# The effects of "effort after meaning" on recall: Differences in within- and between-subjects designs

**FRANKLIN M. ZAROMB AND HENRY L. ROEDIGER III**
*Washington University, St. Louis, Missouri*

In four experiments, we examined free recall of ambiguous sentences with or without corresponding cues to facilitate comprehension, using Auble and Franks's (1978) paradigm to examine effort after meaning (Bartlett, 1932). The ambiguous sentences were studied without cues, with cues meaningfully embedded in them, with cues provided shortly before the sentence (precue), or with cues following the sentence after several seconds (delayed cue). When these conditions were manipulated within subjects, the process of cue integration in the pre- and delayed-cue conditions enhanced recall, relative to the no-cue and embedded-cue conditions. Furthermore, in a test condition in which subjects first attempted to recall the cues alone, recall was also best in the delayed-cue condition. The effects described above did not occur when the cue presentation conditions varied between subjects, and on a test of order reconstruction, there was even an advantage for sentences studied in the embedded-cue condition over those studied in the delayed-cue condition. The dissociative effects of experimental design on sentence recall and order reconstruction suggest that effort after meaning might enhance memory for study items at the cost of impairing memory for temporal order information.

What is the relationship between comprehension and memory? Ebbinghaus (1885/1964) began the scientific study of memory by attempting to avoid this question. By studying memory for three-letter trigrams under controlled experimental conditions, he assumed that he would not have to be concerned about the variability in how subjects understood or related to the stimuli. Nonetheless, Ebbinghaus noted that the lists of syllables he memorized

> exhibit very important and almost incomprehensible variations as to the ease or difficulty with which they are learned. It even appears from this point of view as if the differences between sense and nonsense material were not nearly as great as one would be inclined a priori to imagine. (p. 23)

Others who followed in Ebbinghaus's footsteps measured differences in the meaningfulness of nonsense syllables (e.g., Archer, 1960; Glaze, 1928; Hull, 1933) in order to afford researchers better experimental control over these presumed variations in learning (see also Jenkins, 1985). Even the plainest materials seem to require an effort to find meaningfulness in order to comprehend and remember them (e.g., Prytulak, 1971).

In contrast to Ebbinghaus's (1885/1964) approach, Bartlett (1932) presumed a direct relationship between processes involved in comprehension and acts of remembering. Specifically, Bartlett proposed that individuals engaged in a process of "effort after meaning," of which he wrote,

it is fitting to speak of every human cognitive reaction—perceiving, imaging, remembering, thinking, and reasoning—as an effort after meaning. Certain of the tendencies which the subject brings with him into the situation with which he is called upon to deal are utilized so as to make his reaction the "easiest," or the least disagreeable, or the quickest and least obstructed that is at the time possible. When we try to discover how this is done we find that always it is by an effort to connect what is given with something else. (p. 44)

In experiencing the world, humans presumably bring their past knowledge to bear on understanding current happenings. When it comes to remembering past occurrences, Bartlett argued that people attempt to make their recollections conform to cultural expectations, prior background knowledge, or the current context.

To directly examine effort after meaning as it relates to memory, Auble and Franks (1978) developed an ingenious experimental paradigm in which subjects heard and later attempted to recall ambiguous sentences, such as "The clothes were ruined because the sign vanished," that were presented with or without disambiguating cues (i.e., wet paint). Where cues were provided, their conditions of presentation varied. In a control condition, the cue was meaningfully embedded within the sentence (i.e., "The clothes were ruined because the wet paint sign vanished"). In other conditions, the sentence preceded the cue. Sometimes, the

**F. M. Zaromb, fmzaromb@artsci.wustl.edu**

cue occurred immediately after the sentence, whereas in other cases, the cue was delayed by 5 sec. Auble and Franks found superior recall in the last condition, in which subjects had to puzzle over the meaning of the sentences for several seconds before hearing the cues. They took this as evidence for a memorial benefit of effort toward comprehension or, in Bartlett's terms, effort after meaning.[1]

In subsequent work, Auble, Franks, and Soraci (1979) observed that the memorial benefit of effort after meaning occurs even when sentences are learned that are relatively easy to understand, such as "The colors appeared because the rain stopped." Auble et al. found that even though most subjects were able to figure out the meaning of these sentences without hearing the corresponding cue words (e.g., *rainbow*), recall was still enhanced in the delayed-cue condition, relative to the embedded-cue condition, and nearly identical results were obtained with more ambiguous sentences. Thus, Auble et al. argued that the effect may be attributed to an "aha!" reaction that occurs when an initially ambiguous stimulus is suddenly comprehended, and it does not matter how difficult the original stimulus is or how long it may take to understand, so long as it is at least momentarily ambiguous and comprehension is ultimately achieved.

Surprisingly, very little follow-up research has been done to directly identify the cognitive processes that underlie effort after meaning. The terms *effort after meaning* (Bartlett, 1932) and *effort toward comprehension* (Auble & Franks, 1978) suggest that some form of cognitive effort is involved in attempting to understand a to-be-learned stimulus and that this effortful process directly influences retention. If so, what is the exact nature of this effort, and what factors influence the amount and quality of effort devoted to stimulus comprehension? Furthermore, how does effort after meaning relate to the cognitive effort required to perform other encoding tasks, such as item generation, semantic elaboration, or anagram problem solving (e.g., Craik & Tulving, 1975; Jacoby, 1978; Slamecka & Graf, 1978; Tyler, Hertel, McCallum, & Ellis, 1979)? The latter question is underscored by the fact that cognitive effort is poorly understood in general, and prior research examining the effects of cognitive effort on memory performance has yielded mixed results (Ellis, Thomas, & Rodriguez, 1984; Zacks, Hasher, Sanft, & Rose, 1983).

By the same token, if the cognitive operations or phenomenological experience associated with achieving comprehension directly influence retention, as is implied by the explanation of the "aha!" effect (Auble et al., 1979), the exact nature of this reaction remains unclear. Is it the integration of the new stimulus with the subject's prior knowledge to form a coherent mental representation? Is it the distinct phenomenological experience of comprehension, with all that it might entail in terms of arousing attention, conscious awareness, or further reflection or enhancing one's mood? Or is it a combination of these elements or something else altogether?

One starting point for empirically studying these fundamental questions is to consider the similarities and differences between the memorial advantage of effort after meaning and other, more extensively studied mnemonic

effects in the psychological literature. In this regard, the memorial benefit of effort after meaning does appear to be similar to the generation effect, which is a robust finding and has engendered a sizable literature, in which there is a memorial advantage for information that is actively generated rather than passively heard, seen, or read (e.g., Jacoby, 1978; Slamecka & Graf, 1978; for recent reviews, see Bertsch, Pesta, Wiscott, & McDaniel, 2007; Mulligan & Lozito, 2004). On a general level, due to the active or synthetic nature of comprehension, effort after meaning appears to be an active process. More specifically, the Auble and Franks (1978) paradigm is similar to generation tasks that require subjects to solve verbal or mathematical problems, as opposed to passively receiving the solutions (Crutcher & Healy, 1989; Erdelyi, Buschke, & Finkelstein, 1977; Jacoby, 1978). The purpose of the present research is to shed light on the processes underlying the memorial advantage of effort after meaning by asking whether the same variables that modulate generation effects will also determine the effects of effort after meaning in the Auble and Franks paradigm. We first will review relevant prior literature, before describing our own research.

## Effort After Meaning and Generation

Wills, Soraci, Chechile, and Taylor (2000) investigated the intersection between effort after meaning and generation as they relate to memory by reexamining Peynircioğlu's (1989) finding of a generation effect in which recall for pictures generated by "connecting the dots" was superior to that for pictures that were simply traced by subjects. Specifically, Wills and colleagues asked whether this generation effect was due to the presence or absence of an "aha!" reaction. In one experiment, subjects viewed and were later tested on their recall of pictures of objects under three study conditions. One group of subjects generated each picture by connecting dots; another group drew each figure by simply tracing over an already visible outline; and the last group visually scanned the full picture of each object.

To test for possible effects of an "aha!" reaction, subjects were given either a preview or a postview of the final image. Presumably, connecting dots should elicit an "aha!" reaction only in the postview condition, because the final form would remain ambiguous until the drawing was complete, at which point the subjects would finally understand what it was. As for the generation effect observed by Peynircioğlu (1989), the results showed a recall advantage for connecting the dots. More important, this advantage was driven by performance in the postview condition. Wills et al. (2000) argued that the critical feature producing enhanced recall was not so much the fact that the subjects generated the image, because they also did that in the preview condition. Rather, in the postview condition, an "aha!" reaction occurred as the subjects connected the dots, transforming an ambiguous array of dots into a concrete image.

In contrast to Auble et al.'s (1979) account of the "aha!" reaction, Wills et al. (2000) explained their findings in terms of the multiple-cue account of the incongruous item generation effect (Soraci et al., 1999, Soraci et al., 1994). Several studies have shown that the presence and

magnitude of the generation effect depend on the nature of the encoding rule, as well as the number and type of candidate responses that come to mind while subjects attempt to generate target study items. For example, the word fragment "c_p" may be completed with three possible solutions (i.e., *cap*, *cop*, *cup*). Under typical item generation instructions, a subject might be asked to provide a *congruous* response to the following encoding rule: "A policeman: c_p." However, if subjects are instructed to provide an *incongruous* response based on the encoding rule, "Not a policeman: c_p," the word *cop* and at least one of the other two solutions will become mentally activated. Similarly, with the cue, "Not a policeman, not a hat: c_p," all three words—*cop*, *cap*, and *cup*—are likely to come to mind. Soraci et al. (1994) found that a word presented as a fragment such as "c_p" is more likely to be recalled under incongruous than under congruous item generation or read-only study conditions.

Wills et al. (2000) suggested that the memorial advantage observed for the recall and recognition of pictures might be related to a process of generating possible candidate solutions to an ambiguous array of dots before connecting them to create the final image of a concrete object. Although these candidate solutions may be incorrect, they may still serve as retrieval cues to facilitate recall of the correct image. Similarly, Carlin, Soraci, Dennis, Chechile, and Loiselle (2001) observed that pictures that started out of focus and faded in were recalled better than pictures that started in focus and faded out. According to the multiple-cue hypothesis, the recall advantage stemmed from subjects' entertaining and revising candidate solutions as the pictures came into focus (see also Wills, Estow, Soraci, & Garcia, 2006).

The aim of the present experiments was to examine the memorial advantage of effort after meaning by testing whether this effect is determined primarily by the processes involved in suddenly comprehending an ambiguous stimulus (e.g., the "aha!" effect) or by the generation of multiple solutions during study (e.g., the multiple-cue hypothesis). One approach to distinguishing between the two accounts is to vary the placement of the disambiguating cue in relation to its corresponding sentence at study. Experiment 1 compared recall for ambiguous sentences that were studied with cues meaningfully embedded, presented shortly beforehand (precue condition), and presented after either a short (2-sec) or a long (10-sec) delay interval (delayed-cue condition).

