

Semantic specificity in cued recall

HENRY L. ROEDIGER III

Purdue University, West Lafayette, Indiana 47907

and

BETH ADELSON

Harvard University, Cambridge, Massachusetts 02138

In three cued recall experiments, extralist retrieval cues that were congruous in meaning to an encoded target pair of words produced better recall of the targets than when the cue-target relation was incongruous. However, this result, which differs from that of other experiments, depended in some cases on scoring recall of a target pair when either member of the pair was recalled. It is argued that (1) pairs of words are typically stored as higher order units, (2) the best test procedure is to request recall of both members of the pair when an extralist cue is presented, and (3) semantic features provide an important dimension in the mnemonic representation of word events.

One general principle of learning and memory that has repeatedly been "discovered" and relabeled over the years is that the more similar the environments in which an organism learns and is tested on a response, the better performance will be. In the conditioning literature, this is referred to as stimulus generalization. In both classical and instrumental conditioning procedures, it has been shown repeatedly that organisms trained to respond in the presence of a particular stimulus will tend to respond more strongly to other stimuli the more closely they match the original conditioned stimulus on various physical dimensions. For example, Guttman and Kalish (1956) trained pigeons to peck a key upon which was projected light of a certain wavelength. During extinction they tested subjects with 11 wavelengths spread about the original conditioned stimulus and found quite regular generalization gradients: The closer the wave-length of the test stimulus to the original conditioned stimulus, the greater the responding during extinction.

Students of human learning and memory have often proposed a similar principle that governs memory for past events. For example, Hollingworth (1928) promulgated a "principle of reinstatement of stimulating conditions," whereby the successful recall of an experience was thought to depend on, among other things, the completeness of reinstatement at recall of the context in which the event originally occurred. McGeoch (1942) also proposed that one factor determining forgetting

Experiments 2 and 3 were conducted at the University of Toronto, where the first author was a Visiting Associate Professor and the second author was an undergraduate student. We gratefully acknowledge the assistance of M. Ellen McIntosh in conducting Experiment 1. Requests for reprints may be sent to Henry L. Roediger III, Department of Psychology, Purdue University, West Lafayette, Indiana 47907.

was altered stimulating conditions. "Everything learned is learned in response to stimulating or antecedent conditions which are part of the learning situation and specific to it . . . alteration produces a decrement in retention, if as a result of it (a) the stimuli necessary to elicit the originally learned acts are not effectively present or (b) new stimuli are introduced which evoke competing responses in sufficient strength to block the originally learned ones" (McGeoch, 1942, p. 501).

One of the newest variants of this old idea is the encoding specificity hypothesis (or principle), originally proposed by Tulving and Osler (1968). This principle states that "specific encoding operations performed on what is perceived determine what is stored, and what is stored determines what retrieval cues are effective in providing access to what is stored" (Tulving & Thomson, 1973, p. 368). Stated slightly differently, it is "the principle that any retrieval cue is effective only to the extent that its informational contents match the information contained in the trace of the event" (Tulving & Watkins, 1975, p. 262). Encoding specificity thus has much in common with previous ideas, such as stimulus generalization, but there is a shift of focus from some previous conceptions in that one cannot infer the properties of the memory trace of an event from the overt, observable properties of the stimulus. Of course, there is a long history in the study of both human learning (e.g., Underwood, 1963) and animal learning (e.g., Spence, 1936) concerning the problem of distinguishing the nominal stimulus presented to the subject from the functional stimulus that governs behavior.

The evidence that is most easily interpreted in terms of the encoding specificity principle comes from experiments in which the same nominal stimulus is studied under conditions designed to effect two different encod-

ings and is then tested under various retrieval cuing conditions. For example, subjects presented with single target words such as CHAIR recall the targets better in the presence of strong associates such as Table than when no cues are provided (free recall). However, if CHAIR is presented in the context of a weak associate such as Glue, then Table no longer serves as such an effective cue and in fact may produce poorer recall than found in free recall conditions (Thomson & Tulving, 1970). On the other hand, a weak associate such as Glue does not serve as an effective retrieval cue relative to free recall when the target words have been presented singly, but it is effective following presentation of the Glue-CHAIR pair. Such interactions between encoding and retrieval conditions can be interpreted on the assumption that recall results from the degree of overlap or the match between information in the encoded trace and that in the cues in the retrieval environment. When CHAIR is presented by itself, it is encoded in a way that overlaps with the later encoding of Table as a cue; when the Glue-CHAIR pair is presented, the encoding of CHAIR is altered so that there is no longer such great overlap between the encoding of CHAIR and that of Table, but now there is relatively great overlap with information in Glue.

Encoding specificity ideas have been made more precise by borrowing from the stimulus-sampling ideas of Estes (1959) and Bower (1967). Flexser and Tulving (1978) and Pellegrino and Salzberg (1975a, 1975b) have proposed that the internal representation of a stimulus may be regarded as a collection of features or elements that are sampled from the input stimulus. (These features may or may not refer to observable properties of the stimulus.) It is also assumed that in a cued recall or recognition test the internal representation of the cue in the test environment can similarly be represented as a set of features. Successful recall or recognition of the target experience is assumed to occur when there is a sufficient match of features encoded in the trace of the experience and those in the retrieval environment.

