Remembering After a Perception of Discrepancy: Out With the Old, in With the Two

Antonia Kronlund and Bruce W. A. Whittlesea
Simon Fraser University

A surprising validation of expectation experienced during a recognition test induces the perception of discrepancy and a feeling of familiarity. The authors investigated whether that perception also affects memory performance when it is experienced in the original encounter with a stimulus. Target words were presented in a study phase, half in a context thought to induce the perception of discrepancy. In a subsequent recognition test, that earlier experience increased the accuracy of subjects' discrimination. However, when the subsequent task required a once-versus-twice judgment, that experience caused an illusion of reoccurrence for words presented once. The authors concluded that a perception of discrepancy in an initial encounter may be a valuable aid to later recognition but can also cause systematic memory errors under some circumstances.

Keywords: fluency, discrepancy, familiarity, encoding, learning

As illustrated by Mandler's (1980) famous example, an encounter with a marginal acquaintance often does not cause a subjective experience of familiarity[1] when that person is met in an expectable context (e.g., encountering one's butcher in his shop), whereas meeting that person in an unusual context (e.g., on the bus) may cause a powerful feeling of familiarity. Fluency of processing per se does not seem adequate to explain this phenomenon, because the fluency of processing the gestalt of the person's face could be expected to be as great (or greater) in the former situation as the latter. Instead, Whittlesea and Williams (1998) suggested that it is the surprise caused by an experience of fluency in the latter case that is the direct source of the feeling of familiarity. They suggested that the experience of fluent processing in a context in which it should not be expected gives rise to a perception of discrepancy, signaling to the person that there is an inconsistency in their current experience, which must be resolved. In such a situation, one plausible way of resolving this perception is by attributing the unexpected fluency to a previous encounter with the person. By this account, the attribution of unexpected fluency to a source in the past is the direct cause of the subjective experience of familiarity.

The idea that a perception of discrepancy can cause feelings of familiarity has now been extensively documented and examined by means of a variety of phenomena, including the revelation effect (e.g., Bernstein, Whittlesea & Loftus, 2002; Kronlund & Bernstein, 2006), the mirror effect (e.g., Whittlesea, Kronlund, Joordens, & Hockley, 2005), and recognition memory (e.g., Whittlesea & Leboe, 2003; Whittlesea & Williams, 1998, 2000, 2001a). In all of those studies, an illusion of recognition was caused by manipulating the test trials of the experimental situation in a way that would create a disparity between the subjects' expected and actual processing for each trial. Thus, in all of those studies, circumstances were so arranged that the subjects experienced the perception of discrepancy in the moment that they attempted to perform a recognition decision.

However, in each of these experimental contexts, the disparity between the subjects' expected and actual processing can also potentially be an instance of learning, which can have consequences for later interactions with the same stimuli. In the case of the stem completion paradigm, the presentation of the target word after a stem is not just an occasion for evaluating the goodness of the current event and experiencing a feeling of familiarity; it can be an instance of learning the target word. We expected that the perception of discrepancy might not only be the source of feelings of familiarity when experienced in test but might also be a fairly potent source of later remembering when experienced instead in a study episode. That idea is the basis of the experiments of this article.

Background

The sentence stem completion paradigm (Whittlesea & Williams, 2001b) is a useful way to examine the processes that underlie the perception of discrepancy. For example, Whittlesea

---

---

[1] See Whittlesea & Williams (2000) for a discussion of six different meanings of the term *familiarity*. In this article, we are defining familiarity as "the subjective feeling of having prior experience, whether or not one actually has" (Whittlesea & Williams, 2000, p. 547).

and Williams (2001b) presented single words in a study phase. In the recognition test, they presented studied and new words as terminal words of sentences. The sentence stems were of two types: *high constraint* (able to be completed sensibly by only a small number of words: e.g., "She swept the kitchen floor with a . . .") and *low constraint* (able to be completed sensibly by many words: e.g., "She couldn't find a place to put the . . ."). Thus, stems were completed either by a word shown in study (e.g., *BROOM*) or an equally sensible word that had not been shown in study (e.g., *SPONGE*). The result of interest is that high-constraint stems caused an increase in both hits and false alarms of about 5% relative to low-constraint stems, but only when the stems were separated from the target words by a short pause (about 250–750 ms).

