplet. This involved having the subject study six interfering sentences for the three pairs in a triplet. All of these sentences used transitive verb constructions. For test pairs

[3] The practice sentences had been worded "An item is a category" in Experiment 1. This format was changed in Experiment 2 to check the uninteresting possibility that the facilitation in the practice condition might be due to a surface match between the test probe and the sentence studied. In Experiment 2, item and category order at test is different from the order at study.

Table 4
Example of Materials for Experiment 2

Study condition	Sentences studied
True test pair: python–snake	
Practice	An instance of a snake is a python.
Control	
Interference	The python ate a fish.
	The dog barked at a python.
	The snake waited for a trout.
	The collie attacked the snake.
False test pair: python–dog	
Control	
Interference	An instance of a dog is a collie.
	The trout swam away from a dog.
	An instance of a snake is a python.
	The python ate a fish.
Spurious connection	The dog barked at a python.
	The python attacked a dog.

Note. For each test pair, subjects studied sentences from one condition. The interference condition study sentences presented also include part of the manipulation for other test pairs in the same triplet.

assigned to the false interference condition, the subject also studied two sentences about the item and two sentences about the category. The item sentences used the other two categories from the triplet, and the category sentences used the other two items. One of the other categories for the item would be the category that contained the item, and similarly one of the items for the category would be in that category. In these cases the subject was presented with a sentence in the same form as in the true practice condition. This allowed a design such that the subset relation sentences gave no clue as to how the item and category would later be tested. The other sentences in the false interference condition involved transitive verbs connecting item and category. The final study condition was the false spurious connection. Item–category test pairs assigned to this condition were presented together in two sentences, one with the item as the sentence subject and one with the category as the sentence subject. Two sets of triplets were assigned to each study manipulation, yielding 42 study sentences (6 for true practice, 12 for true interference, 12 for the false interference, and 12 for the false spurious connection). Fourteen buffer sentences constructed from the buffer test pairs were included. All sentences involving transitive verbs were in the past tense.

The test condition manipulation involved the test block in which the item–category pair was first to be presented (Block 1, 3, or 5) and whether the items and categories recombined to change the truth of the probe. Two triplets were assigned to each study condition. Within each study condition, one of the triplets was assigned to the switched test condition and one to the nonswitched condition. The three test pairs within each triplet were then assigned to a test condition determining the block in which it would be introduced. These assignments were randomly made for each subject. The number of test condition manipulations varied with block: On Blocks 1 and 2 there was one, on Blocks 3 and 4 there were two (whether the items were introduced on Blocks 1 or 3), and on Blocks 5 and 6 there were four (whether the items had been introduced in Blocks 1, 3, or 5 or were switched.) In the switched case one or both terms had been introduced in a prior block.

For the pleasantness ratings, all words (96) to be used in any categorization judgment were included. Two sets of 48 words were constructed with an item and its true category assigned to different sets. Half of the subjects were shown each word set first.

Subjects. The subjects were 42 adults, 18–30 yr old, from the Yale University community. Most were undergraduates. In addition, data from 9 other subjects were not included, due to computer problems (6), low study accuracy (1), or low test accuracy (2). Each subject was given course credit or was paid $5 for participating.

Procedure. The computer setup was as in Experiment 1.

The procedure had a few changes from the first experiment. In the completion phase, there were only three passes through the 56 study sentences, and subjects had to type only the last word of the sentence, which was always an item or category to be used in the categorization judgments. These changes were introduced to compensate for time increases created by other aspects of the design. Following the study phase, subjects rated the 96 words composing the test pairs as to their pleasantness, on a scale of 1 to 9. Though the subjects were not told this beforehand, they were then shown the same words again and were asked to give each word the same rating as they had the first time. These 192 rating responses took about 15 min. In the test phase there were two differences from Experiment 1. Since new items were introduced in Blocks 3 and 5, the test block length changed. For Blocks 1–6, the numbers of trials, including warm-up items, were 18, 18, 33, 33, 48, and 48, respectively. Subjects had been told when new items would be presented and were reminded immediately before Blocks 3 and 5. Before Block 5, subjects were warned that half of the item–category pairings would be switched.

Results

Subjects were 98% correct in the final pass through the completion phase. For each subject, RTs for correct responses in each condition in the verification phase (Study Manipulation × Test Condition × Block × Truth) were averaged. To avoid effects of extremes, times longer than 2.5 sec (less than 1% of data) were set equal to 2.5 sec. Figures 1 and 2 present a large portion of the data. Blocks 1 and 2, Blocks 3 and 4, and Blocks 5 and 6 were collapsed, and the data are not presented from the switched-truth conditions. Each curve in these figures presents a different test condition, with the curves beginning at the block the test conditions were introduced. The standard errors of the times in these figures were 20.1 msec for tubes in Blocks 1 and 2, 16.0 msec for trues in Blocks 3 and 4, 23.3 msec for trues in Blocks 5 and 6, 26.3 msec for falses in Blocks 1 and 2, 21.2 msec for falses in Blocks 3 and 4, and 31.3 msec for falses in Blocks 5 and 6. Standard errors were calculated from the Subject × Condition interaction for each block. These standard errors from each block were pooled to get the standard errors reported for pairs of blocks. These overall standard errors were used in forming the statistical tests. It is important to realize in interpreting these figures that times become more variable in Blocks 5 and 6 because they are only half as many observations per condition. The other half of the pairings were switched.