The issue that the present paper addresses is the nature of the features that are (by assumption) sampled at study and test. What sorts of features are encoded in the study and test situations for simple events such as presentation of words? A common assumption has been that words are encoded in terms of their meanings and, thus, that the features are semantic ones (e.g., Pellegrino & Salzberg, 1975a, 1975b). There have been several explications of encoding specificity phenomena, particularly the recognition failure of recallable words (Tulving & Thomson, 1973) in terms of a semantic interpretation of encoding specificity (e.g., Martin, 1975; Reder, Anderson, & Bjork, 1974). According to one version of this view, words have several senses, and when one studies a word in a list, an association is formed between a particular representation of the word and the fact

that it occurred in the list. The context at encoding helps determine what representation (or node) is tagged as having occurred in the list. This occurrence information must be accessible at the time of test for recall or recognition to be successful. If the context at the time of test leads to a node or representation of the word other than that encoded at study, recall or recognition of the word will not result. This particular version of the semantic interpretation of encoding specificity has been criticized by Tulving and Watkins (1977) on several grounds.

The semantic interpretation of encoding specificity was originally proposed by Tulving and Thomson (1973, p. 365) and can be described in the feature-sampling language of Flexser and Tulving (1978) and Pellegrino and Salzberg (1975a, 1975b). For example, with regard to the cued recall experiment of Thomson and Tulving (1970) discussed earlier, one might argue that when one studies CHAIR presented by itself in a list, semantic features regarding furniture are encoded. Sufficiently similar features are encoded when Table is later presented as a cue to provide overlap and thus to allow recall of CHAIR. However, when CHAIR is accompanied by Glue at input in a Glue-CHAIR pair, the semantic features encoded about CHAIR are a different subset from when it is presented alone. For example, the encoded features may be related to how furniture is made. Thus the features encoded from Table may not now overlap sufficiently with those in the representation of CHAIR to allow recall of CHAIR when Table is presented as a retrieval cue. However, presentation of Glue will provide sufficient overlap with the encoded representation of CHAIR to allow its recall.

The semantic feature interpretation of encoding specificity is consistent with the findings described above, but there are difficulties with the idea. In particular, attempts in several studies to manipulate the degree of semantic congruence between study and test of an item have shown surprisingly little influence on cued recall (Goldstein, Schmitt, & Scheirer, 1978; Tulving, 1974). Tulving (1974) presented target words for study in the presence of specific context words and then tested them with strong preexperimental associates of the targets as extralist retrieval cues. In two of three encoding conditions, the context words were intended to bias a meaning of the target word that was congruous or incongruous with the extralist cue used at recall. For example, a target word such as DAY was studied in the context of either Evening or Sun and then tested with presentation of Night as an extralist cue. It was expected that the encoding of DAY when accompanied by Evening would be more congruous with the meaning of the cue Night than when the context word was Sun. However, in two experiments there was little effect of the congruity variable; there was slight facilitation in cued recall in the congruous relative to the incongruous conditions, but the facilitation was rela-

tively small and not statistically reliable. Tulving (1974, p. 778) suggested that "specificity of encoding of word events in episodic memory transcends the semantic meaning of words."

The experiments by Goldstein et al. (1978) were based on logic similar to that of Tulving's (1974), but in their experiments the changes in meaning between study and test would seem to be more drastic since they employed homographs, words with at least two distinct meanings. If semantic features of words are encoded, one might expect that when one studies the pair Strawberry-JAM with the knowledge that recall for JAM will be requested later, a retrieval cue such as Raspberry would be more effective than cues tapping different meanings of JAM, such as Log or Traffic. In an experiment based on this logic, Goldstein et al. (1978, Experiment 1) found no difference in cued recall between conditions with cues intended to bias the same meaning of a target homograph or a different meaning. However, in a second experiment, in which subjects were given no special instructions at study (as in Experiment 1) or were told to generate rhymes, sentences, or images for items in a pair, recall cues tapping meanings similar to those induced at study produced better recall than those tapping different meanings, but only in the conditions in which subjects generated images or sentences when studying the word pairs. It was concluded (Goldstein et al., 1978, p. 18) that for subjects in the usual encoding specificity paradigm (i.e., without special instructions at encoding) "little semantic processing manifests itself in the utilization of the episodic trace . . . or at least that the semantic processing occurring is small relative to that under other conditioning" (sic).

The conclusion that there is little semantic processing of words or word pairs under intentional learning conditions seems to fly in the face of much contrary evidence. For example, subjects tend to falsely recognize semantically associated words in a recognition test more than unrelated lures (Underwood, 1965), and strong associates serve as effective retrieval cues relative to free recall when words are presented singly at input (Thomson & Tulving, 1970). Also, in recognition memory experiments varying the study and test context of homographs by biasing different meanings does result in large effects on performance (Light & Carter-Sobell, 1970; Thomson, 1972). These facts and others (e.g., Barclay, Bransford, Franks, McCarrell, & Nitsch, 1974) argue for a representation of word events in terms of their meaning. Yet if events are represented in terms of their meaning, why does the semantic congruity between the study conditions and test conditions not affect recall in the expected manner in the cued recall experiments of Goldstein et al. (1978) and Tulving (1974)?

The three experiments reported here address this question and propose a tentative answer. We will argue

that in experiments like those of Goldstein et al. (1978) and Tulving (1974) there is semantic encoding of the context word and target pairs and that the semantic congruity of the retrieval cue to the encoded episode affects recall. However, in order to observe such an effect, one must consider the target episode to the pair of items presented, not simply the member of the pair that the experimenter designates as the target. As will be shown, when subjects are presented word pairs, they often confuse the experimenter-designated target with the context word and recall the latter. When recall is scored appropriately, a semantic congruity effect is obtained even when not present in recall of targets.