According to the multiple-cue hypothesis, extending the interval between sentence and cue should improve recall performance, relative to the shorter cue delay, because subjects have additional time to generate potential cues in their efforts to understand the ambiguous sentences, which may, in turn, aid subsequent retrieval. Furthermore, a delayed-cue condition should elicit greater recall performance than should a precue condition, because the earlier presentation of the cues should constrain or possibly preempt the generation of comprehension-facilitating cues during study. By contrast, if the processes involved in achieving comprehension are critical to enhancing memory, the length of the sentence–cue delay interval should not differentially affect recall performance, and to the extent that a precue is itself somewhat ambiguous to subjects until they actually hear the disambiguating sentence, there should also not be a difference in recall performance between the pre- and delayed-cue presentation conditions.

## Effort After Meaning and Experimental Design

On the basis of studies of congruous and incongruous item generation effects, it may also be possible to distinguish between the two accounts of the memorial advantage of effort after meaning through the manipulation of experimental design. There is a curious tendency for the generation effect (with verbal stimuli) to disappear in between-subjects and blocked-list within-subjects designs (Begg & Snider, 1987; Hirshman & Bjork, 1988; Slamecka & Katsaiti, 1987). By contrast, incongruous item generation effects have been demonstrated in between-subjects designs (Soraci et al., 1999; Soraci et al., 1994). The influence of experimental design is by no means limited to the generation effect, for a host of other established memory phenomena, such as the word frequency, bizarreness, enactment, and perceptual interference effects, are modulated by experimental design, or list composition, as well (for a recent review, see McDaniel & Bugg, 2008). Recent studies have demonstrated that variations in experimental design can also reveal striking dissociations on measures of recall and recognition, as well as temporal order memory, providing converging evidence for the view that unusual encoding conditions, such as generation or bizarre imagery, attract superior processing of individual stimulus features at the cost of disrupting the encoding of relations among stimuli or other features of the current episodic context (DeLosh & McDaniel, 1996; McDaniel & Bugg, 2008; Nairne, Riegler, & Serra, 1991; Serra & Nairne, 1993). Thus, the manipulation of experimental design has become an important tool for testing theories of human memory.

In Auble and Franks's (1978) paradigm, the memorial benefit of effort after meaning has been demonstrated only in within-subjects designs using tests of free recall. Thus, it is worth asking whether this memorial advantage persists across variations in experimental design and types of memory tests. Experiment 2 tested whether the original findings of Auble and Franks and Auble et al. (1979) would extend to a between-subjects experimental design and test of order reconstruction. Experiments 3 and 4 also used within- and between-subjects experimental designs to further examine the memorial advantage of effort after meaning. Specifically, Experiment 3 compared recall for sentences studied with or without disambiguating cues. If the resolution of ambiguity is critical to remembering the sentences, recall for sentences studied with cues should exceed that for sentences studied without cues, even though subjects might exert a considerable amount of mental effort attempting to understand ambiguous sentences in the latter condition, perhaps even generating potential retrieval cues, as predicted by the multiple-cue hypothesis. Last, Experiment 4 tested whether sentence recall is associated with the ability to remember the corresponding disambiguating cues.

## EXPERIMENT 1

In Experiment 1, we examined the effects of effort after meaning on recall by comparing recall performance for ambiguous sentences studied in precue, delayed-cue, or embedded-cue presentation conditions in a within-subjects design. According to the multiple-cue hypothesis, extending the interval between sentence and cue presentation should enhance recall, because subjects have additional time to generate potential retrieval cues in their efforts to understand the ambiguous sentences. Furthermore, the earlier presentation of the cue in the precue condition should make the sentence readily comprehensible (similar to the embedded-cue condition) and constrain or even preempt the generation of comprehension-facilitating cues during study.

Alternatively, one might argue that because the cue and sentence are presented separately in the precue condition, effort after meaning is still required to integrate an initially ambiguous cue with its corresponding sentence to achieve comprehension. Consistent with the "aha!" reaction account, it is possible that providing a short interval between sentence and cue, regardless of their order of presentation, improves recall, because both conditions involve the mental transformation of an originally ambiguous stimulus into a comprehended representation. Thus, if recall performance in the pre- and delayed-cue conditions is superior to that in the embedded-cue condition, this would suggest that the separation of sentence and cue, regardless of presentation order, is important to producing a memorial advantage of effort after meaning.

### Method

**Subjects**. Fifty-four Washington University undergraduates participated for either course credit or payment.

**Materials and Procedure**. The stimuli were 24 difficult-to-comprehend sentences and their corresponding cues sampled from Auble et al. (1979; several additional sentences and cues were created; see Appendix A). The sentences and cues were voice recorded on an IBM computer using Sound Forge 5.0 software (Sonic Foundry, Inc.). E-Prime experimental software (Psychology Software Tools, Inc.) was used for stimulus presentation and recording subjects' keyboard responses.

Sentence–cue presentation during study was manipulated within subjects. Six sentences were presented with cues meaningfully embedded (embedded-cue condition). For example, the sentence, "The home was small because the sun came out" (clue: IGLOO), was presented as "The IGLOO was small because the sun came out." Six sentences were preceded 2 sec earlier by cues (precue condition). The remaining sentences were followed by cues at 2- or 10-sec intervals (delayed-cue 2- and 10-sec conditions, six sentences each). Sentence order was randomized, with the constraint that, across subjects, each sentence was presented an equal number of times in each of the four conditions. The total study time for each sentence was 20 sec, which included 5 sec for sentence presentation and 2 sec for cue presentation (see Appendix B).

The experimenter initially instructed the subjects that they would study and later be asked to recall a series of ambiguous sentences and that, as they learned each sentence, additional clue words might be heard either prior to or following the sentence to facilitate comprehension. Following auditory presentation of each sentence and the possible cue, the subjects were prompted by a computer display to indicate how well they understood the sentence by pressing one of three keys: "1" if they understood the sentence, "2" if they did not understand the sentence, and "3" if they were unsure. The subjects were allowed 5 sec to press a key.

Following a 5-min number-matching distractor task, the subjects attempted to type as many and as much of the sentences as they could remember in any order into a blank Microsoft Word document for 10 min. The experimenter instructed the subjects to mark the passing of each minute by typing the current interval into the document following the most recent response. This was done in order to measure cumulative recall performance (how recall unfolds over time; see Wixted & Rohrer, 1994). The experimental session lasted about 50 min.

### Results

**Comprehension**. Sentence comprehension was measured as the proportion of sentences rated as being understood (i.e., given a rating of 1). The sentences were generally understood, as indicated by the high rates of perceived comprehension in the embedded-cue ($M = .74$, $SD = .23$), precue ($M = .77$, $SD = .22$), delayed-cue 2-sec ($M = .81$, $SD = .20$), and delayed-cue 10-sec ($M = .79$, $SD = .20$) conditions. An ANOVA confirmed that comprehension ratings did not significantly differ across the cue presentation conditions [$F(3,159) = 1.08$, $MS_e = .04$, n.s., $\eta_p^2 = .02$].

**Sentence recall**. The recall results were scored in terms of the number of sentences or close paraphrases recalled per subject. For example, a subject received full credit (1 point) for the response, "The woman couldn't see well because the blade was broken," to the presented sentence, "The woman could not see because the blade was bent," even though a couple of the words differed between the two sentences. Full credit was also given to paraphrases that shared fewer identical words but were clearly consistent with the meaning of the study sentences. Thus, the recalled sentence and cue, "The food was tasty because the threads were there (spiderweb)," received full credit, because it clearly corresponded to the presented sentence, "The breakfast was excellent because the thread was sticky." Moreover, because each sentence consisted of two clauses, partial credit was given if only one of the two clauses of a sentence was correctly recalled. For example, subjects received partial credit (0.5 points) for recalling the following sentence fragments in isolation: "The breakfast was excellent because . . ." or ". . . because the signs came up." Although the results of the present experiments were based on scoring done by the experimenter, a second independent rater, who was blind to the experimental design, scored recall protocols for 48 subjects, and the measures of recall performance derived from the two sets of scored data correlated nearly perfectly ($r = .98$). Last, we measured and plotted cumulative recall performance during the testing period for each of the sentence presentation conditions in all four experiments. We will withhold discussion of these cumulative recall functions until the General Discussion section.

Recall was highest in the delayed-cue 10-sec condition ($M = .43$, $SD = .22$), followed by the precue ($M = .40$, $SD = .20$) and delayed-cue 2-sec ($M = .37$, $SD = .19$) conditions, and was poorest for sentences with embedded cues ($M = .26$, $SD = .26$). An ANOVA indicated a significant effect of cue presentation condition on sentence

recall [$F(3,159) = 8.40$, $MS_e = .04$, $\eta_p^2 = .14$]. Individual pairwise comparisons revealed that, as compared with the embedded-cue condition, recall was significantly higher for sentences studied in the precue 2-sec [$t(53) = 3.78$, $SEM = .04$, $d = 0.57$], delayed-cue 2-sec [$t(53) = 3.34$, $SEM = .03$, $d = 0.58$], and delayed-cue 10-sec [$t(53) = 4.59$, $SEM = .04$, $d = 0.83$] conditions. For these and subsequent sets of individual comparisons, the Type I error rate was controlled at .05 using the false discovery rate (FDR) procedure (Benjamini & Hochberg, 1995; Benjamini & Yekutieli, 2001). However, recall performance in the delayed-cue 10-sec condition was not significantly different from that in the precue condition [$t(53) = 0.91$, $SEM = .04$, n.s., $d = 0.12$] or that in the delayed-cue 2-sec condition [$t(53) = 1.65$, $SEM = .04$, n.s., $d = 0.29$]. Last, the difference in recall between the delayed-cue 2-sec and precue conditions was not significant [$t(53) = 0.79$, $SEM = .04$, n.s., $d = 0.12$]. Thus, recall was superior in the pre- and delayed-cue conditions, relative to the embedded-cue condition, but there were no significant differences between the pre- and delayed-cue conditions.[2]

## Discussion

The results of Experiment 1 demonstrated that requiring subjects to puzzle over the meaning of sentences for several seconds, prior to comprehension, enhanced recall, relative to an embedded-cue condition in which the sentences were readily understood. These data are consistent with the results of prior studies and corroborate the notion of a memorial advantage for effort after meaning (Auble & Franks, 1978; Auble et al., 1979). In addition, recall performance was similarly enhanced when the cues were presented before their corresponding sentences, which suggests that it is not the placement of the cue in relation to the sentence but, rather, the temporal separation of the sentence and cue, regardless of presentation order, that produces the memorial benefit of effort after meaning.

Taken together, these results appear inconsistent with the multiple-cue hypothesis, which predicts that more cues should be generated in the delayed-cue conditions, thereby providing a greater number of retrieval cues to aid subsequent recall performance. Rather, these results are consistent with the view that the processes involved in the sudden comprehension of an ambiguous stimulus (e.g., the "aha!" effect) enhance memory, because both the pre- and delayed-cue conditions required subjects to integrate the sentences with their corresponding cues in order to achieve comprehension and produced superior recall performance, as compared with the embedded-cue condition, in which sentences were readily comprehended.

## EXPERIMENT 2

In Experiment 2, we examined whether the original findings of Auble and Franks (1978) and Auble et al. (1979) would extend to a between-subjects experimental design and test of serial order memory. Two conditions (embedded cue and delayed cue 10 sec) were used from Experiment 1. For one group of subjects, the cue presentation conditions varied within subjects, and for the remaining two groups, the conditions varied between subjects. All three groups studied one list of sentences, followed by a test of free recall, and then another list, followed by an order reconstruction test.