This pattern of recognition performance that uses sentence stems is highly stable and has been replicated repeatedly (Whittlesea, 2002a, 2002b). Whittlesea and Williams (2001b) suggested that it results from a *perception of discrepancy*. They suggested that that perception, as well as the associated feeling of familiarity, develop in three steps, involving expectation, inference, and attribution. The high-constraint stems cause the person to develop a strong but indefinite expectation about what is coming next (i.e., a readiness to incorporate one of a small number of concepts, without actually predicting any one of them). The pause causes an experience of uncertainty, a fleeting sense of suspense or realization that one does not know exactly what is coming. This experience of uncertainty causes the target word, when it is shown, to feel not just coherent with the stem, but surprisingly well fitting. In the context of a recognition experiment, the subjects unconsciously attribute the surprise to a prior experience of the target word, consciously experiencing the feeling of familiarity.

The phrase "waiting for the other shoe to drop" provides an example of how a validation (rather than a violation) of expectation can be surprising. In the original story, a man comes in late at night to an inn. He sits on his bed and drops one of his shoes on the floor, before realizing that the others in the inn were asleep. He then takes the other one off much more carefully and quietly puts it on the floor. As he's falling asleep, he hears a shout from the room below him, "I can't sleep, waiting for you to drop the other shoe!" The phrase, as it is used today, means that one is patiently waiting for something to happen that is expected. The waiting period is a period of uncertainty; when the outcome occurs, the person experiences a surprising resolution from not knowing exactly when it will occur. Thus, the experience of a surprising validation is termed the *perception* of discrepancy, and not simply discrepancy, because the term is not meant to describe the stimulus (a discrepant stimulus would not be an expected one); rather, it is meant to describe the subjective feeling of surprise, arising from some sort of uncertainty given the context.

There are several kinds of evidence that the effect depends on an *indefinite expectation*. Whittlesea and Williams (2001b) demonstrated that any manipulation that would cause a definite expectation prevented the effect. For example, the effect does not occur if the target words are completely predictable from the stems (e.g., "Row, row, row your *BOAT*"). There is also convergent evidence that validation of indefinite expectations causes surprise, rather than a heightened sense of fluency. When subjects were asked to judge the predictability of target words presented after constraining stems, subjects judged words to be less predictable when

presented with rather than without a pause. That observation is the opposite of what could be expected from an increase in fluency of processing; instead, it is in line with the idea that the pause caused the words to feel somewhat surprisingly well fitting (Whittlesea, 2002b).

The Current Paradigm

In Whittlesea and Williams' experiments, the perception of discrepancy was induced during the test session; the issue was how that perception could contribute to feelings of remembering something that happened earlier. However, the perception of discrepancy also serves an alerting function, drawing increased attention to the stimulus. For that reason, it may also serve as an occasion of learning. In the butcher example, the perception of discrepancy created by the encounter of the butcher on the bus may cause that particular encounter to later be accurately discriminated from other, prior encounters with the butcher. To demonstrate this effect, in Experiment 1 we induced the perception of discrepancy during the study session. We observed that that manipulation increased success in remembering in a subsequent recognition test, indicative of more effective encoding during the study session.

The recognition test in that study used the standard technique of showing a list of words, half of which had been studied and half of which had not. Although useful for answering the question posed about surprise and learning, that procedure fails to provide discriminating evidence about how the person uses that knowledge in the remembering test. One obvious possibility is that more effective representations cause more effective retrieval of earlier experience during the test. However, as alluded to earlier, people may often make a recognition decision based on a feeling of familiarity, resulting from an attributional process rather than actual retrieval. The effects on these two processes of factors manipulated during the test can be disambiguated by the false alarm rate associated with each experimental condition: Retrieval could not increase the false alarm rate, because false alarms are reports of events that did not occur and so could not be retrieved. In contrast, because, as argued earlier, the feeling of familiarity results directly from an attributional process, that feeling could be induced by misleading the person's evaluation of their own processing, leading to additional false alarms.