To simplify the data presentation in these figures, we have omitted error rates. In general, errors were low (.022 for Blocks 1 and 2 true, .019 for Blocks 3 and 4 true, .019 for Blocks 5 and 6 true, .067 for Blocks 1 and 2 false, .032 for Blocks 3 and 4 false, .024 for Blocks 5 and 6 false) and somewhat unrelated to reaction times for trues (r = .26), and positively correlated with reaction times for falses (r = .74). In no case is there a claim based on reaction time that would be compromised by consideration of error rates.

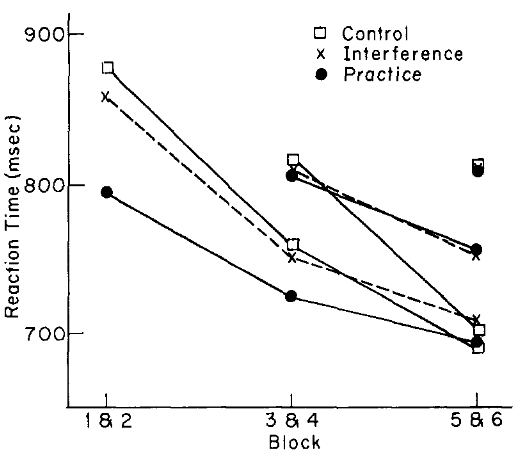

Figure 1. Mean reaction times for true judgments in Experiment 2. (Separate functions are plotted for items introduced first in Block 1, for items introduced first in Block 3, and for items introduced first in Block 5.)

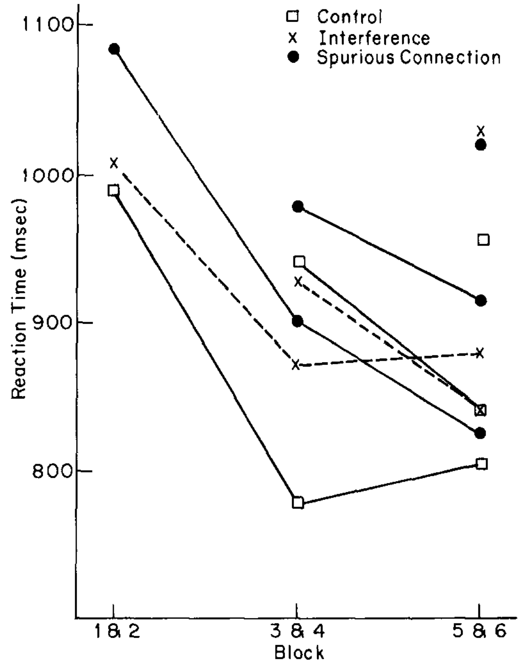

Figure 2. Mean reaction times for false judgments in Experiment 2. (Separate functions are plotted for items introduced first in Block 1, for items introduced first in Block 3, and for items introduced first in Block 5.)

Let us consider the data in the true condition (Figure 1) first. Collapsing the data in Figure 1 over everything but study condition resulted in a 767-msec mean in the practice condition, 781 msec in the control condition, and 787 msec in the interference condition. With an 8.4-msec *SE* for these pooled data, there was a marginally significant difference between the practice condition and the mean of the interference and control conditions. Thus, it does appear that the practice condition was facilitating relative to the control, but there is no significant evidence that the interference conditions hurt. As Figure 1 illustrates, there was a significant interaction, $F(4, 1230) = 2.38$, $p < .05$, between the three pairs of blocks and the three study manipulations. For the first pair of blocks, the difference between the mean of the control and interference

(869 msec.) was significantly, $t(164) = 3.01$, $p < .01$, slower than the practice (795 msec). In contrast, there was virtually no difference between the mean of the control and interference conditions (766 msec) and the practice condition (761 msec) on later trials. Thus, the overall effect of study, which was significant, reflected what happened on the first trials. It should be noted that Experiment 1 found the study manipulation still significant for trues on later blocks, although that study manipulation was reduced.

There appears to be some tendency for the control condition to speed up relative to the interference condition. Looking at the first two blocks when material was first introduced, the control material had a mean 845 msec and the interference material had a mean 832 msec. Looking at the third, fourth, fifth, and sixth blocks, after the material has been introduced, the control had a mean 716 msec and the mean of the interference condition was 739 msec. This interaction was marginally significant, $t(1230) = 1.52$, $p < .15$. To the extent that this interaction is real, it suggests that we did not entirely succeed in our efforts to familiarize subjects with the material in the control condition.