If recall is enhanced in the delayed-cue condition in the within-subjects design, but not in the between-subjects design, this result would suggest that the memorial benefit of effort after meaning is comparable to other effects that are sensitive to variations in experimental design, such as the generation effect. In addition, the order reconstruction task provides the opportunity to further examine this comparison by testing specific theoretical predictions. Of particular interest is the question of whether study conditions that are designed to elicit greater effort after meaning enhance memory for ambiguous sentences at the cost of impairing memory for the temporal order in which the sentences occur over the course of study. The item-order account of the generation effect (Nairne et al., 1991; Serra & Nairne, 1993), which has also been extended to successfully account for the word frequency, bizarreness, enactment, and perceptual interference effects (McDaniel & Bugg, 2008), proposes that variations in recall performance across experimental designs may be accounted for by a processing trade-off that occurs when an unusual encoding condition enhances processing of individual stimulus features at the expense of encoding temporal order information about the presentation of the events.

Within the present context, we considered effort-after-meaning conditions to represent unusual encoding. Therefore, the item-order account predicts that retention of order information in the within-subjects design should be equivalent across all sentence presentation conditions. This prediction arises because increased effort after meaning should disrupt the encoding of order information throughout the list of sentences, whereas memory for the contents of the individual sentences should be enhanced in the conditions that elicit greater effort after meaning (the delayed-cue condition). By contrast, in the between-subjects design, greater processing of sentences in the delayed-cue list condition should impair encoding of temporal order information, relative to the embedded-cue list condition. Enhanced order memory in the embedded-cue condition should facilitate sentence recall and significantly reduce, if not eliminate, the memorial advantage of effort after meaning.

On the other hand, if the multiple-cue account is correct, recall of sentences in the delayed-cue condition should be enhanced in both types of experimental designs, without necessarily impairing temporal order memory. In other words, one's effort after meaning for ambiguous stimuli at study should generate alternative cues that may be used to facilitate later retrieval of the correct sentence representation across experimental designs, just as incongruous item generation effects have been demonstrated with generate and read conditions distributed between subjects (e.g., Soraci et al., 1994).

## Method

**Subjects**. Ninety Washington University undergraduates participated for either course credit or payment.

**Materials and Procedure**. The stimuli were 32 sentences, and their corresponding cues were sampled from Auble et al. (1979; see Appendix C). On the basis of ratings obtained in the original Auble et al. study, 13 sentences were difficult to comprehend, in that the majority of subjects reported not understanding their meaning independently of the cues (Sentences 1–13). Three additional sentences (14–16) created in our laboratory were included as difficult sentences.[3]

There were two sentence–cue presentation conditions during the first study trial. For 30 subjects, eight of the sentences had embedded cues (embedded-cue condition), and the remaining eight sentences were followed by cues after a 10-sec interval (delayed-cue 10-sec condition). Half of the sentences in each condition were easy to comprehend, and the other half difficult. Sentence order was randomized, with the constraint that, across subjects, each sentence was presented an equal number of times in each of the two conditions.

For the remaining 60 subjects, the two cue presentation conditions were distributed between subjects, with 30 subjects assigned to each of two groups. The subjects in the embedded-cue group heard only sentences with cues meaningfully embedded. The subjects in the delayed-cue 10-sec group heard only sentences followed by cues after a 10-sec delay interval. Half of the sentences for each group were easy to comprehend, and the other half difficult. One concern with varying the conditions of cue presentation between subjects was that in the delayed-cue condition, knowing that they would always hear a cue for each and every sentence, the subjects might exert little effort to attempt to understand the sentences. To address this concern, all the subjects were instructed to try to discern the meanings of the sentences. The total study time for each sentence was 20 sec, which included 5 sec for sentence presentation and 2 sec for cue presentation (see Appendix B).

The experimenter initially instructed the subjects that they would study and later be asked to recall a series of sentences in two separate study and test trials. The subjects in the mixed-list and delayed-cue groups were informed that additional cue words might be heard, to facilitate sentence comprehension. Following auditory presentation of each sentence and the possible cue on the first trial, the subjects were prompted by a computer display and were allowed 10 sec to indicate how well they understood the sentence by typing a number between 0 and 100. After hearing 16 sentences, the subjects attempted to type as many of the sentences as they could remember in any order for 10 min. The experimenter instructed the subjects to mark the passing of each minute by typing the current interval following the most recent response.

The same study phase procedure was then repeated on the second trial, using the remaining 16 sentences. This time, auditory presentation of each sentence and the possible cue was followed by a surprise order reconstruction test. The subjects were given a sheet listing all 16 sentences in random order and were asked to number them 1–16 in the exact order of presentation for 5 min. The experimental session lasted about 1 h.

## Results

**Comprehension**. As is shown in Table 1, both the easy and difficult sentences were generally understood across all subject groups following the presentation of either embedded or delayed cues. Nevertheless, the easy sentences were still perceived as more comprehensible than the difficult sentences. For the within-subjects group, an ANOVA indicated that the easy sentences received higher comprehension ratings than did the difficult sentences [$F(1,29) = 65.91$, $MS_e = 60.69$, $\eta_p^2 = .69$]. However, there was neither a significant effect of cue presentation condition [$F(1,29) = 1.03$, $MS_e = 107.74$, n.s., $\eta_p^2 = .03$] nor an interaction between the two factors ($F < 1$). Similar results were obtained for the between-subjects groups, in that the easy sentences received higher comprehen-

**Table 1**
**Mean Judgments of Comprehension for the Easy- and Difficult-to-Comprehend Sentences Studied in Embedded-Cue and Delayed-Cue Presentation Conditions in Trial 1 of Experiment 2**

| | Within Subjects | | | | Between Subjects | | | |
| | Easy | | Difficult | | Easy | | Difficult | |
| Condition | M | CI | M | CI | M | CI | M | CI |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Embedded cue | 91.4 | 2.0 | 79.6 | 4.2 | 92.0 | 3.1 | 79.8 | 5.2 |
| Delayed cue 10 sec | 93.2 | 2.0 | 81.9 | 2.1 | 92.7 | 3.1 | 76.3 | 7.4 |

Note—Confidence interval (CI) values for the within-subjects portion are 95% CIs calculated according to the method of Loftus and Masson (1994). CI values for the between-subjects portion are standard 95% CIs.

sion ratings than did the difficult sentences [$F(1,58) = 49.88$, $MS_e = 121.96$, $\eta_p^2 = .46$]. In addition, there was neither a significant effect of cue presentation condition ($F < 1$) nor a significant interaction between the two factors [$F(1,58) = 1.07$, $MS_e = 121.96$, n.s., $\eta_p^2 = .02$]. Thus, judgments of comprehension were influenced primarily by the linguistic features and/or semantic contents of the individual sentences, and not by variations in the conditions of cue presentation (e.g., effort after meaning).

**Sentence recall**. Sentence recall performance for the within-subjects group was higher in the delayed-cue 10-sec condition ($M = .57$, $SD = .21$) than in the embedded-cue condition ($M = .38$, $SD = .23$) [$t(29) = 4.85$, $SEM = .04$]. An ANOVA further compared recall for both easy and difficult to comprehend sentences across conditions, and the results showed a significant effect of cue presentation condition [$F(1,29) = 23.50$, $MS_e = .05$, $\eta_p^2 = .45$]. However, there was neither a significant effect of sentence comprehension difficulty nor an interaction between the two factors ($Fs < 1$). By contrast, for the between-subjects groups, recall performance was similar for the embedded-cue ($M = .47$, $SD = .15$) and delayed-cue ($M = .49$, $SD = .13$) conditions. An ANOVA indicated that there was neither a significant effect of cue presentation condition ($F < 1$) or sentence difficulty ($F < 1$) nor a significant interaction between the two factors [$F(1,58) = 2.67$, $MS_e = .03$, n.s., $\eta_p^2 = .04$].

In order to directly compare within- and between-subjects effects of cue presentation condition on sentence recall, Erlebacher's (1977, 1978) ANOVA procedure was used.[4] The ANOVA revealed that although there was a nonsignificant effect of experimental design ($F < 1$), there was a main effect of cue presentation condition [$F(1,72) = 15.92$, $MS_e = .02$, $\eta_p^2 = .27$] and a significant interaction between experimental design and cue presentation condition [$F(1,72) = 10.17$, $MS_e = .02$, $\eta_p^2 = .19$].

**Order reconstruction**. Table 2 reports order reconstruction performance, as measured by the proportion of sentences correctly identified in their study serial position. As is predicted by the item-order theoretical framework, order reconstruction was best for the embedded-cue group ($M = .45$, $SD = .19$), followed by the group who studied a mixed list of sentences presented with embedded or delayed cues ($M = .41$, $SD = .23$), and was poorest for the delayed-cue (10-sec) group ($M = .36$, $SD = .16$). Importantly, performance for the embedded-cue group was significantly better than that for the delayed-cue group

**Table 2**
**Input–Output (I/O) Correspondence Scores Calculated According to the Method of Asch and Ebenholtz (1962) and the Proportions (Prop.) of Sentences Correctly Identified in Their Serial Position During List Presentation in the Mixed-List, Embedded-Cue, and Delayed-Cue List Presentation Conditions in Experiment 2**

| | I/O Correspondence | | Order Reconstruction | |
|---|---|---|---|---|
| Condition | M | CI | Prop. | CI |
| Mixed list | .51 | .05 | .41 | .08 |
| Embedded cue | .59 | .07 | .45 | .07 |
| Delayed cue 10 sec | .49 | .06 | .36 | .06 |

Note—CI, standard 95% confidence interval.

$[t(58) = 2.08, SEM = .05, d = 0.51]$. Order reconstruction scores did not differ significantly between the mixed-list and embedded-cue groups $[t(58) = 0.80, SEM = .05,$ n.s., $d = 0.19]$ or between the mixed-list and delayed-cue groups $[t(58) = 1.01, SEM = .05,$ n.s., $d = 0.25]$.

In addition, Figure 1 plots order reconstruction performance for the embedded-cue and delayed-cue groups as a function of serial position. The serial position curves showed marked primacy and recency effects, and order reconstruction was better in the embedded-cue condition for 12 out of the 16 serial positions. An ANOVA revealed significant effects of serial position $[F(15,870) = 28.29, MS_e = .15, \eta_p^2 = .33]$ and cue presentation condition $[F(1,58) = 4.34, MS_e = .51, \eta_p^2 = .07]$, but a nonsignificant interaction between the two factors $[F(15,870) = 1.20, MS_e = .15,$ n.s., $\eta_p^2 = .02]$.