The logic of Experiment 1 is similar to that in the experiments of Goldstein et al. (1978) and Tulving (1974), as well as those of Murphy and Wallace (1974) to be discussed below. Subjects studied pairs of words, one of which (the target) was a homograph, and were then tested with synonyms of the homographs as retrieval cues. The capitalized target word (e.g., ADDRESS) was paired with lowercase context word and then tested with a synonym cue for one of its meanings (Speech). There were four types of context words: A strong synonym of one meaning of the target (Talk), a weaker synonym of the same meaning of the target (Oration), a synonym that biased a different meaning of the target (Residence), or a word unrelated to the target (Hole). It was expected that recall of the word pair would be enhanced in the strong-synonym encoding condition relative to all others, but that recall should also be better in the weak-synonym condition than in the different-synonym condition. Thus manipulation of the synonym context words was intended to provide a gradient of semantic features similar to those in the retrieval cue, albeit a gradient that was ordinal in nature.

EXPERIMENT 1

Method

Subjects. The participants were 120 introductory psychology students at Purdue University who received partial credit for a course requirement by serving in the experiment. They were tested in groups of 10 to 30 in sessions lasting approximately 45 min.

Design and Materials. The study lists consisted of 20 pairs of words. One word in each pair, presented in capital letters, was the to-be-remembered word or target word. The other word in each pair, the context word, was presented in lowercase letters above the target word. The target words were homographs chosen such that each had at least three synonyms for one of its meanings and one synonym for its other meaning. Two people judged the three synonyms that biased one of the meanings as being either strong synonyms (quite close in meaning) or weak synonyms (less close in meaning). For each target word, two synonyms were judged as strong and one was judged as weak. One of the strong synonyms for each target word was chosen to be its extralist retrieval cue. The other three synonyms (two for one meaning, one for the other) and one unrelated word served as the context words for each target word in one of four different encoding conditions. The four encoding conditions were defined as (1) a strong synonym of the first meaning of the target word (strong synonym), (2) a weak synonym of the

first meaning of the target word (weak synonym), (3) a synonym of the second (different) meaning of the target word (different synonym), and (4) a word unrelated to the target word (unrelated word). For example, when the word ADDRESS served as the target, the strong-synonym condition was represented by the pair Talk-ADDRESS, the weak-synonym condition by the pair Oration-ADDRESS, the different-synonym conditions by the pair Residence-ADDRESS, and the unrelated word condition by the pair Hole-ADDRESS. The retrieval cue for ADDRESS was always Speech. Complete materials used in Experiment 1 are presented in Table 1.

Four lists were prepared such that, across the lists, each of the 20 target words appeared in each of the four encoding conditions. Within each list, five word pairs represented each encoding condition. The order of presentation of pairs was governed by a randomized block design so that in each successive block of four pairs, all encoding conditions were represented.

The lists just described were always the third list that subjects saw and recalled. Two other practice lists were also prepared so that the relation between the word pairs in these lists reflected the relation present between the context words and target words in the experimental lists. Subjects were given cued recall tests for the target words in the practice lists, with the context word serving as cues. The reason for the practice lists was to encourage subjects to encode the targets in relation to the context words.

Procedure. Participants were given general instructions informing them of the nature of their task. They were told that they would see word pairs with one word in uppercase letters and one word in lowercase letters. They were told that they would be tested for recall of only the capitalized words. However, they were also told to pay attention to the lowercase words and the relation between the two words, since this would later help them recall the target words.

The word pairs in all three lists were presented visually by means of a slide projector at the rate of 3 sec/pair. After presentation of each practice list, subjects attempted recall of the capitalized targets in response to the lowercase context words that were provided as cues on the first two pages of the recall booklets. Recall of practice lists will not be discussed further here.

Before presentation of the third, critical list, subjects were told that they would see another list of word pairs and that their task was still the same: They were to remember the capitalized word and note the relation between it and its corresponding context word. After presentation of the list, subjects were asked to turn to the third page of their recall booklets. They were given new instructions that informed them that their task was still to recall the capitalized target words but that this time the recall cues would not be the context words paired with target words. They were told that each recall cue should still remind them of one target word and that, if it did, they should write the target word in the space next to the cue. Subjects were told that if they remembered target words that were not in response to a recall cue, they were to write them at the bottom of the page. Five minutes were allowed for recall, within which subjects worked at their own pace. The word pairs and recall cues were presented to all subjects in the same order.

Results and Discussion

The primary results of interest from Experiment 1 are recall of the critical lists in which encoding conditions were varied. A stringent scoring criterion was used in the results reported here whereby a subject received credit for recalling a target word only if it was correctly paired with its retrieval cue. However, the conclusions do not change if a lenient scoring criterion is used, since few words were recalled to either the "wrong" retrieval cue or to no cue at all.

The mean number of target words recalled (of a possible six) under the four different encoding conditions were strong synonym, 2.64; weak synonym, 2.40; different synonym, 1.26; and unrelated word, 1.32. Encoding conditions reliably affected recall [$F(3,363) = 42.49, p < .001, \text{MSe} = 1.09$]. Newman-Keuls tests indicated that recall under the strong-synonym and

Table 1
Materials Used in Experiments 1 and 3

Target	Retrieval Cue	Context Word at Study			
		Strong Synonym	Weak Synonym	Different Synonym	Unrelated Word
ADDRESS	speech	talk	oration	residence	hole
APPEAL	plea	beg	canvass	attraction	insane
ARMS	weapons	guns	munitions	limbs	leopard
BALL	dance	party	reception	sphere	curfew
BANK	shore	beach	ridge	vault	soup
BUG	insect	fly	vermin	microphone	root
CHANGE	alter	vary	transform	coins	control
COUNT	add	compute	enumerate	earl	shame
DRAFT	draw	sketch	diagram	breeze	estate
FILM	flick	movie	cinema	coating	stem
GRAFT	swindle	corruption	defraud	join	white
JUNK	trash	waste	wreckage	boat	purpose
PACK	knapsack	bundle	bale	herd	melt
PUPIL*	learner	student	scholar	eye	magic
RACE	dash	sprint	competition	people	predict
RULER	queen	king	sovereign	yardstick	lullabye
SOUND	tone	note	timbre	healthy	muscle
SPELL	trance	charm	talisman	write	prairie
SPIRITS*	ghosts	fairies	demons	alcohol	guest
WATCH	stare	look	examine	clock	ego

*Deleted in Experiment 3.

weak-synonym encoding conditions differed reliably from that under the other two conditions. No other conditions differed reliably.