Collapsing the false data in Figure 2 over everything but study condition, the means were 884 msec in the control condition, 926 msec in the interference condition, and 954 msec in the spurious connection condition. With an *SE* of 11.0 msec, the differences among all three conditions were significant. The *t*s for differences among the adjacent pairs were 2.70 and 1.80 (one-tailed tests were used for these planned comparisons). ACT's predictions of the interference condition slower than the control and the spurious connection condition slowest were confirmed.

This pattern of the data held up for every block. In Blocks 1 and 2, control was 980 msec, interference was 1,008 msec, and spurious connection was 1.083 msec (*t*s of .75 and 2.02). In Blocks 3 and 4 control was 859 msec, interference was 900 msec, and spurious connection was 939 msec (*t*s of 1.93 and 1.84). In Blocks 5 and 6 control was 868 msec, interference was 917 msec, and spurious connection was 920 msec (*t*s of 1.92 and .12). The individual differences were not always significant, but they were always in the right direction. A test of whether there was an interaction between blocks and study was not significant, $F(4, 1230) < 1$. This contrasts with the true data, which indicated that the study manipulation had no effect after the first two blocks.

It appears that there was no significant interaction between specific practice and study condition in the false data. The means for the first two tests on any test probe, averaged over block, were control, 959 msec; interference, 989 msec; and spurious connection, 1,027 msec. The means for the third, fourth, fifth, and sixth blocks were 808 msec, 864 msec, and 880 msec. The interaction between blocks and study condition, as reflected in these numbers, was not significant, $F(2, 1230) < 1$. The false data contrast with Experiment 1, which showed a significant change in the false study effects with practice. However, this trend in Experiment 1 depended on contrasts involving the control or practice condition. In this experiment efforts were taken to avoid the initial lack of familiarity in the control condition, and there was no practice condition for the false materials. Although clearly not significant, the control material did speed up slightly more (151 msec) than the interference (125 msec) or the spurious connection (147 msec). This is consistent with the weak indication in the true data that the control material was not quite equal in familiarity.

Figures 1 and 2 clearly indicate that the speedup in reaction time is at least partially due to specific practice and not general speedup. Table 5 presents the analysis for the data on Blocks 5 and 6, dividing the data according to whether the test item was switched, and if unswitched whether it was introduced on Blocks 1, 3, or 5. Subjects were 37 msec faster on material that they had been practicing since the beginning of the experiment than on items introduced in Block 3, $t(656) = 2.32$, $p < .05$, and 110 msec faster on these items than on items introduced in Block 5, $t(656) = 6.92$, $p < .001$). The switched data were not classified according to the block of introduction. Because these data came from items and cate-

Table 5
Mean Reaction Times (in msec) in Experiment 2 for Blocks 5 and 6

	Introduced on Trial			Switched on Trial 5
	1	3	5	
True	696 (.008)	740 (.000)	822 (.048)	883 (.039)
False	837 (.012)	866 (.016)	1,003 (.044)	966 (.071)
M	766 (.010)	803 (.008)	913 (.045)	925 (.050)

Note. Error rates are in parentheses.

gories that were introduced on different blocks, such a classification would not be meaningful. The critical question is how these switched items fare relative to items introduced in Block 5. The new items introduced as trues on Block 5 were dealt with faster than the true items switched from falses on Block 5, $t(1312) = 3.67$, $p < .001$. In contrast, the new falses on Block 5 were dealt with slower than the falses switched from trues on Block 5, $t(1312) = 1.83$, $p < .05$. The explanation of this interaction is that the switched trues came from originally false items in which there were two interfering study conditions, whereas the switched falses came from originally true items in which there was a practice condition. Thus, comparing new trues with switched trues also compares the effects of practice with the effects of interference. The opposite is the case in comparing new falses with switched falses. The only way to compare the switched items with new items is to look at an average over trues and falses. These averages represent an equal frequency of interference versus practice study manipulations for both the new items and the switched items. As can be seen, average performance was similar in the two conditions. The lack of an effect of switching contradicts the hypothesis that subjects' speedup was due to simple storage of a response. If so, the switched truth condition should have resulted in response interference and poorer performance. Thus the subjects' speedup in the nonswitched condition appears to depend on some speedup in the actual judgment process, consistent with ACT's expectation that practice would facilitate retrieval of the relevant information.