That same logic cannot be applied to examine the effects of factors manipulated during the study session, because false alarms refer to items that were not shown in study and so could not have been influenced by those factors. To investigate the effects of a perception of discrepancy occurring during study on the processes of remembering during test, we therefore modified the standard procedure: In Experiments 2 and 3, we presented all test words in the study session but presented them either once or twice. The test question thus changed from "Did this item occur in the study session?" to "Did this item occur twice in the study session?" Because all words presented in test had been presented in study, we could apply our manipulation of choice to half of the study items and examine the effects of that manipulation on test performance. Moreover, the same general form of logic can be applied in examining the effects of that factor on retrieval versus attribution: Actual retrieval could not make a person say that a word presented only once was presented twice, but the evaluation-and-attribution process that is of interest here could. We thus changed the defi-

nition of *hits* to correctly answering the test question ("Was it presented twice?") and of *false alarms* to falsely claiming that a word shown once was shown twice. We observed that an experience of the perception of discrepancy during the study session increased the rate of false alarms as much as hits, indicating that the effects of that study experience were mediated during test by attributional processes, not retrieval alone.[2]

Our point is not that remembering may be erroneous. Rather, we present this evidence as important to make three major points: (a) to argue that these empirical patterns are regular and consistent and so must be able to be explained by any thorough account of memory; (b) to argue that these effects show an attributional (as opposed to a retrieval) function of memory; and (c) to argue that such attributions can be understood through the theoretical construct we call the perception of discrepancy.

Experiment 1: Learning With and Without a Perception of Discrepancy

In previous studies using the sentence stem paradigm to study the perception of discrepancy, the stems were presented during the test phase only. We modified the paradigm slightly by presenting whole sentences during the study phase; in test, we presented either the identical sentences or sentences that were terminated by a word originally shown with some other stem (in the latter case, the sentence as a whole still made sense). Half of the stems in study and half in test were of high constraint, the remainder of low constraint. This procedure has been used to study a related source of feelings of familiarity, namely, the perception of integrality (i.e., predictability or unity; Whittlesea, 2002a, 2004; Whittlesea & Koriat, 2006). The relationship between that perception and the perception of discrepancy is discussed later. The standard outcome observed in studies using this procedure is approximately equal numbers of hits for high- and low-constraint sentences but fewer false alarms for low-constraint sentences. However, this effect was not the major focus of our studies; instead, these conditions served as the baseline for examining the effects of the perception of discrepancy as a foundation for later recognition.

To investigate the perception of discrepancy during study, in the current experiment we presented whole sentences in study with a pause inserted between the stem and target word. These sentences also were of high versus low constraint and were preserved or modified in test in the same way as sentences presented without a pause. As described earlier, a perception of discrepancy occurring in test increases both hits and false alarms because the meaning of that perception ("I must have seen this before") is the same whether the target word is actually old or new. That is, when experienced in test, surprise causes a bias shift. However, we expected that the same perception experienced during study would have a different effect, instead affecting the discriminability of old versus new sentences. In principle, that could occur in either of two ways: by increasing attention to the relationship between stem and target word, thereby increasing later discrimination, or by distracting the subject from that relationship, thereby decreasing later discrimination. We predicted that the former would occur because of evidence from prior work: that people encountering complete sentences develop themes or schemas representing the meaning of the sentence as a whole (Whittlesea, 2002b, 2004; Whittlesea & Koriat, 2006).

Method

Subjects. Twenty-nine Simon Fraser University students participated for course credit in Experiment 1A; a different group of 29 students participated in Experiment 1B.

Procedure. The experiments used the 60 frames presented in the Appendix of Whittlesea (2002b). Each frame consists of a pair of stems (e.g., "The policeman identified himself with his . . ." and "Every elected official should carry a . . .") and a pair of terminal words (e.g., *BADGE, CARD*), either of which can sensibly complete either stem. As suggested by that example, in each pair one of the stems could be sensibly completed only by a small number of words (high-constraint stem) and the other could be completed with many words (low-constraint stem).