The results of this experiment indicate that retrieval cues are more effective when they are similar in meaning to the encoded meaning of the target word. When encoded in the context of synonyms, the target words were recalled better in the presence of a retrieval cue synonymous with the same sense of the targets than with a different sense. Thus, unlike the results of experiments by Goldstein et al. (1978) and Tulving (1974), clear evidence is provided here for the proposition that retrieval cues for words in a list are more effective if they tap the encoded meaning of the target word.

Recall did not differ reliably between the strong- and weak-synonym conditions. Thus there was no evidence in recall of target words for a weak sort of generalization gradient such that the closer in meaning the retrieval cue is to the encoded target, the better is recall. However, an analysis of errors provides more information concerning the meaning-specific effectiveness of retrieval cues and also provides evidence that the manipulation of the strong-synonym and weak-synonym contexts did exert an effect on recall. A common error made by subjects was recall of the context word seen at encoding rather than the target word. This fact is not surprising, since in two of the encoding conditions (strong and weak synonym) the retrieval cue was a synonym of the context word seen at input as well as the target. If we suppose that subjects encode the context and target words into a chunk or higher order unit, then at recall the retrieval cue may provide access to the higher order unit, but subjects may have difficulty in discriminating the target word from the context word. If this is the case, the pattern of context word intrusions may be expected to be greater for the strong- and weak-synonym conditions relative to the different-synonym and unrelated word conditions. The mean number of such context word intrusions was 1.14, .42, .12, and .06 for the strong-synonym, weak-synonym, different-synonym, and unrelated word conditions. Thus recall of context words shows an orderly decline with the degree of semantic relatedness assumed to exist between retrieval cues and the encoded trace. Further, unlike recall of target words, evidence provided by context word intrusions reveals a difference between the strong- and weak-synonym conditions; of 27 nontied subjects, 26 recalled more context words in the strong- than in the weak-synonym condition ($p < .001$, sign test).

If the scoring criterion of giving subjects credit for recalling the episode of a pair of words if they recall either the context or target word is adopted, then the proportion of episodes recalled under the four encoding conditions would be .63, .47, .23, and .23 for the strong-synonym, weak-synonym, different-synonym, and unrelated word conditions, respectively. The gradient

between the strong- and weak-synonym encoding conditions is revealed by this lenient criterion. However, regardless of scoring method, the results are consistent with the idea that the semantic relation between an encoded experience and cues in the retrieval environment is important in determining recall.

EXPERIMENT 2

The results of Experiment 1 lend support to the idea that the meaning of target words is an important coding dimension and that the congruence between meaning of the retrieval cue and the encoded trace is important in determining recall. Similar results have been reported by Murphy and Wallace (1974, Experiments 2 and 3), who also remarked on the tendency of subjects to intrude context words at recall. However, the results from Experiment 1 and those of Murphy and Wallace (1974) are all based on the use of homographs as target words. It would be reassuring to find evidence that the semantic congruity between an encoded trace and retrieval cue affects recall of words other than homographs. Such generality is quite likely, since any word, even low-frequency words that are believed to have few meanings (Reider et al., 1974), can be used in different senses and thereby encoded quite differently (e.g., Tulving & Watkins, 1977).

The purpose of Experiment 2 was to extend the generality of the findings of Experiment 1 by employing a design similar to that used in the previous experiment, but with target words that were not obvious homographs. The target words were those used by Tulving (1974), since there was no evidence from that experiment for a semantic congruity effect of target and cue. In Experiment 2 subjects studied 24 target words (e.g., OPEN) in the presence of a context word and were then tested with a retrieval cue that was a strong associate of the target (Closed). The context word presented with the target at study either was identical to the retrieval cue (Closed-OPEN) or was selected to bias a meaning congruous (Shut-OPEN) or incongruous (Country-OPEN) to the meaning of the cue. We expected to find that when subjects were scored as retrieving the target event when they recalled either the context word or target, congruous context words would lead to better recall later than would incongruous ones. A free recall condition was also included to determine the effectiveness of the retrieval cues following the various encoding conditions relative to a free recall baseline.

Method

All subjects saw three lists, two practice lists and one critical list. The practice lists used in Experiment 2 were the same as those used in Experiment 1. The critical list contained 24 word pairs. The pairs consisted of an uppercase target word with a lowercase context word printed above it. The target words and their high-associate retrieval cues were taken from Table 1 of Tulving's (1974) report. At input the target word was presented

with one of three types of context word: A congruous context word, which was a high associate of the target word and its retrieval cue; an incongruous context word, which was a low associate of the target and its cue; and an identical context word, which was the same as the target's retrieval cue. For example, when CHAIR was the target word, Table was the retrieval cue; Furniture was the congruous context word; Glue was the incongruous context word; and Table was the identical context word. Complete materials for Experiment 2 are presented in Table 2. A number of changes were made from Tulving's (1974) materials in selecting context words, particularly in the case of congruous words, in an attempt to vary congruity more effectively than in the original experiments. The changed words are noted in Table 2.