Discussion

Like Experiment 1, Experiment 2 found many examples of transfer from episodic to semantic memory: (a) There was an effect of episodic practice on a semantic judgment for the trues (a replication of Experiment 1). (b) There was a significant interference effect for falses. There was some indication of this in Block 4 of Experiment 1. (c) There was a significant deficit due to spurious intersections (a replication). (d) Practice at making semantic judgments produced a speedup (replication), which was not due to general facilitation of the task or to simple retrieval of a response. (e) The effects of the study manipulation decreased with practice for the trues (a replication). These results are all consistent with the ACT theory. However, again, there are predictions from the ACT theory that failed to be confirmed. The serious failure of ACT's prediction was the lack of a difference between the control and interference conditions for the trues.

Experiment 3

Experiment 1 found no significant difference between control and interference for the trues and only a weak indication of interference on later blocks for falses. In Experiment 2 we had tried to equate for lexical familiarity for a prior rating task. Here we found no significant difference for the trues and a significant interfering effect for the falses. It might be argued that Experiment 2 had not completely equated for lexical familiarity. Subjects rated all words, control and experimental, but they had extra study exposures for the experimental words. The weak tendency for control to speed up relative to the experimental conditions is

Table 6
Reaction Times for Experiment 3

Condition	True		False	
	Fast	Slow	Fast	Slow
Control	850 (.041)	914 (.026)	934 (.037)	969 (.059)
Interference	846 (.018)	878 (.020)	964 (.057)	995 (.082)

Note. Error rates are in parentheses.

consistent with this interpretation. The third experiment tried to better equate prior exposure in the control and interference conditions. To increase the experimental power with which we could examine this issue, control and interference were the only study conditions used. We also increased the number of blocks to 10 to augment experimental power and to get a better estimate of the effect of practice. Because the design constraints permitted it, we also reintroduced the distinction from Experiment 1 of fast versus slow items.

Method

Materials and design. The same 40 test pairs were used as in Experiment 1 (see Appendix A). For each subject and each predicate of these pairs, one sentence was created with that term as subject and another sentence was created with that term involved in the predicate. These study sentences were created so that they used two words from the test set—one in subject position and one in predicate. So, in all, 80 study sentences were created. In addition there were 8 practice test pairs from which 16 study sentences were created. The items were randomly divided in half to form control and interference material for each subject under the constraint that there were an equal number of slow and fast pairs in the interference and control conditions. So there were 10 pairs in each of the following four conditions: control–fast, interference–fast, control–slow, and interference–slow. Along with the practice items, a subject had to study 48 sentences.

Procedure. The computer setup and general procedure were basically the same as in the previous two experiments. The study phase involved study of the material and four attempts at completion. Since each word in the interference pairs appeared in two sentences in each pass through the study material, we had the subject rate each item from the control phase twice in each pass. One rating asked the subject to assign a pleasantness value to the word, and the other rating involved word frequency. The scale for both ratings was 1 to 7. The test phase involved 10 passes through the 40 pairs. In this experiment the test pairs were presented in the simple form "item category." Each test pass was preceded by the 8 practice pairs. There was a short rest after each study pass.

Subjects. The subjects were 39 adults from the Carnegie-Mellon University community. Most were undergraduates. In addition, data from 10 other subjects was not included: 2 for computer failures; 7 because of failing to achieve the learning criterion of 90% correct; and 1 because his reaction times averaged greater than 2 sec. Each subject participated to earn credit for an introductory psychology course or for \$6. The experiment lasted less than 2 hr.

Results

Subjects were 97% correct in the final pass through the completion phase. To avoid effects of extreme times, RTs greater than 2.5 sec (less than 2% of the data) were set to 2.5 sec. Errors (4.3% of the data) were excluded in calculated mean RTs. The 10 blocks were collapsed into groups as follows: Group 1 was Block 1; Group 2 was the average of Blocks 2 and 3; Group 3 was the average of Blocks 4–6; and group 4 was the average of Blocks 7–10. This was done to give greater emphasis to first trials and because reaction times changed more slowly during later trials. The mean reaction times for the four groups were 1,152 msec, 940 msec, 824 msec, and 757 msec.

An analysis of variance was performed using the conditions defined by the factorial combination of groups of blocks, speed, truth, and study manipulations. Separate analyses were performed on reaction times and accuracy using the subjects' mean RT or mean percentage of error in each cell defined by the factorial design. There were highly significant effects of all main variables except study condition. Table 6 presents a breakdown of the data according to speed, truth, and study manipulation. The standard error of the reaction times was

9.1 msec, and for the percentage of error it was .0056.

As is clear from Table 6, there was a highly significant interaction ($p < .001$) between study condition and truth for both RTs and error rates. Subjects performed better in the interference condition for the trues but worse for the falses. For reaction times, the difference for trues was significant $t(38) = 2.20$; $p < .05$, as was the opposite difference for falses, $t(38) = 3.02$, $p < .005$. Similarly, for percentage of error, the difference for trues was significant, $t(38) = 2.59$, $p < .01$, as was the opposite difference for falses, $t(38) = 3.84, p < .001$.