In Experiment 1A, prior to the study phase, the subjects were informed that they would see sentences and that in a subsequent test they would see the identical sentences or ones in which the last word was replaced. A total of 120 sentences was presented in the study phase. Sixty were presented without a pause; on the remaining trials, the stem was presented for 2,500 ms (terminated with an ellipsis, indicating that the sentence was incomplete). The ellipsis was then replaced by the target word for that trial, written in capital letters. (The duration of 2,500 ms is long enough for the average reader to have completed reading the stem; thus, for most readers there would be a pause of variable interval prior to the onset of the target word. This procedure for inducing a pause has also been used by Whittlesea & Williams, 2001b). Crossed with the pause factor, 60 stems were of high constraint and 60 were of low constraint. Target words of each pair were assigned at random to stems for each subject (e.g., producing either "The policeman identified himself with his *BADGE*" or "Every elected official should carry a *CARD*," or vice versa).

At test, all 120 stems were shown again; all were completed by a target word (i.e., there were no pauses at test). Half of the sentences shown with a pause in study and half shown without a pause were completed by the same word as in study; the remainder was completed with the alternate word for that pair. In the latter case, the words were interchanged within a pair, so that both sentences of the pair ("policeman" and "official") were literally novel. Subjects decided whether the sentences were identical to study sentences or whether the last word was changed, striking one of two keys on a button box to record their decision.

Experiment 1B was identical in all ways except that in study trials involving a pause, the subjects were required to make a guess about the target word before it was shown. On such trials, the stem was presented with an ellipsis; the subject typed in a guess about the terminal word, using the space bar to indicate that they were finished; and only then was the target word shown, with the instruction "Actual sentence." The subjects were instructed to use the word shown in that sentence to perform the later recognition task if it differed from the word they guessed.

Results and Discussion

The experimental predictions were unidirectional, as discussed earlier, on theoretical grounds. An alpha of .05 is assumed throughout. Probabilities of claiming "old" for Experiment 1A are reported in the top section of Table 1. Data were subjected to a repeated measures analysis of variance. The subjects discriminated

---

[2] We are not attempting to develop a theory of frequency per se (for a retrospective on this literature, see the review by Zacks & Hasher, 2002); instead, we are using the once-versus-twice judgment task as a tool for examining the consequences of experiencing the perception of discrepancy during learning. The standard recognition task is also a sort of frequency task, test items having been presented at frequencies of zero and one. The task we used is the closest possible procedure that permits examination of retrieval versus attribution.

Table 1
Experiment 1: Judgments of Prior Occurrence

	Condition at test			
	Old		New	
Variable	Condition during study			
	Pause	No pause	Pause	No pause
Experiment 1A				
High constraint	.77 (.03)	.67 (.04)	.30 (.03)	.38 (.04)
Low constraint	.66 (.04)	.66 (.03)	.22 (.03)	.25 (.02)
	Generate	No pause	Generate	No pause
Experiment 1B				
High constraint	.92 (.01)	.59 (.04)	.38 (.04)	.35 (.04)
Low constraint	.88 (.02)	.56 (.04)	.29 (.03)	.24 (.02)

Note. Standard errors are reported in parentheses.

well between old and new words, $F(1, 28) = 160.34$, $MSE = .06$, $\eta^2 = .85$. For sentences that had originally been presented without a pause, the constraint of stems caused very little difference in hits (1%; see the second column of Table 1, top two rows; $F < 1$), whereas false alarms were 13% lower for low-constraint stems compared with high-constraint stems (fourth column, top two rows), $F(1, 28) = 13.58$, $MSE = .02$, $\eta^2 = .32$. That is the same effect observed by Whittlesea (2002b, 2004) and Whittlesea and Koriat (2006) in studies using the same procedure. The meaning of this effect is discussed below.

When study sentences were presented with a pause, a very different pattern of effect was observed. The pause had no influence with low-constraint stems, either when presented with old sentences (Table 1, second row, first vs. second column; $F < 1$) or re-paired sentences (second row, third vs. fourth column; $F < 1$). However, inserting a pause between the stem and target word resulted in 10% greater hits for high-constraint stems (top row, first and second columns), $F(1, 28) = 10.72$, $MSE = .01$, $\eta^2 = .27$. It also resulted in 8% fewer false alarms for sentences with high-constraint stems (top row, third vs. fourth column), $F(1, 28) = 7.67$, $MSE = .01$, $\eta^2 = .22$. That is, presenting a pause before the target word of high-constraint stems in the study phase improved the subjects' discrimination between original and re-paired sentences in test relative to the same conditions without a pause but had no effect on discrimination of sentences containing low-constraint stems.