The three types of context words defined three encoding conditions relative to the retrieval cue: congruous, incongruous, and identical. Three critical lists were formed to counterbalance materials against conditions. Within each list, the order of the different encoding conditions was randomly determined. Each of three groups of 20 subjects received one of the critical lists containing eight target words in each of the three encoding conditions. Thus encoding conditions were varied within subjects. Three additional groups of 10 subjects were presented the same lists and treated identically until after presentation of the third, critical list. Then they were given an unexpected free recall test.

The general instructions, presentation and test conditions of the practice lists, and procedure for cued recall of the critical lists were the same as in Experiment 1. Following presentation of the critical list, all subjects were given a 5-min period for unpaced recall. For subjects recalling the critical list under cued recall conditions, four random orders of retrieval cues were constructed. Five subjects in each of the three groups received one of the cuing orders. Both cued and free recall subjects were instructed to recall only the capitalized target words.

Ninety University of Toronto undergraduates participated in the experiment for pay. Subjects were tested in groups of 8 to 12 in sessions that lasted about 45 min.

Table 2
Materials Used in Experiment 2

Target Word	Retrieval Cue	Context Word	
		Congruous	Incongruous
BALL	tennis	racquet*	whistle
BLUE	sky	heaven	pretty*
CHAIR	table	furniture*	glue
COAT	lining	inside*	covering
COLD	hot	warm*	ground
DAY	night	nighttime*	sun
DIRTY	clean	spotless*	barn
FLOWER	bloom	blossom*	fruit
FOOD	eat	swallow*	moth
GREEN	grass	lawn	cheese
HAND	finger	knuckle*	tool
HARD	soft	pliable*	glass
HIGH	low	sunken*	hope
LARGE	small	little*	stomach
LIGHT	dark	black	head
MAN	woman	female*	command
NEED	want	desire	bath
OPEN	closed	shut	country
ROUND	square	rectangle*	cabbage
SHORT	long	elongated*	stem
SLOW	fast	quick*	memory
WATER	lake	pond*	whiskey
WET	dry	arid*	cave
WIND	blow	whistle*	noise

*Different from words used in Tulving's (1974) experiment.

Table 3
Proportion of Words Recalled in the Conditions of Experiment 2

Encoding Condition	Cued Recall			Free Recall	
	T	C	E	T	E
Identical	.84		.84	.22	.24
Congruous	.65	.16	.81	.25	.27
Incongruous	.64	.06	.70	.29	.32

Note—T = target word; C = context word; E = either word.

Results and Discussion

The results pertaining to recall of the critical list are contained in Table 3. Consider first the cued recall of target words. A greater number of words were recalled when the context word paired with the target at study was identical to the retrieval cue (6.73 of 8 possible, or 84%) than in the conditions in which it was congruous with the cue (5.20, or 65%) or incongruous with it (5.10, or 64%). There was no apparent difference in recall of target words in the congruous and incongruous encoding conditions. These impressions were confirmed by an analysis of variance that revealed reliable variation among conditions [$F(2,177) = 14.29$, $p < .001$, $MSe = 3.52$] and Newman-Keuls tests that indicated that the only reliable difference in recall was between the identical condition and the other two conditions. Thus, despite our efforts to improve on the materials used by Tulving (1974) to make the congruous and incongruous encoding conditions more different, we still quite nicely replicated his effect: A retrieval cue semantically congruous with the encoding of a target word was no more effective than a cue tapping an ostensibly incongruous encoding of the target.

An examination of intrusions showed that, once again, subjects often recalled context words rather than target words and that this tendency occurred more frequently in the congruous than in the incongruous encoding condition. There was a mean of 1.25 context word intrusions in the congruous condition and .48 in the incongruous condition ($p < .001$ by sign test). When subjects were given credit for recalling the episode of a pair of words when they recalled either the context word or the target word, an effect of semantic congruity between the cue and target emerged. In the identical, congruous, and incongruous conditions, subjects recalled 84%, 81%, and 70% of the episodes, respectively. An analysis of variance revealed reliable variation among cuing conditions [$F(2,177) = 7.21$, $p < .005$, $MSe = 2.99$], and Newman-Keuls tests pinpointed the only reliable difference between the incongruous condition and the other two. Thus, whereas recall in the congruous encoding condition resembled that in the incongruous condition when subjects were scored only on target words, when they were scored on recall of either word in a pair, performance in the congruous conditions was much closer to that in the identical condition and reliably different from that in the incongruous condi-

tion. As in Experiment 1, evidence for the importance of semantic congruity between encoding and retrieval was revealed by the less stringent scoring criterion.

Recall under cued recall conditions was much greater than that under free recall conditions for all three encoding conditions whether subjects were scored on the basis of target word recall or recall of word pairs. Interpreted by the encoding specificity principle, this result indicates that even in the case of the nominally "incongruous" encoding conditions, there was rather great overlap between information in the trace and cue.

EXPERIMENT 3

The conclusion drawn from the results of Experiments 1 and 2 was that, contrary to the inference drawn from some other studies, the effectiveness of a retrieval cue depends on its semantic relation to an encoded trace. However, part of the evidence for this proposition was obtained only when a scoring criterion was adopted by which subjects were given credit for recalling the episode of a pair of words when they either recalled the target word (as instructed) or intruded the context word. In Experiment 3 three conditions of Experiment 1 were replicated, but at testing subjects were provided with extralist retrieval cues and told to try to recall both members of a pair of words to each cue. They were also instructed to report just one member of a pair if that was the only item they could recall. Mueller (1979) has also used a pair recall technique for other purposes. Thus the purpose of Experiment 3 was to replicate the semantic congruity effect of Experiment 1 when subjects were explicitly instructed to recall both members of a pair.