In contrast to Experiment 1, this experiment did obtain a significant effect of the speed variable on reaction times for both trues, $t(38) = 5.44$, $p < .001$, and falses, $t(38) = 3.08$, $p < .005$. As for accuracy, there was no significant effect of speed for trues, $t(38) = 1.16$, but there was for falses $t(38) = 4.20$, $p < .001$, and there was a significant interaction between trues and falses with respect to the effects of speed on accuracy, $F(1, 38) = 13.27, p < .001$. As predicted by ACT, the effect of this speed variable decreases with practice. (This is in contradiction to the conclusions of Shoben et al., 1978). The Block × Speed interactions were highly significant for reaction time, $F(3, 114) = 17.42$, and significant for percentage correct, $F(3, 114) = 3.67$. The mean difference between fast and slow in Block 1 was 121 msec, whereas for Blocks 7–10 it was 10 msec. Unlike previous experiments there was no significant interaction between practice and the study manipulations for reaction time, $F(3, 114) = .54$, for the Block × Study interaction, and, $F(3, 114) = .58$, for the Block × Study × Truth interaction. However, there was significant Block × Study × Truth interaction for accuracy, $F(3, 114) = 5.32$, $p < .005$, such that the initial accuracy differences disappeared with practice.

Discussion

Like the past experiments, this experiment has produced numerous demonstrations of transfer of episodic experiences to semantic memory retrieval: (a) The interference study manipulation proved to have a facilitating effect on the judgment of true semantic statements. Previous experiments had failed to get significant results. (b) The interference study manipulation proved to have an interfering effect on the judgment of false semantic memory statements (a replication). (c) Practice in the experiment served to reduce differences between fast and slow semantic statements (a replication). Indeed, in this experiment unlike the prior two, there was no major instance of an episodic manipulation that failed to have some impact on semantic memory judgments. The ACT theory predicted all of these effects except the first. ACT predicts that the interference manipulation should interfere. This prediction failed in the previous two experiments, but there was reason to attribute the failure to lack of recent practice at lexical encoding. It is clear that this explanation cannot apply in the current experiment. Moreover, rather than no effect of the interference manipulation, it proved to be facilitating—both in terms of reaction time and error rates.

General Discussion

Before commenting on the general implications of these experiments, it is worth commenting on the small size of the effects on which these conclusions are based. These conclusions are based on reaction time differences between conditions of as little as 20 msec and error rate differences as small as 1.5%. Particularly with respect to reaction times where the means are as large as 1 sec, these seem like very small differences. However, in some cases these comparisons are based on as many as 4,000 observations. The fact that these episodic manipulations are weak should not be surprising because we were trying to manipulate strongly encoded semantic memories. Thus, the only reason why this research was successful was because it was designed to be able to clearly detect such small differences.

It is also worth reviewing what we think are the empirical conclusions from these experiments: (a) Committing to episodic memory a true semantic fact facilitates its later semantic retrieval. (b) Committing to episodic memory irrelevant information about concepts does not interfere with and can facilitate later recognition of semantic facts about these concepts. (c) Committing to episodic memory irrelevant information about concepts interferes with later rejection of false semantic facts about these concepts. (d) Committing to episodic memory facts that introduce spurious connections between two concepts makes it harder (than in c) to reject a false semantic fact connecting these concepts. (e) Practice at making semantic judgments leads to a speedup that is not due to general task facilitation or to simple retrieval of a stored response. It is true that all of these results did not appear significant in all experiments, but there were certain methodological difficulties with the first experiment, and it is to be expected statistically that certain real differences will prove insignificant on occasion. However, we feel that the weight of the three experiments strongly supports these conclusions.

It is clear that these results offer little support for the semantic–episodic distinction. The experiments provided abundant examples of transfer from episodic memory to semantic memory. Many of these transfer effects are in accord with principles established for episodic memory (practice effects, interference effects for falses) and for semantic memory (effects of spurious connections on falses). To be sure, some semantic–episodic model could be concocted to account for these results, but the character of that model is not immediately obvious. Certainly, the model of Shoben et al., the most directly applicable to this situation, is contradicted by many of these results. In line with arguments made by Anderson (1976, 1978), it is not possible to produce empirical data that will reject all possible semantic–episodic models. In absence of this possibility, we have to content ourselves with the observation that these results are inconsistent with the spirit and semantic–episodic models (i.e., that there should not be transfer between the two memories and that these memories should obey different laws) and are inconsistent with the known semantic–episodic models.