**Input–output correspondence**. To provide a more direct test of the item-order account, Table 2 also reports input–output correspondence, the degree to which subjects
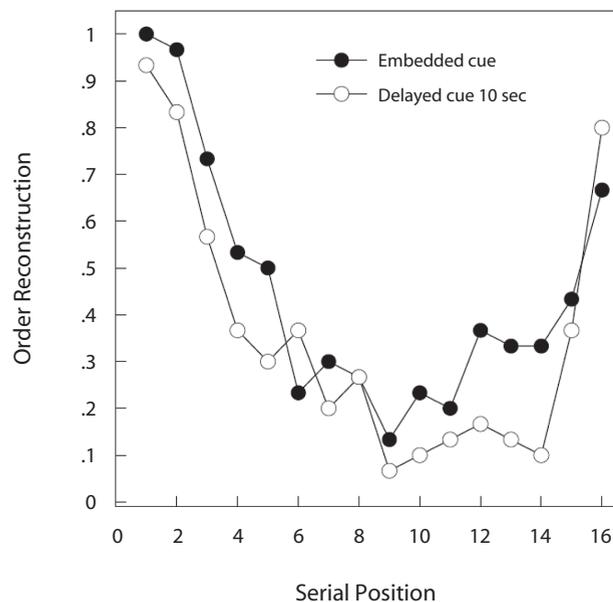


**Figure 1. Order reconstruction performance plotted as the proportion of sentences correctly assigned to each serial position for the embedded- and delayed-cue conditions distributed between subjects in Experiment 2.**

recall sentences in their order of study presentation, calculated according to the method of Asch and Ebenholtz (1962). Again, as predicted by the item-order theoretical framework, input–output correspondence was highest for the embedded-cue group $(M = .59, SD = .19)$, followed by the group who studied a mixed list of sentences presented with embedded or delayed cues $(M = .51, SD = .14)$, and was poorest for the delayed-cue 10-sec group $(M = .49, SD = .15)$. Importantly, input–output correspondence for the embedded-cue group was significantly greater than that for the delayed-cue group $[t(58) = 2.23, SEM = .04, d = 0.58]$. Scores did not differ significantly between the mixed-list and embedded-cue groups $[t(58) = 1.74, SEM = .04,$ n.s., $d = 0.48]$ or between the mixed-list and delayed-cue groups $[t(58) = 0.63, SEM = .04,$ n.s., $d = 0.14]$.

**Discussion**

The results of Experiment 2 replicated the finding that when conditions of cue presentation are distributed within subjects, requiring subjects to puzzle over the meaning of sentences for several seconds prior to comprehension enhances recall, relative to an embedded-cue condition in which the sentences are readily understood. These data are consistent with the results of Experiment 1 and prior studies and corroborate the notion of a memorial advantage for effort after meaning (Auble & Franks, 1978; Auble et al., 1979). By contrast, when the cue presentation conditions in the present experiment were distributed between subjects, recall was similar for sentences studied with embedded and delayed cues. Moreover, temporal order memory, as measured by performance on the order reconstruction task, as well as the degree of input–output correspondence in free recall, was superior in the embedded-cue condition, relative to the delayed-cue condition.

Somewhat surprisingly, sentences that were considered relatively easy to understand were recalled no better than difficult sentences, yet both types of sentences were remembered better in the delayed-cue condition. This finding is consistent with data reported by Auble et al. (1979) and suggests that the resolution of ambiguity, regardless of degree, is important in determining the memorial advantage of effort after meaning.

Taken together, these findings are similar to those observed in studies of the generation effect, with regard to how experimental design affects the phenomenon. More important, the results directly follow from predictions of the item-order theoretical framework, which maintains that variations in recall performance across experimental designs may be accounted for by a processing trade-off that occurs when an unusual encoding condition, such as increased effort after meaning, enhances processing of individual stimulus features at the expense of encoding temporal order information. In the within-subjects design, the item-order account correctly predicted superior recall for sentences in the delayed-cue condition, because order memory should be equivalent across the two sentence presentation conditions and sentences in the delayed-cue condition presumably benefited from enhanced sentence-specific processing, relative to the embedded-cue sentences. By contrast, in the between-subjects design, order memory was enhanced for

sentences with embedded cues that required relatively little effort after meaning at study, as compared with the delayed-cue condition, thereby compensating for the difference in recall that was observed in the within-subjects design.

It should be noted that Experiments 1 and 2 demonstrated the effects of effort after meaning on recall only under conditions in which the sentences were generally understood during study. Yet, if the resolution of ambiguity is critical in determining the effects of effort after meaning on recall, it is important to ask whether the above-mentioned effects would occur under conditions in which the sentences might or might not remain ambiguous.

## EXPERIMENT 3

In Experiment 3, recall for ambiguous sentences studied with disambiguating cues was compared with recall for sentences studied without cues. In one condition, subjects studied ambiguous sentences without the benefit of cues to aid comprehension. In a second condition, subjects were provided with cues to facilitate comprehension just prior to hearing the sentences themselves. In the last two conditions, cues were provided to the subjects either 2 or 6 sec following presentation of their corresponding sentences.[5] On the basis of the previous findings that suggest that the resolution of ambiguity is critical in determining the effects of effort after meaning on recall performance, recall for sentences studied with cues to aid comprehension should exceed that of sentences studied without cues, even though subjects might exert a considerable amount of mental effort attempting to understand ambiguous sentences in the latter condition.

It is established that meaningful stimuli tend to be remembered better than poorly understood or nonsensical materials, and this memorial advantage has been observed in between-subjects experimental designs (e.g., Bransford & Johnson, 1972; Johnson, Doll, Bransford, & Lapinski, 1974). Thus, in Experiment 3, we examined again whether the memorial advantage of effort after meaning would extend to a between-subjects design, this time by comparing cue presentation conditions that facilitated comprehension with a no-cue condition in which sentences remained ambiguous. For one group of subjects, the four cue presentation conditions varied within subjects, and for the remaining four groups, the conditions varied between subjects.

### Method

**Subjects**. One hundred twenty-eight Washington University undergraduates participated for either course credit or payment.

**Materials and Procedure**. The stimuli were the same sentences and their corresponding cues that were used in Experiment 1 (see Appendix A). There were four conditions of sentence–cue presentation during the study phase. For the within-subjects group ($n = 64$), six of the sentences were presented without cues (no-cue condition). Six of the sentences were immediately preceded by cues (precue condition). In the last two conditions, sentences were followed by cues at 2- or 6-sec intervals (delayed-cue 2-sec and 6-sec conditions). Sentence order was randomized, with the constraint that, across subjects, each sentence was presented an approximately equal number of times in each of the four conditions. For the remaining subjects, the cue presentation conditions were distributed between subjects, with 16 subjects assigned to one of four groups.

The subjects in the no-cue group studied only sentences presented without cues that would facilitate comprehension. The precue group studied sentences that were immediately preceded by cues. And the two delayed-cue groups heard sentences followed by cues at 2- and 6-sec delay intervals, respectively.[6] The total presentation time for each sentence was 15 sec, which included 5 sec for sentence and 2 sec for cue presentation (see Appendix B). Study instructions, judgments of comprehension, number-matching exercise, and recall test were performed in the manner described in Experiment 1. The experimental session lasted about 45 min.

### Results

**Comprehension**. Sentence comprehension was measured as the proportion of sentences rated as being understood in each of the sentence presentation conditions and subject groups. For the within-subjects group, the rate of perceived comprehension was highest for sentences in the delayed-cue 6-sec condition ($M = .87$, $SD = .14$), followed by the precue ($M = .82$, $SD = .16$) and delayed-cue 2-sec ($M = .76$, $SD = .20$) conditions, and was lowest for sentences studied without cues ($M = .21$, $SD = .24$). An ANOVA compared sentence comprehension ratings, and the results showed a significant effect of cue presentation condition [$F(3,189) = 191.84$, $MS_e = .03$, $\eta_p^2 = .75$]. Individual pairwise comparisons revealed that perceived comprehension for sentences in all three cue conditions was significantly higher than that in the no-cue condition [$t(63) = 17.20$, $SEM = .04$, $d = 2.99$; $t(63) = 14.96$, $SEM = .04$, $d = 2.49$; and $t(63) = 20.86$, $SEM = .03$, $d = 3.36$, for comparisons of the precue, delayed-cue 2-sec, and delayed-cue 6-sec conditions with the no-cue condition, respectively]. Comprehension did not differ significantly between the precue and delayed-cue 2-sec conditions [$t(63) = 1.72$, $SEM = .03$, n.s., $d = 0.33$]. However, sentences in the delayed-cue 6-sec condition were more likely to be rated as being understood than were sentences in the delayed-cue 2-sec condition [$t(63) = 3.77$, $SEM = .03$, $d = 0.64$], as well as those in the precue condition, although the latter difference did not reach the conventional level of statistical significance [$t(63) = 2.39$, $SEM = .02$, $p = .06$, $d = 0.33$].

For the between-subjects groups, the rate of perceived comprehension was highest for sentences in the precue condition ($M = .87$, $SD = .13$), followed by the delayed-cue 6-sec ($M = .79$, $SD = .15$) and delayed-cue 2-sec ($M = .75$, $SD = .25$) conditions, and was lowest for sentences studied without cues ($M = .32$, $SD = .22$). An ANOVA also indicated a significant effect of cue presentation condition on comprehension ratings [$F(3,60) = 25.67$, $MS_e = .04$, $\eta_p^2 = .56$]. Individual comparisons revealed that comprehension for sentences in all of the cue conditions was significantly greater than that in the no-cue condition (all $t$ values $> 4.00$, all $d$ values $> 1.46$). No other differences were statistically significant.

**Sentence recall**. Table 3 reports sentence recall performance for both the within- and between-subjects groups. Sentence recall for the within-subjects group was highest in the delayed-cue 6-sec condition ($M = .43$, $SD = .21$), followed by the precue ($M = .36$, $SD = .23$) and delayed-cue 2-sec ($M = .33$, $SD = .22$) conditions. Recall was poorest for sentences studied without cues ($M = .18$,

**Table 3**
**Proportions (Prop.) of Sentences Correctly Recalled in No-Cue, Precue, and Delayed-Cue Presentation Conditions in Experiment 3**

| Condition | Within Subjects | | Between Subjects | |
|---|---|---|---|---|
| | Prop. | CI | Prop. | CI |
| No cue | .18 | .03 | .32 | .06 |
| Precue | .36 | .06 | .40 | .06 |
| Delayed cue 2 sec | .33 | .06 | .39 | .07 |
| Delayed cue 6 sec | .43 | .05 | .27 | .07 |

Note—Confidence interval (CI) values for the within-subjects portion are 95% CIs calculated according to the method of Loftus and Masson (1994). CI values for the between-subjects portion are standard 95% CIs.

$SD = .15$). An ANOVA compared sentence recall across conditions, and the results showed a significant effect of cue presentation condition [$F(3,189) = 20.01$, $MS_e = .03$, $\eta_p^2 = .24$]. Individual pairwise comparisons revealed that, as compared with the no-cue condition, recall was significantly higher for sentences in the precue [$t(63) = 5.97$, $SEM = .03$, $d = 0.93$], delayed-cue 2-sec [$t(63) = 4.60$, $SEM = .03$, $d = 0.80$], and delayed-cue 6-sec [$t(63) = 8.05$, $SEM = .03$, $d = 1.37$] conditions. Furthermore, recall in the delayed-cue 6-sec condition was higher than that in the delayed-cue 2-sec condition [$t(63) = 3.26$, $SEM = .03$, $d = 0.46$], as well as that in the precue condition, although the last difference did not reach the conventional level of statistical significance [$t(63) = 2.03$, $SEM = .03$, $p = .14$, $d = 0.32$]. Last, the difference in recall between the delayed-cue 2-sec and precue conditions was not significant [$t(63) = 0.73$, $SEM = .04$, n.s., $d = 0.13$].