Method

All subjects again saw and recalled three lists, two practice lists and one critical list. The practice lists were those used in the previous experiments. The critical list contained 38 word pairs, but only 18 pairs were to be tested later. The 18 response words of the critical list were the same as those used in Experiment 1, with the exception of PUPIL and SPIRITS, which were randomly eliminated (see Table 1). The 18 word pairs represented three encoding conditions: strong synonym (Talk-ADDRESS), weak synonym (Oration-ADDRESS), and different synonym (Residence-ADDRESS). The unrelated word condition of Experiment 1 was not used, since it produced results quite similar to that of the different-synonym condition.

Three lists were prepared so that each critical target word was represented in each of the encoding conditions across lists. In each list six word pairs represented each encoding condition. Three groups of 24 subjects saw and attempted recall of each of the critical lists under cued recall conditions. Three additional groups of eight subjects saw and attempted recall of the same lists under free recall conditions.

The procedure was essentially the same as that of the previous experiments for study and test of the practice lists. Before presentation of the third, critical list, subjects were told that they would see a set of 38 word pairs and that they should again note the relation between words in a pair. Subjects were also told that the word pairs would be presented on either blue or white backgrounds, although they were not told that they

should pay special attention to words on either color. The 18 critical words that were later to be tested were presented on clear slides between two groups of 10 filler pairs that were on blue slides. This was done in an attempt to reduce the problem of ceiling effects in recall of only 18 critical target pairs. The filler slides were 20 pairs used in the critical list of Experiment 2.

After presentation of the critical list, subjects in the cued recall conditions were told that for this list the recall cues would not be the same as the context words they had seen at input. They were told that each recall cue they would see was intended to remind them of one word pair and, if it did, they should write both members of the pair next to the cue. Subjects were also told that they should write only one member of a pair if that was the only one they could think of. They were further instructed that the recall cues were only intended to remind them of words presented on white backgrounds and that these were the only pairs that need be recalled. Finally, subjects were told that if they recalled words or word pairs but could not match them to any cue, they should write them off to the right-hand side of the paper. The target words and the recall cues for the critical lists were presented in the same order for each of the three groups of subjects. Since the encoding conditions of the target words were systematically rotated across the three lists, order of presentation and order of recall were completely balanced across the three conditions across lists. The recall test lasted 5 min.

Free recall subjects were also given a 5-min period in which to recall as many pairs of words from the critical list as possible. They were also instructed that they need recall only the pairs that had white backgrounds and that they should recall only one word if they could not remember both words of a pair.

The 96 subjects served for pay in groups of 5 to 13 in experimental sessions lasting approximately 45 min. The experimental session concluded with the experimenter explaining the purpose of the experiment and answering subjects' questions.

Results and Discussion

The results presented in Table 4 are the proportion of words recalled in each of the encoding conditions for cued recall and free recall of the context words, the target words, and either the context or target words. If one considers the word pair to be encoded as a higher order unit, then scoring subjects for recall of either item in the unit as a measure of recall of the higher order unit is analogous to scoring subjects as recalling a category in categorized list recall whenever they recall at least one word from the category (Cohen, 1963).

Considering first cued recall of the target words, the results of Experiment 3 closely replicated those of the first experiment in that recall of homographs was greater when the target word was encoded in a manner semantically congruous rather than incongruous with the sense of the word elicited by the recall cue. However, there was little difference between cued recall of target

Table 4
Proportion of Words Recalled in the Conditions of Experiment 3

Encoding (Synonym) Condition	Cued Recall			Free Recall		
	C	T	E	C	T	E
Strong	.37	.36	.43	.06	.11	.12
Weak	.32	.33	.37	.19	.17	.23
Different	.06	.08	.08	.13	.15	.15

Note—C = context word; T = target word; E = either word.

words in the strong-synonym and weak-synonym encoding conditions. Examination of context word recall indicated somewhat better recall in the strong-synonym than in the weak-synonym condition, again replicating Experiment 1. An examination of higher order unit recall (recall of either the context word or the target) revealed best recall in the strong-synonym encoding condition, with recall of one or the other word to 43% of the cues, next best recall in the weak-synonym condition (37%), and quite poor recall in the different-synonym condition (8%). An analysis of variance showed reliable variation among conditions [$F(2,213) = 60.32$, $p < .005$, $MSe = 1.56$], and Duncan's tests indicated reliable differences between recall in the different-synonym condition and the other two conditions. There was no statistically reliable difference between recall in the strong- and weak-synonym conditions, but the result would seem to be reliable in that it occurred in both Experiments 1 and 3. The most likely conclusion is that the difference is small but real. At any rate, the main purpose of Experiment 3 was fulfilled: The basic results of Experiment 1 were replicated under conditions in which subjects were instructed to recall both context words and targets to the recall cues.

The comparison of cued recall and free recall can be interpreted within an encoding specificity framework. In the strong-synonym and weak-synonym encoding conditions, the recall cue was semantically similar to the supposed semantic content of the memory trace, and thus one would expect the extralist recall cues to be quite effective relative to free recall. In fact, this was the case: Cued recall of higher order units was almost four times greater than free recall in the strong-synonym condition. On the other hand, in the different-synonym encoding condition the recall cue would be expected to overlap minimally with the semantic features in the trace, and thus one would expect the extralist recall cues to be ineffective relative to free recall. This expectation was borne out, since cued recall in the different-synonym encoding condition was actually reliably worse than free recall both in recall of context words and targets ($p < .01$ by t tests in both cases). Thus, in agreement with a semantic version of encoding specificity, extralist recall cues are effective only when they match the meaning of a homographic target word that was encoded in the same semantic sense. Speech may be strongly associated with the target word ADDRESS preexperimentally, but it will not serve as an effective cue for ADDRESS if ADDRESS is encoded in the context of Residence in the study list. This finding is similar to one reported by Thomson and Tulving (1970).