On the other hand, although these results are clearly in keeping with the spirit of models that make no semantic–episodic distinctions, there is one result that is in serious contradiction to the predictions of the ACT theory that we had in mind. This is the failure to find an interfering effect for trues. It should be noted that this result is contrary to the findings of Lewis and Anderson (1976), who found that episodic information interfered with retrieval of biographical facts. There are, of course, many possible differences between these experiments and those of Lewis and Anderson, but we prefer to attribute the different results to the difference between semantic and biographical memories. There is no reason why a subject would rehearse a biographical fact about an individual while studying a fantasy fact about that individual. For instance, there is no reason to rehearse that Washington crossed the Delaware while studying that Washington was a liberal senator. On the other hand, it seems reasonable that a subject while studying a sentence about a cobra might rehearse the fact that a cobra is a snake. Consistent with this is some unpublished data from our laboratory showing that learning additional, supposedly interfering, information about concepts facilitates making semantic consistency judgments involving the concepts.

It certainly would not be hard to propose a version of a nonsemantic–episodic model that would predict that the implicit rehearsal of semantic information would be sufficiently facilitating to overcome any interfering effect of the study manipulation. In fact, ACT is such a model under certain assumptions about the beneficial effect of the episodic rehearsal relative to the detrimental effect of the interference. For instance, if we assume that the strengthening of semantic information due to rehearsal while studying interfering information is equal to the strengthening of the interfering information, then the following analysis applies: Let a be the strength of the old semantic fact, S be the strength of the other facts about the concepts, and s be the amount by which the old semantic and new episodic information is strengthened. Reaction time to the old semantic fact will be a function of the ratio of the strength of all facts to its strength. Thus verification time will vary with $(S + a + 2s)/(a + s) = 2 + (S - a)/(a + s)$ which is a negative function of s, the amount of episodic study. In this view the better control to assess an interfering effect is the practice condition that involved semantic practice and no interference. There was a significant deficit of the interference condition relative to the practice condition.

Even though the beneficial effect of interfering material on trues can be explained in the ACT framework, it clearly was not predicted. Whether the ACT post hoc explanation is correct or not can only be determined by further research. If the ACT post hoc explanation is correct, the failure of prediction points to a weakness in the theory's ability to analyze the semantic processing of a sentence as it is encoded.

Other Issues

Although this experiment was not specifically designed to compare semantic memory models, it does have implications for a current controversy. To account for the data from various semantic memory experiments, network models (Collins & Loftus, 1975; Glass & Holyoak, 1975) and set-theoretic models (Meyer, 1970; Smith et al., 1974) have been proposed. One may ask how these models could account for the results of the present experiment. Since ACT is the spreading activation network theory, network models may obviously explain the results. As another example, the Collins and Loftus model would need an additional structural assumption, that episodic information is stored with semantic information, but no new processing assumptions.

The necessary modifications for the set-theoretic models are less obvious. Consider the most complete model of this type, the feature comparison model (Rips, Shoben, & Smith, 1973; Smith et al., 1974). In this model, decisions are made by comparing the feature overlap between the test item and category. To account for the present results, study manipulations would have to affect at least one of the following: accessibility of features, decision thresholds, or the relative importance of features. Further, one of the strongest points of the feature comparison model is the ease with which it accounts for relatedness effects on negative judgments. The apparent effect of spurious intersections argues against the feature explanation. Although an extension could presumably be worked out, since these network and feature models may well be isomorphic (Hollan, 1975), it will probably require more involved modifications than the network model.

We would like to conclude with some general comments about the nature of the semantic–episodic distinction and its motivation. Memory distinctions can be of two general kinds, functional or content. Though this classification is not absolute and may better be regarded as a continuum, the purposes and consequences of these two kinds of distinctions are different. A functional distinction implies that the memories being distinguished have different structural properties or, perhaps equivalently, that they show differential properties in basic memory phenomena and operations such as decay, interference, or retrieval. The short-term memory versus long-term memory distinction is believed by many to be an example of this kind of distinction. Short-term memory is generally acknowledged to show faster decay, be more vulnerable to certain types of interference, and involve different retrieval processes.

Content distinctions, however, formalize differences in the types of information being processed. Though these may lead to the discovery of functional distinctions, this is not their sole purpose. The main reason for proposing content distinctions is to gain new insights. This goal may be accomplished in two ways. Simply by conceptually partitioning the information, the classification may lead to new ways of looking at the issues. A detailed consideration of a particular subdomain may suggest possibilities that would not have occurred otherwise. In this case, then, one uses the distinction in the hope of generating new ideas about the problem.

One may also use content distinctions to try to discover general principles or mechanisms. For example, the observed differences between words and unrelated letter strings on tasks such as memory span were instrumental in developing the widely used notion of "chunking" (Miller, 1956).[4] When manipulations with different kinds of information lead to different results, there are two possible reasons. One possibility is that the memory representations of the two information types (or the operations acting on them) are qualitatively different. A second possibility, however, is that the results are just two manifestations of the same general mechanism acting on slightly different kinds of information. If the latter is the case, a comparison of the two results may be an important clue to understanding the underlying mechanism.