By contrast, when the cue presentation conditions were distributed between subjects, recall performance was highest in the precue condition ($M = .40$, $SD = .11$), followed by the delayed-cue 2-sec ($M = .39$, $SD = .13$) and no-cue ($M = .32$, $SD = .12$) conditions. Surprisingly, recall was poorest in the delayed-cue 6-sec condition ($M = .27$, $SD = .14$). An ANOVA indicated a significant, albeit small, effect of cue presentation condition on sentence recall [$F(3,60) = 3.79$, $MS_e = .02$, $\eta_p^2 = .16$]. However, individual pairwise comparisons revealed significant differences in recall performance only between the precue and delayed-cue 6-sec conditions [$t(30) = 2.96$, $SEM = .04$, $d = 1.05$] and between the two delayed-cue conditions [$t(30) = 2.49$, $SEM = .05$, $d = 0.89$]. No other comparisons were statistically significant.

In order to directly compare within- and between-subjects effects of cue presentation condition on sentence recall, Erlebacher's (1977, 1978) ANOVA procedure was used. Because this procedure requires an equal number of subjects for the within-subjects portion of the experiment and for each group in the between-subjects portion, data from only the first 16 subjects tested in the within-subjects portion of the experiment were included in the analysis. The ANOVA revealed that although there was no main effect of experimental design ($F < 1$), there was a main effect of cue presentation condition [$F(3,80) = 6.55$, $MS_e = .03$, $\eta_p^2 = .28$]. This effect was qualified, however, by a significant interaction between experimental design and cue presentation condition [$F(3,80) = 8.14$, $MS_e = .03$, $\eta_p^2 = $

.33], which was due in large part to the no-cue condition's yielding poor recall performance in the within-subjects group, but not among the between-subjects groups. Conversely, the delayed-cue 6-sec condition provided greatest recall in the within-subjects group, but not among the between-subjects groups, where recall was even lower.

## Discussion

The results of Experiment 3 demonstrated that when the four conditions of cue presentation (no cue, precue, delayed cue 2 sec, and delayed cue 6 sec) were distributed within subjects, providing cues to facilitate comprehension improved recall for ambiguous sentences, even though sentences studied without cues required the subjects to exert a considerable amount of effort after meaning. These data are consistent with the results of prior studies and corroborate the notion that eventual comprehension of the to-be-learned stimuli is a critical factor in determining the memorial advantage for effort after meaning (Auble & Franks, 1978; Auble et al., 1979).

Somewhat surprisingly, with the conditions distributed between subjects, neither providing a cue nor increasing the delay interval between sentence and cue presentation improved recall performance. Contrary to the established finding that comprehensible or meaningful verbal materials are remembered better than poorly understood or nonsensical materials (e.g., Bransford & Johnson, 1972), these data demonstrate that under certain conditions, comprehension does not aid recall. Even more surprisingly, the delayed-cue 6-sec condition, which elicited the best recall performance in the within-subjects design, demonstrated the poorest recall performance in the between-subjects design. Why the delayed-cue 6-sec group exhibited such a low level of recall, relative to the other groups and, especially, the delayed-cue 2-sec group, is unclear.

## EXPERIMENT 4

In Experiment 4, we examined the memorial advantage of effort after meaning by asking whether sentence recall is associated with the ability to remember the corresponding disambiguating cues. To the extent that the resolution of ambiguity is critical in determining the effects of effort after meaning on recall, effort after meaning should promote the integration of an ambiguous sentence with its corresponding cue during study to facilitate comprehension and sentence recall. One prediction of this view is that one should observe a direct relation between sentence and cue recollection.

Experiment 4 tested the effects of effort after meaning on recall in both within- and between-subjects experimental designs by focusing on subjects' recall of the sentence cues, and not the sentences themselves. In addition, the subjects attempted to recall the sentences (as well as the cues) following recall of the cues alone, in order to examine memory for the sentences following successful and failed retrieval of cues. If sentence recall is directly related to cue recollection, one would first expect to observe a similar pattern of cue and sentence recall performance in the three cue presentation conditions when manipulated both within and between subjects. Moreover, one would expect to observe a tendency for

the subjects to recall the cues associated with sentences that are correctly recalled and, conversely, a tendency to recall the sentences that correspond to correctly recalled cues.

## Method

**Subjects**. Ninety-six Washington University undergraduates participated for either course credit or payment.

**Materials and Procedure**. The stimuli and procedures were similar to those used in Experiment 3, with the following exceptions. First, there were three sentence–cue presentation conditions during the study phase. For half of the subjects ($n = 48$), eight of the sentences were immediately preceded by cues (precue condition). Eight of the sentences were followed by cues after a 2-sec interval (delayed-cue 2-sec condition). Last, eight sentences were followed by cues after a 10-sec interval (delayed-cue 10-sec condition). The order of sentence and cue presentation conditions was randomized, with the constraint that, across subjects, each sentence was presented an approximately equal number of times in each of the three conditions.

For the remaining 48 subjects, the three cue presentation conditions were distributed between subjects, with 16 subjects assigned to each of three groups. The subjects in the precue group studied only sentences that were immediately preceded by cues. The subjects in the delayed-cue 2-sec group heard only sentences followed by cues after a 2-sec delay interval. And the subjects in the delayed-cue 10-sec group heard only sentences followed by cues after a 10-sec delay interval. The total presentation time for each sentence was 20 sec, which included 5 sec for sentence and 2 sec for cue presentation (see Appendix B).

The experimenter initially instructed the subjects that they would study and later be asked to recall a series of sentences. The subjects were further informed that additional cue words would be heard to facilitate sentence comprehension. Following auditory presentation of the sentences and possible cues and completion of the number-matching exercise, the experimenter instructed the subjects to type as many of the cues as they could remember into a blank Microsoft Word document for 5 min. When the 5 min elapsed, the subjects closed the Word document and opened a new blank document for sentence recall. Last, the subjects attempted to type as many and as much of the sentences (along with the cues) as they could remember for 10 min. The experimenter instructed the subjects to mark the passing of each minute during both recall periods by typing the current interval into the document following the most recent response. The experimental session lasted about 1 h.

## Results

**Comprehension**. Sentence comprehension was measured as the proportion of sentences rated as being understood. When the conditions of cue presentation varied within subjects, the rate of perceived comprehension was highest in the precue condition ($M = .83$, $SD = .16$), followed by the delayed-cue 10-sec ($M = .82$, $SD = .19$) and 2-sec ($M = .78$, $SD = .18$) conditions. An ANOVA indi-

cated that the effect of cue presentation condition on comprehension ratings did not reach the conventional level of statistical significance [$F(2,94) = 2.48$, $MS_e = .02$, $p = .09$, $\eta_p^2 = .05$]. Similarly, when the conditions of cue presentation were distributed between subjects, the rate of perceived comprehension was highest for sentences in the precue condition ($M = .82$, $SD = .11$), followed by the delayed-cue 2-sec ($M = .72$, $SD = .20$) and 10-sec ($M = .69$, $SD = .27$) conditions. An ANOVA confirmed that the comprehension ratings did not significantly differ across the cue presentation conditions [$F(2,45) = 1.71$, $MS_e = .04$, n.s., $\eta_p^2 = .07$].

**Cue recall**. Table 4 reports cue recall performance, as measured by the proportion of sentence cues correctly recalled during the initial recall period (Test 1). When the cue presentation conditions were manipulated within subjects, cue recall was highest for sentences in the delayed-cue 10-sec condition, followed by sentences in the precue and delayed-cue 2-sec conditions. However, the results of an ANOVA contrasting the effects of cue presentation condition on cue recall performance were not significant [$F(2,94) = 1.97$, $MS_e = .03$, n.s., $\eta_p^2 = .04$]. When the cue presentation conditions were distributed between subjects, cue recall performance was similar across conditions ($F < 1$).

Since the subjects were instructed to recall cues during the second recall period (Test 2), Table 4 also reports cumulative cue recall performance across the two recall periods (Tests 1 and 2). Thus, when the conditions of cue presentation varied within subjects, an ANOVA indicated a significant effect of cue presentation condition on cumulative cue recall performance [$F(2,94) = 6.57$, $MS_e = .02$, $\eta_p^2 = .12$]. Individual pairwise comparisons revealed that recall of the cues presented in the delayed-cue 10-sec condition was superior to that of cues in the precue condition [$t(47) = 3.56$, $SEM = .03$, $d = 0.54$], as well as in the delayed-cue 2-sec condition [$t(47) = 3.05$, $SEM = .03$, $d = 0.63$]. There was no difference in cue recall for the delayed-cue 2-sec and precue conditions [$t(47) = 0.15$, $SEM = .03$, n.s., $d = 0.05$]. By contrast, with the conditions of cue presentation distributed between subjects, cue recall performance was similar across all three conditions ($F < 1$).

**Sentence recall**. When the conditions of cue presentation varied within subjects, sentence recall was highest for the delayed-cue 10-sec condition ($M = .51$, $SD = .15$), followed by sentences in the delayed-cue 2-sec ($M = .40$,

---

**Table 4**
**Proportions (Prop.) of Cues Correctly Recalled During
the First and Second Recall Trials in Precue and
Delayed-Cue Presentation Conditions in Experiment 4**

| | Within Subjects | | | | Between Subjects | | | |
| | Test 1 | | Tests 1 and 2 | | Test 1 | | Tests 1 and 2 | |
| Condition | Prop. | CI | Prop. | CI | Prop. | CI | Prop. | CI |
|---|---|---|---|---|---|---|---|---|
| Precue | .33 | .06 | .38 | .07 | .31 | .06 | .39 | .07 |
| Delayed cue 2 sec | .32 | .05 | .37 | .06 | .28 | .05 | .35 | .07 |
| Delayed cue 10 sec | .38 | .04 | .48 | .05 | .32 | .05 | .40 | .07 |

Note—Confidence interval (CI) values for the within-subjects portion are 95% CIs calculated according to the method of Loftus and Masson (1994). CI values for the between-subjects portion are standard 95% CIs.

$SD = .20$) and precue ($M = .38$, $SD = .21$) conditions. An ANOVA indicated a significant effect of cue presentation condition [$F(2,94) = 8.05$, $MS_e = .03$, $\eta_p^2 = .15$]. Individual pairwise comparisons revealed that recall of sentences presented in the delayed-cue 10-sec condition was superior to that of sentences in the precue condition [$t(47) = 3.99$, $SEM = .03$, $d = 0.71$], as well as the delayed-cue 2-sec condition [$t(47) = 3.69$, $SEM = .03$, $d = 0.62$]. There was no difference in sentence recall for the delayed-cue 2-sec and precue conditions [$t(47) = 0.42$, $SEM = .04$, n.s., $d = 0.10$]. By contrast, with the conditions of cue presentation distributed between subjects, sentence recall performance was similar across all three conditions ($M = .39$, $SD = .13$, in the delayed-cue 10-sec condition; $M = .34$, $SD = .14$, in the delayed-cue 2-sec condition; and $M = .36$, $SD = .15$, in the precue condition; $F < 1$).