Further evidence concerning cued recall in the three encoding conditions may be obtained by considering conditional probabilities of context word and target recall, which are shown in Table 5. In all three conditions there was a strong tendency to recall both the

Table 5

Recall	Synonym		
	Strong	Weak	Different
Context Word Target	.81	.84	.71
Context Word Target	.11	.06	.00
Target Context Word	.81	.87	1.00
Target Context Word	.11	.08	.03
Context Word First*	.61	.60	.21

*Proportion of times context word was recalled first.

context word and the target to the extralist cue or to recall neither. For example, if the target word was recalled, the probability that the context word was also recalled was .81, .84, and .71 in the strong-synonym, weak-synonym, and different-synonym conditions, respectively. The probability of recalling the context word if the target was not recalled was .11, .06, and .00 for the same conditions. As can be seen from Table 5, the results are similar when recall of targets is conditionalized on context word recall. Thus recall of the word pairs tends to occur in an all-or-none manner when subjects are instructed to recall both members of the pair to an extralist cue.

Another question of interest is whether or not the context word mediates recall of the target in any of the encoding conditions, or whether information in the extralist cue matches better the informational contents of the stimulus or response in the higher order unit representing the pair. Subjects were told to recall the words to a cue in any order. If we assume that the output order reflects the order in which the words occurred to the subjects, then for the strong-synonym and weak-synonym conditions there is evidence that the effect of the extralist cues tended to be mediated through, or overlap more with, the representation of the stimulus. In cases in which both members of the pair were recalled, the context word was recalled first 61% of the time in the strong-synonym encoding condition and 60% of the time in the weak-synonym encoding condition. This finding obtains despite the fact that subjects were expecting at the time of study to recall only the target words. On the other hand, in the different-synonym condition subjects seemed to mediate recall of context words through recall of targets. Subjects never recalled a context word when they did not also recall a target, and on those occasions when both were recalled, the target was recalled first 79% of the time. Thus, when subjects studied pairs like Residence-ADDRESS and then received cues like Speech, they recalled the target ADDRESS infrequently, but they never recalled Residence unless they also recalled ADDRESS.

GENERAL DISCUSSION

The results of the present experiments provide evi-

dence that semantic factors play a large role in cued recall. In both Experiments 1 and 3 it was found that recall cues for homographs produced much better recall when they matched the sense of the word induced by the encoding context than when they did not. Murphy and Wallace (1974) reported a similar result. In Experiment 3 cued recall of homographs encoded in a sense different from that represented by the recall cue was actually worse than free recall of the same targets. Evidence was also found that, even when the two encoding contexts induced the same interpretation of the homograph, recall was better when the context word was more nearly synonymous with the target and extralist cue than when it was somewhat less synonymous with these items. However, the latter finding depended upon scoring subjects as having recalled the episode of a word pair if they recalled either member of the pair.

It may be that failures to find a semantic congruity effect in other studies (Goldstein et al., 1978; Tulving, 1974) were partly due to scoring subjects strictly on correct recall of the target word. In the case of congruous encoding of a context word and target (i.e., congruous with respect to the later recall cue), subjects may form a higher order unit of the word pair, and in cued recall subjects may produce the "wrong" target from the experimenter's viewpoint. In fact, there was evidence of greater intrusion of context words in both Experiments 1 and 2 as the congruity between the encoding context and the recall cue increased. The argument that previous failures to find a semantic congruity effect in cued recall are due to the scoring technique adopted is bolstered by the fact that in Experiment 2, which replicated the cued recall aspect of Tulving's (1974) experiments, no congruity effect was found when Tulving's scoring system was adopted (subjects were given credit for recalling the episode only if they recalled the target), whereas a congruity effect was found when subjects were given credit for recalling the episode when either the context word or the target was recalled. The pair recall method used in Experiment 3, in which subjects are asked to recall both members of a pair to an extralist cue, seems more informative and generally preferable to the usual practice of requesting recall of only the member of the pair designated as the target by the experimenter.

One unresolved issue is why the results of Experiments 1 and 3 did not replicate those of Goldstein et al. (1978, Experiment 1). In our experiments subjects recalled homographs better when the retrieval cues biased a similar rather than a different meaning of the target, regardless of the scoring criterion employed. Goldstein et al. (1978) found no such effect except when subjects were given special elaborative processing instructions at study (Experiment 2). It is difficult to ascertain why there was this discrepancy in outcome, except to note that the sets of materials used were quite different in the two studies. An examination of

the materials used by Goldstein et al. (1978) permits the speculation that subjects may in some cases have recalled targets to cues other than those intended by the experimenter to be appropriate, because cues with similar meanings were included.¹ For example, subjects studied Florida-PALM in some conditions and were cued with either Tropical to tap a similar meaning of PALM or Sweaty or Clammy to tap a different meaning. Recall of PALM to Tropical may have been diminished because subjects were given other cues that could also have overlapped with the encoding of PALM in the Florida-PALM pair, such as Hawaiian and Ocean. Other instances of the same nature also are present in their materials. If recall was affected by such problems in matching cues to targets, it may have taken the more elaborative processing to produce even the relatively weak semantic congruity effects found in Experiment 2. An important consideration in selecting materials for cued recall experiments is that cues overlap or match only one target. This could be determined in pretests in which subjects are given the cues and the targets and asked to match them. Items upon which there is not good agreement could be discarded. Although such a procedure was not explicitly followed in the present experiments either, it may have been that the cues overlapped each other less in our lists. Whether or not differences in materials between the experiments produced the discrepancy in results, it would appear that the semantic congruity effect obtained in the present experiments is reliable, in that Murphy and Wallace (1974) obtained similar effects.