The point of outlining this classification of distinctions is to propose that the semantic–episodic distinction has been generally accepted as a functional distinction, when it should have been considered as a content distinction. The reasons for this acceptance with little empirical support are unclear. Perhaps the two areas of research initially involved such different paradigms, measures, and issues, that a functional distinction seemed only reasonable. This perceived functional distinction led to people "working in" semantic or episodic memory. In turn, this situation has resulted in generally separate literature with little thought given to the similarities and relations. We hope that one outcome of this research will be that researchers will consider more carefully the connection between the two memories. If there are no functional differences, all of the principles from one research tradition should transfer to the other. This would mean both a considerable simplification in our understanding of human memory and an increase in what we know about memory.

[4] We thank Paul Kline for suggesting this example.

Reference Notes

1. McCloskey, M. *Search and comparison processes in fact retrieval and question-answering.* Unpublished manuscript, Johns Hopkins University, 1978.
2. Glucksberg, S., & McCloskey, M. *Advantages of total ignorance: It's harder to remember that you don't know than to discover that you don't know.* Paper presented at the meeting of the Psychonomic Society, Phoenix, Arizona, November 8–10, 1979.

References

Anderson, J. R. Retrieval of propositional information from long-term memory. *Cognitive Psychology*, 1974, *5*, 451–474.

Anderson, J. R. *Language, memory, and thought.* Hillsdale, N.J.: Erlbaum, 1976.

Anderson, J. R. Arguments concerning representations for mental imagery. *Psychological Review*, 1978, *85*, 249–277.

Anderson, J. R., & Bower, G. H. *Human associative memory.* Washington, D.C.: Hemisphere Press, 1973.

Anderson, J. R., & Reder, L. M. Negative judgments in and about semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1974, *13*, 664–681.

Atkinson, R. C., Herrmann, D. J., & Wescourt, K. T. Search processes in recognition memory. In R. L. Solso (Ed.), *Theories in cognitive psychology.* Hillsdale, N.J.: Erlbaum, 1974.

Clark, H. H. The language-as-fixed effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 335–359.

Collins, A. M., & Loftus, E. F. A spreading-activation theory of semantic processing. *Psychological Review*, 1975, *82*, 407–428.

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, *8*, 240–247.

Collins, A. M., & Quillian, M. R. Experiments on semantic memory and language comprehension. In L. W. Gregg (Ed.), *Cognition in learning and memory.* New York: Wiley, 1972. (a)

Collins, A. M., & Quillian, M. R. How to make a language user. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press, 1972. (b)

Fischler, I. Semantic facilitation without association in a lexical decision task. *Memory and Cognition*, 1977, *5*, 335–339.

Flexser, A. J., & Tulving, E. Retrieval independence and recall. *Psychological Review*, 1978, *85*, 153–171.

Freedman, J. L., & Loftus, E. F. Retrieval of words from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 107–115.

Glass, A. L., & Holyoak, K. J. Alternative conceptions of semantic theory. *Cognition*, 1975, *3*, 313–339.

Glass, A. L., Holyoak, K. J., & O'Dell, C. Production frequency and the verification of quantified statements. *Journal of Verbal Learning and Verbal Behavior*, 1974, *13*, 237–254.

Hall, J. F. *The psychology of learning.* Philadelphia, Pa.: Lippincott, 1966.

Hayes-Roth, B. Evolution of cognitive structures and processes. *Psychological Review*, 1977, *84*, 260–278.

Hollan, J. D. Features and semantic memory: Set-theoretic or network model? *Psychological Review*, 1975, *82*, 154–155.

Holyoak, K. J., & Glass, A. L. The role of contradictions and counterexamples in the rejection of false sentences. *Journal of Verbal Learning and Verbal Behavior*, 1975, *14*, 215–239.

King, D. R. W., & Anderson, J. R. Long-term memory search: An intersecting activation process. *Journal of Verbal Learning and Verbal Behavior*, 1976, *15*, 587–605.

Kintsch, W. *The Representation of meaning in memory*. Hillsdale, N.J.: Erlbaum, 1974.

Kintsch, W. Memory representations of text. In R. L. Solso (Ed.), *Information processing and cognition*. Hillsdale, N.J.: Erlbaum, 1975.

Kjeldergaard, P. M. Transfer and mediation in verbal learning. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N.J.: Prentice-Hall, 1968.

Landauer, T. K., & Meyer, D. E. Category size and semantic-memory retrieval. *Journal of Verbal Learning and Verbal Behavior*, 1972, *11*, 539–549.

Lewis, C. L., & Anderson, J. R. Interference with real world knowledge. *Cognitive Psychology*, 1976, *8*, 311–335.

Lockhart, R. S., Craik, F. I. M., & Jacoby, L. Depth of processing, recognition, and recall. In J. Brown (Ed.), *Recall and recognition*. London: Wiley, 1976.

Loftus, E. F. Activation of semantic memory. *American Journal of Psychology*, 1974, *86*, 331–337.