The fact that the subjects were given an additional opportunity to recall cue words during the sentence recall phase made it possible to measure the conditional probability of recalling sentence cues during either the first or the second recall period, given the correct recall of their corresponding sentence. As is shown in Table 5, the subjects were able to recall appropriate cues for over 95% of the sentences they correctly recalled. Similarly, the subjects recalled the corresponding sentences for about 90% of the cues that were correctly recalled. Seldom did a subject recall a sentence (or cue) without being able to recollect its corresponding cue (or sentence).

## Discussion

The results of Experiment 4 demonstrated that the pattern of recall performance for both cues and sentences was similar in both within- and between-subjects experimental designs, even though the cues were studied only under incidental learning instructions. For subjects tested in a within-subjects design, cue and sentence recall was superior for the condition in which the subjects puzzled over the meaning of the sentences for 10 sec before hearing the disambiguating cues. This is somewhat surprising, in that the previous experiments failed to demonstrate a significant recall advantage for studying an ambiguous sentence with a delayed cue, as opposed to hearing the cue before the sentence. The main difference between this and the previous experiments is that the subjects initially focused their retrieval on the cues alone, which, in turn, may have facilitated sentence recall,

thereby enabling the subjects to more fully capitalize upon the encoding advantage of effort after meaning. Why this advantage should occur only in the longer delayed-cue condition and not in the precue and shorter delayed-cue conditions is unclear. By contrast, with a shift to a between-subjects design, cue and sentence recall was similar across the cue presentation conditions. In addition, the finding that the subjects recollected the appropriate cues for nearly every sentence they correctly recalled, and vice versa, suggests a direct relationship between sentence and cue memory.

## CUMULATIVE RECALL DATA

As was described in the Method section for Experiment 1, in all our experiments, recall was measured over time, so that cumulative recall curves could be produced. This procedure is advantageous, because one can ascertain whether the patterns of effects seen at the end of the recall period (i.e., the measure usually reported) hold over time. Psychologists have generally ignored the importance of recall time as a variable, despite evidence that it can greatly matter in the outcome of experiments (e.g., Roediger & Thorpe, 1978).

The cumulative recall data for all four experiments are provided in Figure 2. The figure serves as a useful summary of the main points of our article and also shows that, in these effort-after-meaning experiments, conclusions generally hold irrespective of recall time. That is, the trends that we have shown to be statistically significant generally hold from the first minute on. Figure 2 (panels A–C) shows that, relative to the embedded-cue and no-cue control conditions, all the conditions in which the cue and sentence were separated (and hence, effort after meaning was required to achieve comprehension) showed greater recall in the within-subjects design typically used in prior research. Furthermore, the effects consistently disappear in between-subjects designs, irrespective of the amount of recall time.

## GENERAL DISCUSSION

In four experiments, puzzling over the meaning of to-be-learned ambiguous sentences prior to comprehension enhanced recall, relative to studying sentences that were more readily understood or not understood, a result that is consistent with the original findings of a memorial advantage of effort toward comprehension, otherwise termed *effort after meaning* or the *"aha!" effect* (Auble & Franks, 1978; Auble et al., 1979). However, this memorial advantage was observed only when the conditions of cue presentation were distributed within subjects, in a mixed-list design. When the conditions of cue presentation were distributed between subjects, the effect was not observed. In fact, as was reported in Experiment 3, the sentences that were recalled the best in the within-subjects design—the sentences studied with the longest delayed-cue interval—were recalled worse than sentences studied without cues in the between-subjects design.

The present findings help to elucidate the theoretical link between effort after meaning and generation. Previous studies have suggested similarities between the processes involved in effort after meaning and two different

**Table 5**
**Conditional Probability of Recalling Cues Given the Correct Recall of Their Corresponding Sentences, and Vice Versa, in Precue and Delayed-Cue Presentation Conditions in Experiment 4**

| | Within Subjects | | | | Between Subjects | | | |
|---|---|---|---|---|---|---|---|---|
| | $P(C \mid S)$ | | $P(S \mid C)$ | | $P(C \mid S)$ | | $P(S \mid C)$ | |
| Condition | M | CI | M | CI | M | CI | M | CI |
| Precue | .95 | .07 | .92 | .07 | .97 | .03 | .89 | .05 |
| Delayed cue 2 sec | .95 | .06 | .93 | .04 | .97 | .03 | .89 | .06 |
| Delayed cue 10 sec | .97 | .04 | .96 | .04 | .96 | .04 | .91 | .04 |

Note—Confidence interval (CI) values for the within-subjects portion are 95% CIs calculated according to the method of Loftus and Masson (1994). Values for the between-subjects portion are standard 95% CIs.
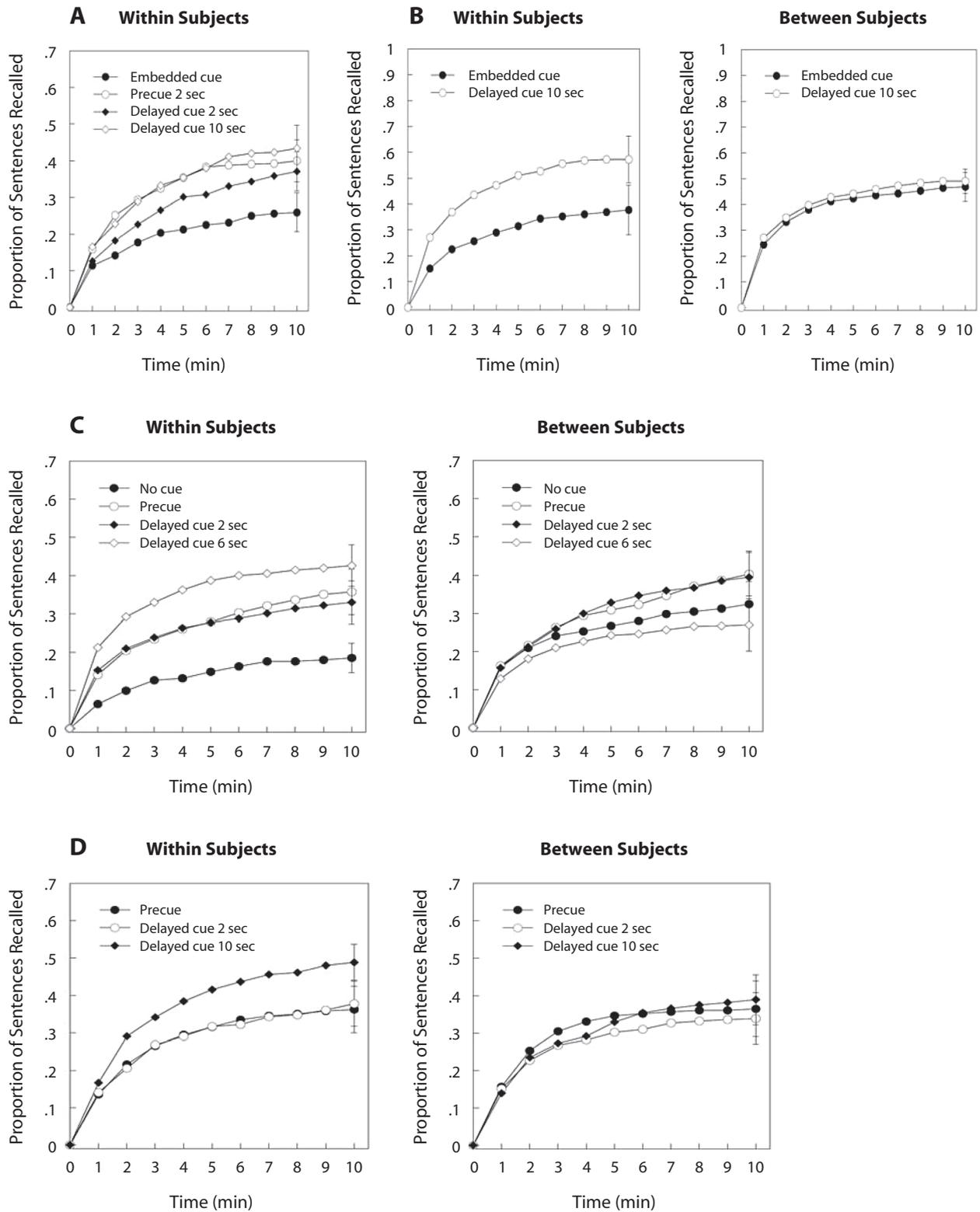
Figure 2. (A–D) Cumulative recall performance in Experiments 1–4, respectively. Error bars for the final recall value in the within-subjects experiments are 95% confidence intervals calculated according to the method of Loftus and Masson (1994). Error bars for the final recall value in the between-subjects experiments are standard 95% confidence intervals.

types of generation. Consistent with the view of Auble et al. (1979), effort after meaning and generation may be similar to the extent that both involve a mental transition from a state of ambiguity or noncomprehension to sudden comprehension. By contrast, Wills et al. (2000) proposed that effort after meaning is comparable to incongruous item generation, in that both involve the generation of multiple candidate responses during study that may serve as retrieval cues to aid recall. The present findings violate predictions derived from the multiple-cue account.

First, in Experiment 1, there was a recall advantage for sentences studied with precues, as compared with sentences with embedded cues, and recall performance in the precue condition was comparable to that in the delayed-cue conditions. If anything, the precue condition is designed to constrain the generation of multiple sentence meanings during study. Second, in Experiment 2, sentences that were considered relatively easy to understand were recalled no better than difficult sentences, yet both types of sentences were remembered better in the delayed-cue condition. This finding is consistent with data reported by Auble et al. (1979) and suggests that the resolution of ambiguity, regardless of degree, is critical in determining the memorial advantage of effort after meaning and, perhaps, other generation effects as well. For instance, in Experiment 1 in Jacoby (1978), subjects studied and later attempted to recall a list of word pair associates under two conditions, reading and construction. The subjects either read the associates ("foot–shoe") or constructed the associates by completing a word fragment ("foot–s__e"). During study, the word pairs were presented either once or twice within a list, and the repetitions occurred in massed and spaced intervals. Furthermore, study instructions varied for the repeated items, so that subjects having first read a given word pair might have to either read or construct the pair on the second presentation. Jacoby found that very minimal construction greatly enhanced cued recall of word pair associates, even for once-presented items. Somewhat surprisingly, subjects who constructed a word fragment immediately after having read that same word showed no greater memory for the generated words than for comparable read items. Jacoby argued that for these massed items, generation was a "trivial" task. Alternatively, one might argue that foreknowledge of the word fragments might have suppressed an "aha!" reaction.

Last, in Experiments 2–4, the memorial benefit of effort after meaning did not occur in between-subjects designs. Indeed, all of these results are contrary to those observed with incongruous item generation (Soraci et al., 1994). This does not completely rule out the multiple-cue account, however. It is possible that in group learning contexts and under study conditions that allow more time for individuals to actively contemplate the meaning of ambiguous materials, multiple candidate solutions are generated and aid in subsequent recall performance (Wills et al., 2006).