The effect of inserting a pause in a sentence shown in study is thus very different from its effect when inserted in a sentence shown for the first time in test: In the latter case, as described earlier, it results in greater hits and false alarms for target words following high-constraint stems. The latter effect has been interpreted to result from a perception of discrepancy occurring during the test (Whittlesea & Williams, 2001b; Whittlesea, 2002b). We argue that the effect of the pause in study is also mediated by that perception; the difference in effect of that perception is a consequence of when it is experienced. We suggest that seeing a high-constraint stem in study followed by a pause induces an indefinite expectation about its target word, exactly as the same experience does when that procedure is adopted in test. When in the test, the surprise occasioned by seeing a high-constraint stem completed with a sensible termination after a period of uncertainty

causes the subject to experience a feeling of familiarity; when instead in the study phase, the same perception of discrepancy causes the subject to attend more closely to the relationship between the stem and target word, resulting in more hits and fewer false alarms in the subsequent test. This effect does not occur with low-constraint stems, because those stems do not arouse any expectation and so do not occasion any surprise.

However, there might be an alternate explanation. The combination of a high-constraint stem and a pause might stimulate subjects to guess the termination before it is actually presented. That would also lead to better discrimination in that condition (the read/generate effect; Slamecka & Graf, 1978) and would not entail a role for the perception of discrepancy.

Experiment 1B was conducted to test between these alternatives. It was identical in all ways to Experiment 1A, except that on pause trials, the pause was extended indefinitely until the subject had made a guess about what word would be used to complete the sentence.[3] They made correct guesses on about 40% of trials presenting high-constraint stems but less than 5% of trials presenting low-constraint stems, confirming the intended difference in constraint of those stems. As shown in the bottom section of Table 1, the resulting pattern of recognition claims was very different from that observed in Experiment 1A. First, the increase in hits for old high-constraint stems was much larger in Experiment 1B (33%; third row of Table 1, first vs. second column) than in Experiment 1A (10%; top row, first vs. second column); further, whereas the pause had no effect on claims about old sentences with low-constraint stems in Experiment 1A, the generation requirement caused 32% more hits for the same sentences in Experiment 1B (fourth row of Table 1, first vs. second column). In consequence, we now observed no reliable effect of constraint, $F(1, 28) = 2.25$, $MSE = .02$, $\eta^2 = .07$, but a highly reliable effect of the pause manipulation, $F(1, 28) = 109.92$, $MSE = .03$, $\eta^2 = .80$. Moreover, whereas presenting a pause between a stem and target

---

[3] This guessing procedure was also used by Whittlesea and Williams (2001b). However, they used it in conjunction with stems seen for the first time in test. The effect of the procedure in that case was to eliminate the false familiarity effect associated with high-constraint stems, whether the subject guessed the correct target word or not.

word in the study phase of Experiment 1A caused an 8% decrease in false alarms for high-constraint stems shown in test with a changed target word (top row of Table 1, third and fourth columns) there was no sign of such an effect in Experiment 1B. Instead, for sentences with changed target words, we now observed a reliable effect of constraint of stem, $F(1, 28) = 14.45$, $MSE = .02$, $\eta^2 = .35$, but no reliable effect of the pause manipulation, $F(1, 28) = 1.48$, $MSE = .04$, $\eta^2 = .05$.

We will not perform an extended interpretation of these results; our only concern with this experiment was to observe whether deliberate guessing caused a pattern of data similar to that in the earlier study. It seems clear that what happens when the subjects guess about terminations is very different from what occurs when they have only a short pause between the stem and target word. We concluded that the results of Experiment 1A were not the result of generating guesses but instead the product of the perception of discrepancy in reading the study word.