Within the context of theories that consider the memory trace to be composed of a bundle of features or attributes and the retrieval process to be a matching of features in the retrieval environment to those in the trace, the present results are consistent with the idea that an important set of features is semantic. It is unlikely that we shall, in the near future, be able to characterize these features precisely, and we are even further from being able to provide systematic generalization gradients along a semantic dimension. The fact that in the present experiments (at least Experiment 1) cued recall was better following strong- than following weak-synonym encoding conditions at least provides some comfort that such gradients exist. Other investigators have also found that the effectiveness of retrieval cues depends on the contextual meaning of material, even when the material is not explicitly homographic in nature (e.g., Barclay et al., 1974).

Despite the importance of semantic factors in cued recall, as demonstrated in the present experiments, other findings lead one to believe that additional features are also quite important, although we know little about these features. For example, Slamecka and Barlow (1979) have recently reported three experiments in which subjects received repetitions of homographs with various contextually biasing words. In Experiment 3 subjects

saw pairs like King-Ruler, and then later, either the pair was repeated intact or Ruler was presented with a context word biasing the same sense (Leader-Ruler) or a different sense (Pencil-Ruler). The question of interest was the relative ordering of these three conditions in cued recall when the cue was the first context word presented (i.e., King). Slamecka and Barlow (1979) found that cued recall was best in the intact pair case and did not differ in the other two. That is, it did not matter whether the repetition of the homograph was accompanied by a context word that biased the same meaning as originally encoded or by a word that biased a different meaning. The benefit from the repetition was the same in both cases. This finding argues against the role of semantic features in producing this repetition effect and for what Slamecka and Barlow refer to as "surface features" or the "outward appearance" of the item. Other evidence from their experiments makes it clear that semantic factors were important in their cued recall task, but nonetheless, there seems to have been a large nonsemantic component as well. An important task for future research is the identification and elucidation of the other features that govern recall.

REFERENCES

- BARCLAY, J. R., BRANSFORD, J. D., FRANKS, J. J., MCCARRELL, N. S., & NITSCH, K. Comprehension and semantic flexibility. *Journal of Verbal Learning and Verbal Behavior*, 1974, **13**, 471-482.
- BOWER, G. H. A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 1). New York: Academic Press, 1967.
- COHEN, B. H. Recall of categorized word lists. *Journal of Experimental Psychology*, 1963, **66**, 227-234.
- ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 2). New York: McGraw-Hill, 1959.
- FLEXNER, A. J., & TULVING, E. Retrieval independence in recognition and recall. *Psychological Review*, 1978, **85**, 153-171.
- GOLDSTEIN, E., SCHMITT, J. C., & SCHEIRER, C. J. Semantic effects in encoding specificity: A levels of processing approach. *Memory & Cognition*, 1978, **6**, 13-19.
- GUTTMAN, N., & KALISH, H. I. Discriminability and stimulus generalization. *Journal of Experimental Psychology*, 1956, **51**, 79-88.
- HOLLINGWORTH, H. L. *Psychology: Its facts and principles*. New York: Appleton, 1928.
- LIGHT, L. L., & CARTER-SOBELL, L. Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 1-11.
- MARTIN, E. Generation-recognition theory and the encoding specificity principle. *Psychological Review*, 1975, **82**, 150-153.
- MCGEOCH, J. A. *The psychology of human learning*. New York: Longmans, Green, 1942.
- MUELLER, C. W. *Recognition failure of recallable words: Methodology or memory trace?* Unpublished doctoral dissertation, University of Toronto, 1979.
- MURPHY, M. D., & WALLACE, W. P. Encoding specificity: Semantic change between storage and retrieval cues. *Journal of Experimental Psychology*, 1974, **103**, 768-774.
- PELLEGRINO, J. W., & SALZBERG, P. M. Encoding specificity in associative processing tasks. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, **1**, 538-548. (a)
- PELLEGRINO, J. W., & SALZBERG, P. M. Encoding specificity in cued recall and context recognition. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, **1**, 261-270. (b)
- REDER, L. M., ANDERSON, J. R., & BJORK, R. A. A semantic interpretation of encoding specificity. *Journal of Experimental Psychology*, 1974, **102**, 648-656.
- SLAMECKA, N. J., & BARLOW, W. The role of semantic and surface features in word repetition effects. *Journal of Verbal Learning and Verbal Behavior*, 1979, **18**, 617-627.
- SPENCE, K. W. The nature of discrimination learning in animals. *Psychological Review*, 1936, **43**, 427-449.
- THOMSON, D. M. Context effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 497-511.
- THOMSON, D. M., & TULVING, E. Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, 1970, **86**, 255-262.
- TULVING, E. Recall and recognition of semantically encoded words. *Journal of Experimental Psychology*, 1974, **102**, 778-787.
- TULVING, E., & OSLER, S. Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology*, 1968, **77**, 593-601.
- TULVING, E., & THOMSON, D. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, **80**, 352-373.
- TULVING, E., & WATKINS, M. J. Structure of memory traces. *Psychological Review*, 1975, **82**, 261-275.
- TULVING, E., & WATKINS, O. C. Recognition failure of words with a single meaning. *Memory & Cognition*, 1977, **5**, 513-522.
- UNDERWOOD, B. J. Stimulus selection in verbal learning. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes*. New York: McGraw-Hill, 1963.
- UNDERWOOD, B. J. False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, 1965, **70**, 122-129.

NOTE

1. We would like to thank Endel Tulving for pointing out this possibility.

(Received for publication June 13, 1979;
revision accepted September 28, 1979.)