If the resolution of ambiguity is critical in determining the effects of effort after meaning, why should spending more time puzzling over the meaning of an ambiguous sentence prior to hearing a disambiguating cue enhance recall, relative to hearing the cue immediately before or after the sentence, as was demonstrated in Experiment 4? One possible explanation is that additional time engaging in effort after meaning might allow subjects to develop "insight" into solving the problem at hand. In contrast to the more general "aha!" reaction described by Auble et al. (1979), which can occur even when individuals encounter relatively easy-to-comprehend stimuli, insight typically refers to a process by which a problem solver overcomes a challenging mental impasse by suddenly acquiring a new perspective on how to reach the solution.

Many researchers have argued that developing insight into solving verbal (e.g., riddles) or nonverbal (e.g., mathematics) problems is a unique subjective experience supported by distinct cognitive and neural processes (Duncker, 1945; Köhler, 1925; Luo & Niki, 2003; Maier, 1931; Schooler, Ohlsson, & Brooks, 1993). Moreover, some researchers have maintained that solutions to problems involving insight are remembered better than those derived without insight (Luo & Niki, 2003). Analogously, in the delayed-cue condition of the Auble and Franks (1978) paradigm, subjects might develop insight when trying unsuccessfully for an extended period of time to figure out the meaning of an ambiguous sentence; the presentation of the disambiguating cue suddenly enables them to reorient their thinking in order to achieve comprehension (Luo, Niki, & Phillips, 2004).

The main finding that the effects of effort after meaning are modulated by experimental design has significant theoretical and practical implications. To explain how effort after meaning could improve recall, Auble et al. (1979) suggested three speculative hypotheses. First, under conditions that increase effort after meaning, subjects may be encoding multiple traces of each sentence at study—one trace for the noncomprehended sentence and a second trace for the comprehended sentence. Such an explanation suggests that effort after meaning is comparable to repeated study with varied encoding. Second, sentences requiring greater effort after meaning might be distinctively evaluated against a background of noncomprehended or readily understood sentences, and such distinctive processing might enhance memory for the studied material. Third, sentence recall may be enhanced due to *integration processes* involved in transforming the initial noncomprehended representation of the sentence into the second comprehended representation. Consistent with the encoding specificity principle (Tulving & Thomson, 1973) and the transfer-appropriate processing framework (Bransford, Franks, Morris, & Stein, 1979), recall test conditions may be more similar to the study conditions that enhance effort after meaning to the extent that both require the use of partial sentence information to generate or retrieve a complete representation. More recently, Wills et al. (2000) extended findings of the "aha!" effect to memory for pictorial stimuli and offered the fourth alternative explanation, which is that one's effort after meaning generates alternative, albeit incorrect, cues that may still be used to facilitate later episodic retrieval (see also Wills et al., 2006).

Of these four explanations, only the distinctive-processing explanation readily lends itself to accommodating the observed influence of experimental design. Yet there are several drawbacks to invoking the concept of distinctiveness

(for a review, see Schmidt, 2008). First, it is unclear how to properly classify the various conditions of cue presentation in terms of their distinctiveness, or lack thereof. On the one hand, one could argue that sentences requiring greater effort after meaning are processed more distinctively than are readily understood sentences that reflect the more common experience of day-to-day discourse processing, because the former processing condition is the more unusual. On the other hand, one could just as easily argue that sentences studied in the no-cue condition are more distinctive, because it is more uncommon for a person to process totally ambiguous speech. Furthermore, to simply designate the condition that exhibits superior recall as the more distinctive offers nothing more than a circular definition of distinctiveness, not an explanation of a mnemonic phenomenon.

The multifactor (Hirshman & Bjork, 1988; McDaniel, Wadill, & Einstein, 1988) and item-order accounts (McDaniel & Bugg, 2008; Nairne et al., 1991; Serra & Nairne, 1993) provide a more explicit theoretical approach. According to both accounts, memory for sentences that require greater effort-after-meaning benefits from enhanced elaboration of sentence content information, or item-specific encoding. In addition, according to the multifactor account, when the sentences and cues are presented separately, effort after meaning may also enhance processing of the sentence in relation to the cue. However, this enhancement in item-specific memory and cue–target relational processing occurs at the expense of forming associations among the sentences in the study list, or interitem or whole-list organizational processing.

The item-order account maintains that enhanced item-specific encoding occurs at the expense of encoding temporal or order information regarding sentence presentation within the scope of the entire study list. Indeed, the findings from Experiment 2 confirm the predictions of the item-order theoretical framework. In the within-subjects design, the item-order account correctly predicted superior recall for sentences in the delayed-cue condition, because order memory should be equivalent across the two sentence presentation conditions, and sentences in the delayed-cue condition presumably benefited from enhanced within-sentence (item) processing, relative to the embedded-cue sentences. By contrast, in the between-subjects design, order memory was enhanced for sentences with embedded cues that required relatively little effort after meaning at study, as compared with the delayed-cue condition, thereby compensating for the difference in recall that was observed in the within-subjects design.

In addition to being of theoretical significance, the question of whether the effects of effort after meaning generalize across experimental designs is of practical importance, as well. To the extent that the present findings apply to educational settings, the sensitivity of effort after meaning to experimental design may help educators refine their pedagogy to improve learning in the classroom. For example, although few would object to educators regularly encouraging their students to actively strive to comprehend subject material, the learning benefits of the techniques used to elicit such efforts may largely depend on a host of factors, such as the predictability of the timing and the nature of the information educators provide to help students better understand difficult course material. Specifically, educators may refine techniques such as embedding problems in texts and lectures by having students engage in problem solving to the point where they have reached a mental impasse before providing hints at the solution.

In sum, we find that the effort-after-meaning paradigm (Auble & Franks, 1978) is similar to the generation effect paradigm (Jacoby, 1978; Slamecka & Graf, 1978), in that both are sensitive in the same way to the type of experimental design used to study these phenomena and reveal similar dissociations in item and order memory. Painting with a broad brush, both phenomena are most clearly shown in within-subjects designs and disappear in between-subjects designs. A better theoretical understanding of these effects must await further research. In particular, future research should be aimed at specifying the mechanisms captured by the phrase *effort after meaning*. Several possibilities exist: Effort after meaning may take more time to accomplish, may consume greater processing resources, or may promote better knowledge integration and mental organization. Resolution of these matters lies in the future.

## AUTHOR NOTE

## REFERENCES

ARCHER, E. J. (1960). A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, **74**(10, Whole No. 497).

ASCH, S. E., & EBENHOLTZ, S. M. (1962). The process of free recall: Evidence for non-associative factors in acquisition and retention. *Journal of Psychology*, **54**, 3-31.

AUBLE, P. M., & FRANKS, J. J. (1978). The effects of effort toward comprehension on recall. *Memory & Cognition*, **6**, 20-25.

AUBLE, P. M., FRANKS, J. J., & SORACI, S. A. (1979). Effort toward comprehension: Elaboration or "aha!"? *Memory & Cognition*, **7**, 426-434.

BARTLETT, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge: Cambridge University Press.

BEGG, I., & SNIDER, A. (1987). The generation effect: Evidence for generalized inhibition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 553-563.

BENJAMINI, Y., & HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, **57**, 289-300.

BENJAMINI, Y., & YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**, 1165-1188.

BERTSCH, S., PESTA, B. J., WISCOTT, R., & McDANIEL, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, **35**, 201-210.

BRANSFORD, J. D., FRANKS, J. J., MORRIS, C. D., & STEIN, B. S. (1979). Some general constraints on learning and memory research. In L. S.

Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 331-354). Hillsdale, NJ: Erlbaum.

BRANSFORD, J. D., & JOHNSON, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning & Verbal Behavior*, **11**, 717-726.

CARLIN, M. T., SORACI, S. A., DENNIS, N. A., CHECHILE, N. A., & LOISELLE, R. C. (2001). Enhancing free-recall rates of individuals with mental retardation. *American Journal on Mental Retardation*, **106**, 314-326.

CRAIK, F. I. M., & TULVING, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, **104**, 268-294.

CRUTCHER, R. J., & HEALY, A. F. (1989). Cognitive operations and the generation effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 669-675.

DELOSH, E. L., & MCDANIEL, M. A. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 1136-1146.

DUNCKER, K. (1945). On problem solving. *Psychological Monographs*, **58**, 1-113.

EBBINGHAUS, H. (1964). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Dover. (Original work published in 1885)

ELLIS, H. C., THOMAS, R. L., & RODRIGUEZ, I. A. (1984). Emotional mood states and memory: Elaborative encoding, semantic processing, and cognitive effort. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 470-482.

ERDELYI, M., BUSCHKE, H., & FINKELSTEIN, S. (1977). Hypermnesia for Socratic stimuli: The growth of recall for an internally generated memory list abstracted from a series of riddles. *Memory & Cognition*, **5**, 283-286.

ERLEBACHER, A. (1977). Design and analysis of experiments contrasting the within- and between-subjects manipulation of the independent variable. *Psychological Bulletin*, **8**, 212-219.

ERLEBACHER, A. (1978). The analysis of multifactor experiments designed to contrast the within- and between-subjects manipulation of the independent variables. *Behavior Research Methods & Instrumentation*, **10**, 833-840.

GLAZE, J. A. (1928). The association value of nonsense syllables. *Journal of Genetic Psychology*, **35**, 255-269.

HIRSHMAN, E., & BJORK, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 484-494.

HULL, C. L. (1933). The meaningfulness of 320 selected nonsense syllables. *American Journal of Psychology*, **45**, 730-734.

JACOBY, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning & Verbal Behavior*, **17**, 649-667.

JENKINS, J. J. (1985). Nonsense syllables: Comprehending the "almost incomprehensible variation." *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 455-460.

JOHNSON, M. K., DOLL, T. J., BRANSFORD, J. D., & LAPINSKI, R. H. (1974). Context effects in sentence memory. *Journal of Experimental Psychology*, **103**, 358-360.

KÖHLER, W. (1925). *The mentality of apes.* London: Routledge & Kegan Paul.

LOFTUS, G. R., & MASSON, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, **1**, 476-490.

LUO, J., & NIKI, K. (2003). Function of hippocampus in "insight" of problem solving. *Hippocampus*, **13**, 316-323.

LUO, J., NIKI, K., & PHILLIPS, S. (2004). Neural correlates of the "Aha! reaction." *NeuroReport*, **15**, 2013-2017.

MAIER, N. R. (1931). Reasoning in humans: II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology*, **12**, 181-194.

MCDANIEL, M. A., & BUGG, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review*, **15**, 237-255.

MCDANIEL, M. A., WADILL, P. J., & EINSTEIN, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory & Language*, **27**, 521-536.

MULLIGAN, N. W., & LOZITO, J. P. (2004). Self-generation and memory. In B. Ross (Ed.), *Psychology of learning and motivation* (Vol. 45, pp. 175-214). San Diego: Elsevier.

NAIRNE, J. S., RIEGLER, G. L., & SERRA, M. (1991). Dissociative effects of generation on item and order information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 702-709.

PEYNIRCIOĞLU, Z. (1989). The generation effect with pictures and nonsense figures. *Acta Psychologica*, **70**, 153-160.