Experiment 2: Remembering after a Perception of Discrepancy: Occam's Razor or a Double-Edged Sword?

As discussed earlier, inducing a perception of discrepancy in test biases the recognition process, increasing both hits and false alarms (e.g., Whittlesea, 2002b), whereas the results of Experiment 1A demonstrate that inducing a perception of discrepancy in study instead increased later discrimination, increasing hits but reducing false alarms. That effect was observed by re-presenting sentence stems at test. In Experiment 2, we investigated the generality of this phenomenon to isolated words at test. To examine the effect of the perception of discrepancy during study on hits and false alarms for words presented without sentence stems at test, we used a once-versus-twice task.

Words were presented either once or twice in study. In the first variant of this study (Experiment 2A), once-presented words were presented either alone or following a high-constraint stem and pause; twice-presented words were presented either in isolation on both occasions or once alone and once following a high-constraint stem and pause. Target words were presented in isolation in the test phase; subjects were asked, "Did this word occur twice in the study phase?" When a word was presented twice, a "yes" response was a hit; when a word was presented only once, a "yes" response was a false alarm.

Method

Subjects. Twenty-two Simon Fraser University students participated in Experiment 2A and 19 in Experiment 2B for course credit.

Procedure. The stimuli consisted of the high-constraint stems used in Experiment 1, together with one or the other of the target words used with those stems in that study. Subjects in Experiment 2A were told that they would get an unspecified memory test following a study session. During the study phase, half of the target words were presented in isolation for 1,000 ms each. The remaining target words were presented after their respective constraining stems. Each stem was presented for 3,000 ms before presenting the target word, thus allowing for a pause between reading the stem and seeing the target word. The length of the pause was variable, depending on the speed of reading of each individual subject. The target word was then presented in capital letters along with its stem for 1,000 ms. Crossed with the manipulation of isolated versus sentence presentation, half of the target words were presented once, half twice. In

the latter condition, the stem was presented with the target word on only one occasion and in isolation on the other. At random, half of the words in this condition were presented in isolation first and later in a sentence, the remainder in the reverse sequence.

Thus, the four conditions of Experiment 2A were single presentation of a word in isolation; single presentation in a sentence; double presentation of a word, both times in isolation; and double presentation, once in isolation and once in a sentence. Assignment of words into conditions and presentation order was randomly determined for each subject. In the test phase, subjects were re-presented with all target words shown in study (in a freshly randomized order). Subjects indicated whether they had seen the target word once or twice in study by pressing a key on a button box to record their decision.

Experiment 2B was similar, except that a complete sentence was shown on every trial in the study phase. Half of these sentences were shown once and half twice (and so were their target words). Crossed with that manipulation, both of the occurrences of a twice-presented sentence, or the only occurrence of a once-presented sentence, was presented with a pause on half of the trials. On trials meant to not have a pause, the stem and target word were presented together for 4,000 ms. On the trials with a pause, the stem was presented for 3,000 ms. The target word, in capital letters, was then presented along with the stem for 1,000 ms.

Thus, the four conditions of Experiment 2B were single presentation of a sentence without a pause; single presentation of a sentence with a pause; double presentation of a sentence, with a pause on both occasions; and double presentation of a sentence, without a pause on either occasion. The test was similar to Experiment 2A: All target words were shown in isolation, and the subjects indicated whether each word had occurred in one or two study presentations.

Results and Discussion

In Experiment 2A, subjects were successful at the discrimination task, judging twice-presented words to have occurred twice about 23% more often than once-presented words, $F(1, 21) = 92.07$, $MSE = .01$, $\eta^2 = .81$ (see Table 2, first two columns). Presenting a target word in a sentence increased accuracy of reporting that it had occurred twice, by 7%, $F(1, 21) = 5.84$, $MSE = .02$, $\eta^2 = .24$. However, it also increased the false alarm rate (claiming "twice" for once-presented words) by 9%, $F(1, 21) = 6.19$, $MSE = .02$, $\eta^2 = .22$.