McKoon, G., & Ratcliff, R. Priming in episodic and semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1979, *18*, 463–480.

Meyer, D. E. On the representation and retrieval of stored semantic information. *Cognitive Psychology*, 1970, *1*, 242–300.

Meyer, D. E., & Schvaneveldt, R. Meaning, memory, and mental processes. *Science* 1976, *192*, 27–33.

Miller, G. A. The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, *63*, 81–97.

Muter, P. Recognition failure of recallable words in semantic memory. *Memory and Cognition*, 1978, *6*, 9–12.

Neely, J. H. Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, 1977, *106*, 226–254.

Ortony, A. How episodic is semantic memory? In R. C. Schank & B. L. Nash-Webber (Eds.), *Theoretical issues in natural language processing*. Cambridge, Mass.: Bolt Beranek & Newman, 1975.

Perlmutter, J., Harsip, J., & Myers, J. L. The role of semantic knowledge in retrieval from episodic long-term memories: Implications for a model of retrieval. *Memory and Cognition*, 1976, *4*, 361–368.

Postman, L. Tests of the generality of the principle of encoding specificity. *Memory and Cognition*, 1975, *3*, 663–672.

Proudfoot, R. G. YEPS: Running real-time experiments on a timesharing system. *Behavior Research Methods and Instrumentation*, 1978, *10*, 291–296.

Quillian, M. R. Semantic memory. In M. Minsky (Ed.), *Semantic information processing*. Cambridge, Mass.: MIT Press, 1968.

Rips, L. J., Shoben, E. J., and Smith, E. E. Semantic distance and the verification of semantic relations, *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 1–20.

Schaeffer, B. & Wallace, R. The comparison of word meanings. *Journal of Experimental Psychology*, 1970, *86*, 144–152.

Schank, R. C. The structure of episodes in memory. In D. G. Bobrow & A. M. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.

Shoben, E. J., Wescourt, K. T., & Smith, E. E. Sentence verification, sentence recognition, and the semantic-episodic distinction. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, *4*, 304–317.

Slamecka, N. J. Differentiation versus unlearning of verbal associations. *Journal of Experimental Psychology*, 1966, *71*, 822–828.

Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 1974, *81*, 214–241.

Thorndyke, P. W., & Bower, G. H. Storage and retrieval processes in sentence memory. *Cognitive Psychology*, 1974, *5*, 515–543.

Townsend, J. T. Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. Hillsdale, N.J.: Erlbaum, 1974.

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, 1976.

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

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

Wilkins, A. Conjoint frequency, category size, and categorization time. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 382–385.

Appendix A

Table A1
Item–Category Pairs Used in Experiments 1 and 3

Fast	Slow	Fast	Slow
Trues		Falses	
Cobra–Snake	Shirt–Clothing	Cat–State	Idaho–City
Spaniel–Dog	Fireman–Profession	Cancer–Fuel	Pepper–Beverage
Jazz–Music	Cotton–Cloth	Robin–Vegetable	Sheet–Animal
Harvard–College	France–Country	Termite–Bird	Oak–Flower
Chair–Furniture	Cake–Pastry	London–Artwork	Petroleum–Ship
Polka–Dance	Shilling–Money	Diamond–Tree	Etching–Dwelling
Football–Sport	Arson–Crime	Hammer–Jewel	Rose–Insect
Gun–Weapon	Grape–Fruit	Trout–Disease	Eggnog–Fish
Car–Vehicle	Geology–Science	Cruiser–Tool	Zinc–Spice
Aunt–Relative	Doll–Toy	Carrot–Bedding	Tent–Metal

Appendix B

Table B1
Material Used in Experiment 2

	Category			Category	
Item	True	False	Item	True	False
Reno	City	Country	Zinc	Metal	Spice
Idaho	State	City	Diamond	Jewel	Metal
France	Country	State	Pepper	Spice	Jewel
Python	Snake	Dog	Car	Vehicle	Money
Trout	Fish	Snake	Propane	Fuel	Vehicle
Collie	Dog	Fish	Dollar	Money	Fuel
Apple	Fruit	Vegetable	Concerto	Music	Dance
Carrot	Vegetable	Pastry	Waltz	Dance	Artwork
Cake	Pastry	Fruit	Painting	Artwork	Music
Arson	Crime	Weapon	Lawyer	Profession	Science
Chisel	Tool	Crime	Harvard	College	Profession
Gun	Weapon	Tool	Geology	Science	College
Termite	Insect	Cloth	Polo	Sport	Toy
Pelican	Bird	Insect	Cruiser	Ship	Sport
Cotton	Cloth	Bird	Doll	Toy	Ship
Rose	Flower	Animal	Shirt	Clothing	Furniture
Oak	Tree	Flower	Tent	Dwelling	Clothing
Cat	Animal	Tree	Sofa	Furniture	Dwelling

Received March 3, 1980 ■