PRYTULAK, L. S. (1971). Natural language mediation. *Cognitive Psychology*, **2**, 1-56.

ROEDIGER, H. L., III, & THORPE, L. A. (1978). The role of recall time in producing hypermnesia. *Memory & Cognition*, **6**, 296-305.

SCHMIDT, S. R. (2008). Distinctiveness and memory: A theoretical and empirical review. In J. H. Byrne (Series Ed.) & H. L. Roediger, III (Vol. Ed.), *Learning and memory: A comprehensive reference. Vol. 2: Cognitive psychology of memory* (pp. 125-144). London: Academic Press.

SCHOOLER, J. W., OHLSSON, S., & BROOKS, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General*, **122**, 166-183.

SERRA, M., & NAIRNE, J. S. (1993). Design controversies and the generation effect: Support for an item-order hypothesis. *Memory & Cognition*, **21**, 34-40.

SLAMECKA, N. J., & GRAF, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology*, **26**, 589-607.

SLAMECKA, N. J., & KATSAITI, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory & Language*, **26**, 589-607.

SORACI, S. A., CARLIN, M. T., CHECHILE, R. A., FRANKS, J. J., WILLS, T., & WATANABE, T. (1999). Encoding variability and cuing in generating processing. *Journal of Memory & Language*, **41**, 541-559.

SORACI, S. A., FRANKS, J. J., BRANSFORD, J. D., CHECHILE, R. A., BELLI, R. F., CARR, M., & CARLIN, M. (1994). Incongruous item generation effects: A multiple-cue perspective. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **1**, 67-78.

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

TYLER, S. W., HERTEL, P. T., MCCALLUM, M. C., & ELLIS, H. C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning & Memory*, **5**, 607-617.

WILLS, T. W., ESTOW, S., SORACI, S. A., & GARCIA, J. (2006). The aha effect in groups and other dynamic learning contexts. *Journal of General Psychology*, **133**, 221-236.

WILLS, T. W., SORACI, S. A., CHECHILE, R. A., & TAYLOR, H. A. (2000). "Aha" effects in the generation of pictures. *Memory & Cognition*, **28**, 939-948.

WIXTED, J. T., & ROHRER, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review*, **1**, 89-106.

ZACKS, R. T., HASHER, L., SANFT, H., & ROSE, K. C. (1983). Encoding effort and recall: A cautionary note. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 747-756.

ZAROMB, F. M., KARPICKE, J. D., & ROEDIGER, H. L., III (2009). *The effects of "effort after meaning" on recall and metacognition.* Manuscript in preparation.

**NOTES**

1. In order to minimize confusion, we use the term *effort after meaning* instead of *effort toward comprehension* throughout the article. We prefer *effort after meaning* only because the term highlights connections between the present research and its historical antecedents, even though the phrase *effort after* has the disadvantage of potentially invoking the meaning of *effort following*.

2. According to the sentence-scoring guidelines, it is possible for responses that use different words to receive equivalent scores and, conversely, for responses that are formally quite similar to receive different scores if they do not capture the same ing. For example, a response whose meaning is consistent with that of a study sentence might be granted full credit even though very few words in the response matched words from the original study sentence. On the other hand, a response that shares many of the words from a study sentence but whose overall meaning is inconsistent might not be granted any credit. In order to provide a complementary mea-

sure to the overall sentence recall scores, we scored the number of words precisely recalled per sentence. In this and the subsequent experiments, exactly the same pattern of results and statistical outcomes as that measured for sentence recall was observed in terms of word recall performance. Consistent with the observations of Auble and Franks (1978, Experiment 1), the similar pattern of results that emerged for both measures of sentence and word recall indicate that the effects of effort after meaning on recall are not limited to encoding and/or retrieval at the sentence level.

3. Comprehension ratings collected in a subsequent study (Zaromb, Karpicke, & Roediger, 2009) indicated that one of the sentences was considered relatively easy to understand (Sentence 16; see Appendix B). Nevertheless, the inclusion of this "easy" sentence did not affect the present results. When comprehension ratings and sentence recall performance for the easy and difficult sentences were compared, exactly the same pattern of results and statistical outcomes were obtained as those reported in the Results section with the outlier sentence excluded from the analyses.

4. In this procedure, the mean squares of the main effects of cue presentation condition and experimental design and their interaction are calculated in the normal fashion. However, the error terms for these effects are based on three additional sources of variation: (1) the variation due to subjects, using only the data from the within-subjects portion of Experiment 2; (2) the variation due to the interaction between subjects and cue presentation conditions, using only the data from the within-subjects portion of Experiment 2; and (3) the variation due to subjects within groups, using only the data from the between-subjects portion of Experiment 2. The appropriate error term for the main effect of cue presentation condition and the interaction between cue presentation condition and type of design is the weighted average of the mean squares of 2 and 3. The appropriate error term for the main effect of design type is the weighted average of the mean squares of 1 and 3.

5. We employed a shorter 6-sec interval, rather than the 10-sec interval used in the longer delayed-cue conditions in Experiments 1, 2, and 4, because this experiment was designed to more closely emulate the conditions originally tested in the original Auble and Franks (1978) study, wherein cues were provided following a 5-sec delay.

6. As with Experiment 2, we were concerned that with the two delayed-cue groups, subjects might exert little effort to comprehend the sentences, knowing that the cue would occur shortly after every sentence. To address this concern, we tested two additional groups of subjects and provided them with explicit instructions to try to attempt to figure out the meaning of the sentences prior to hearing their corresponding cues. We found that providing these additional study instructions did not significantly influence recall performance, so we will report only data from the original delayed-cue groups that were given the same intentional learning instructions as the other groups.

## APPENDIX A
### Stimuli for Experiments 1, 3, and 4

| Sentence | Cue |
|---|---|
| 1. The man was hungry because the pole broke. | fisherman |
| 2. The party was stalled because the wire straightened. | corkscrew |
| 3. The breakfast was excellent because the thread was sticky. | spiderweb |
| 4. The person was unhappy because the hole closed. | pierced ears |
| 5. The haystack was important because the cloth ripped. | parachute |
| 6. The needles fell because the bucket was not filled. | Christmas tree |
| 7. The tree grew because the wrinkles opened. | peach pit |
| 8. The clothes were ruined because the sign vanished. | wet paint |
| 9. The street was full of holes because the turning stopped. | cement mixer |
| 10. The notes were sour because the seam split. | bagpipe |
| 11. The family asked for rooms because the light had burned out. | no vacancy |
| 12. The boy was pleased because his rabbit disappeared. | magic act[*] |
| 13. The plan was frustrated because the key locked. | typewriter[*] |
| 14. The food was intact because the plate was broken. | dentures |
| 15. The exam was delayed because the spring broke. | ballpoint pen[*] |
| 16. The mayor was happy because the streets were littered. | ticker tape parade[*] |
| 17. The home was small because the sun came out. | igloo |
| 18. The cut was burned because the clock stopped. | cooking meat |
| 19. The woman could not see because the blade was bent. | windshield wiper |
| 20. The boy spilled his popcorn because the lock broke. | lion cage |
| 21. The man's back ached because the ends were too large. | barbells |
| 22. The smell began because the signs went up. | garbage strike |
| 23. The woman was relieved because the chair was broken. | electric chair[*] |
| 24. The audience cheered because the five balls rotated. | juggler |

[*]Sentence and corresponding cue were created by the experimenter.

## APPENDIX B
### Sequence of Study Trial Events in Experiments 1–4

| Experiment | Condition | Event Sequence |
|---|---|---|
| 3 | No cue | Sentence |
| 1, 2 | Embedded cue | Sentence |
| 3, 4 | Immediate precue | Cue‖ Sentence | * |
| 1 | Precue (2 sec) | Cue | Sentence |
| 1, 3, 4 | Delayed cue (2 sec) | Sentence | Cue | * |
| 3 | Delayed cue (6 sec) | Sentence | Cue |
| 1, 2, 4 | Delayed cue (10 sec) | Sentence | Cue |

```
|  |  |  |  |  |  |  |  |  |  |  |
0   2   4   6   8  10  12  14  16  18  20
                Study Time (seconds)
```

*In Experiment 3, the total study time for each sentence was 15 sec.

## APPENDIX C
### Stimuli for Experiment 2

| Sentence | Cue | Percent* | |
|---|---|---|---|
| **Difficult** | | | |
| 1. The man was hungry because the pole broke. | fisherman | 37 | 23 |
| 2. The audience cheered because the five balls rotated. | juggler | 21 | 14 |
| 3. The cut was burned because the clock stopped. | cooking meat | 21 | 7 |
| 4. The notes were sour because the seam split. | bagpipe | 21 | 0 |
| 5. The haystack was important because the cloth ripped. | parachute | 16 | 17 |
| 6. The smell began because the signs went up. | garbage strike | 10 | 14 |
| 7. The clothes were ruined because the sign vanished. | wet paint | 5 | 7 |
| 8. The home was small because the sun came out. | igloo | 5 | 0 |
| 9. The person was unhappy because the hole closed. | pierced ears | 0 | 27 |
| 10. The food was intact because the plate was broken. | dentures | 0 | 21 |
| 11. The boy spilled his popcorn because the lock broke. | lion cage | 0 | 8 |
| 12. The party was stalled because the wire straightened. | corkscrew | 0 | 0 |
| 13. The breakfast was excellent because the thread was sticky. | spiderweb | 0 | 0 |
| 14. The woman was relieved because the chair was broken. | electric chair | – | 15 |
| 15. The plan was frustrated because the key locked. | typewriter | – | 0 |
| 16. The boy was pleased because his rabbit disappeared. | magic act | – | 79 |
| **Easy** | | | |
| 17. The office was cool because the windows were closed. | air conditioned | 100 | 77 |
| 18. The man got well because the cutting was successful. | operation | 100 | 60 |
| 19. The coffee was sweet because the powder was inserted. | sugar | 100 | 77 |
| 20. The engine stopped because the liquid ran out. | gasoline | 95 | 100 |
| 21. The house tumbled because the insects were satisfied. | termites | 95 | 62 |
| 22. The colors appeared because the rain stopped. | rainbow | 95 | 93 |
| 23. Her toes were pointed because the dance started. | ballet | 95 | 85 |
| 24. The teacher stopped writing because the stick broke. | chalk | 95 | 87 |
| 25. The car was moved because he had no change. | parking meter | 89 | 100 |
| 26. The antelope ran because the spots moved. | leopard | 84 | 75 |
| 27. The numbers were added because the button was pushed. | calculator | 84 | 69 |
| 28. The criminal fled because the red lights flashed. | police car | 84 | 100 |
| 29. The airplane changed altitude because the force was strong. | high wind | 79 | 73 |
| 30. The bicycle was useful because the air was increased. | flat tire | 79 | 87 |
| 31. The betting was interrupted because the wheel was spun. | roulette | 74 | 86 |
| 32. The boy's chest hurt because his friend pulled the straps. | suspenders | 63 | 54 |

*Percentage of subjects who rated each sentence as understandable without the appropriate cue. Numbers in the left column are from Auble, Franks, and Soraci (1979). Numbers in the right column are from Zaromb, Karpicke, and Roediger (2009).