Thus, presenting a sentence stem and pause in study increased the likelihood of judging both once- and twice-presented words to have been seen twice. However, this might not be an effect of the perception of discrepancy, rather, the stem and pause in study may have acted to increase depth of processing, which has been shown to cause an effect similar to this in the once-versus-twice judgment, increasing both false and accurate claims of reoccurrence (Kronlund & Whittlesea, 2005). Experiment 2B was conducted to decide between these alternatives.

In Experiment 2B, the subjects were even more successful in the discrimination task, judging twice-presented words to have occurred twice about 50% more often than once-presented words, $F(1, 21) = 92.07$, $MSE = .01$, $\eta^2 = .81$ (see Table 2, middle two columns). Presenting a pause on both occurrences of twice-presented words increased the number of hits by 24%, $F(1, 18) = 38.61$, $MSE = .01$, $\eta^2 = .68$, but it also increased the number of false alarms by 6%, $F(1, 18) = 6.00$, $MSE = .01$, $\eta^2 = .27$. Because all target words had been seen in complete and constraining sentences in the study phase in this experiment, differences in degree or depth of elaboration seem unlikely. Instead, we con-

Table 2
Experiments 2 and 3: Judgments of Prior Reoccurrence

Actual frequency of presentation	Experiment and condition during study					
	2A		2B		3	
	Stem	No stem	Pause	No pause	Pause	No pause
Once (false alarms)	.38 (.03)	.29 (.03)	.22 (.03)	.16 (.02)	.12 (.02)	.06 (.02)
Twice (hits)	.60 (.02)	.53 (.03)	.81 (.03)	.57 (.03)	.76 (.03)	.47 (.04)

Note. Experiments 2B and 3 were identical with the exception that in Experiment 3, at test, subjects were also shown new items and given the option of responding "seen zero times" in addition to "seen once" or "seen twice." Standard errors are reported in parentheses.

cluded that the effect of the pause was (at least in part, as discussed later) mediated by a perception of discrepancy in the study session.

These experiments demonstrate that experiencing a perception of discrepancy in the original event need not inevitably increase later discrimination, as it did in Experiment 1A; instead, in this case, experiencing that perception in study biased the decision process toward the conclusion that words had been seen twice. The principle governing the effect of a perception of discrepancy on later processing seems clear: It enhances processing of the relationship between the target stimulus and its context in the original event. That assists the subject in a straight recognition test, allowing them to be more sensitive to recurrence and to more easily reject changed presentations. However, that enhanced encoding of the item–context relationship can only be used to determine that the stimulus has been encountered at least once, which is of little help in a once-versus-twice decision, whereby subjects have to rely on other characteristics of the stimulus, such as vividness, clarity, and ease of processing. On the assumption that multiple prior presentations augment such characteristics, influencing the subjective quality of the trace, the subjective quality can sensibly be used to argue back to reoccurrence of prior experience. By enhancing encoding of the item–context relationship, the perception of discrepancy causes false claims about once-presented words because the subjective quality of the trace is used as a heuristic to infer multiple prior experiences. We call this the *subjective quality heuristic*.

A curious sidelight of this study is that two presentations of a word following a stem with versus without a pause causes a much greater difference in hits (24%; Table 2, bottom row, middle two columns) than does double presentation of a word with versus without one presentation following a stem and a pause (7%; bottom row, two left-most columns). That means that the second presentation of a word in a sentence context contributes a great deal to the accuracy of decision. The curious aspect of this is that the second presentation of the sentence is identical to the first, just as the second presentation of a word alone is identical to the first. It seems unlikely that the person would experience the perception of discrepancy (surprise) on a second occurrence of the same sentence, even when the target word is delayed. Instead, we suspect that this effect is mediated by a different mechanism: On the second occasion of seeing a stem, the person develops a definite expectation of what word is about to occur (an idea that is discussed later in detail). The pause then contributes to this learning experience by allowing the subjects the opportunity to actually

generate that word for themselves (an act that is verified by presentation of the target word a moment later).

This act of generation does not assist the subject by enhancing the representation of the relationship between the stem and target word (which again would later only enable them to know that they had encountered that word at least once: useful information in a recognition test, as shown in Experiment 1B, but not in a once-versus-twice decision). Instead, it does so by allowing them to realize, at the moment of generation during the second study presentation, that they are experiencing this word for the second time, thereby encoding the word as a reoccurrence at that time. In this case, a reoccurrence can act as a reminder to subjects that this is actually a repeat, which makes it much easier to decide, at test, that twice-presented words were presented twice (cf. Hintzman, 2004). The effect is not seen in Experiment 2A because although a second presentation is undoubtedly enhanced by the first (just as a second presentation of a word with a stem without a pause is enhanced in Experiment 2B), the subject cannot engage in the act of generation.

Thus, we do not conclude that the entire effect of the pause observed in Experiment 2B is due to the perception of discrepancy. However, the illusion of multiple prior presentations for once-presented words is almost certainly a product of that perception, and the same mechanism also operates in making accurate decisions about twice-presented words (Experiment 2A).

Experiment 3: The Perception of Discrepancy: An Illusion of Reoccurrence

One potential criticism of Experiment 2 is that our result could be simply a replication of Experiment 1's finding, but shifted along the frequency scale, rather than a genuine change in the once-versus-twice estimation.[4] Because the "new" option was not available to subjects, weaker versus stronger memory for a target word could not have lead them to classify the target word as "old" versus "new"; the only way for them to express this difference was by using the "once" versus "twice" response options. Allowing subjects to respond with a "zero" response option, in addition to the "once" and "twice" options, would permit a stronger examination of whether the perception of discrepancy actually led to a genuine illusion of reoccurrence. Showing the same pattern of data as that

---

[4] We thank an anonymous reviewer for pointing this out.

shown in Experiment 2, with the "zero" option, would demonstrate something that is qualitatively different from the demonstration that the perception of discrepancy leads to better memory for the encoding experience (Experiment 1).

This experiment was identical in all ways to Experiment 2B, with the exception that during the test phase, new target words were presented; subjects were asked to discriminate between targets presented zero, once, or twice during study.

Method

Subjects. Twenty Simon Fraser University students participated for course credit.

Procedure. The stimuli and procedure matched those of Experiment 2B in all ways except that stimuli were randomly assigned to be "old" and "new" at test; if old at test, the words were broken down into the four conditions outlined in Experiment 2B. At test, subjects were asked to indicate whether they had seen the target zero times, once, or twice in study by pressing a key on a button box to record their decision.

Results and Discussion

Results are summarized in Table 2 (see right two columns). Two comparisons were carried out to examine the critical results: the effects of the pause manipulation on claims of "twice" for once-presented words (false alarms) and for twice-presented words (hits). As can be seen in the top row, subjects judged once-presented target words to have occurred twice about 6% more often in the pause condition versus the no pause condition, $F(1, 19) = 4.87$, $MSE = .01$, $\eta^2 = .19$. Subjects also judged twice-presented targets to have occurred twice 29% more often in the pause condition than in the no pause condition, $F(1, 19) = 53.10$, $MSE = .02$, $\eta^2 = .74$. Thus, even with the option of "zero" presentations, subjects are more likely to claim "twice" to once-presented words with a pause.

General Discussion

Until now, the perception of discrepancy has been investigated only within the context of test trials. That is important because one of the prime functions of that perception is to alert the person to a potential prior experience of a stimulus by creating a feeling of familiarity in the moment. However, the present experiments reveal a second, potentially more important, role of that perception: that it can cause the person to integrate the components of an event to a greater degree than they would otherwise. That can increase the accuracy of later recognition, at least when tested in recognition, as shown in Experiment 1.

Illusory Remembering Versus Real Remembering

Experiments 2 and 3 demonstrate that the perception of discrepancy in an original experience can later lead to an illusion of familiarity of an event that did not take place; the studies by Whittlesea and Williams (2001b) and Whittlesea (2002a, 2002b) demonstrate that the perception of discrepancy occurring at the time the person attempts to remember can also cause that illusion. Moreover, the perception of discrepancy induced by the insertion of a pause between stems and target words in all of those studies was itself illusory: Normatively, the target words merely fit their

context well. The subjects' surprise on experiencing coherent target words after a pause and constraining stem is a result of their interpretation of their performance, not a product of the stimulus itself (that is why we speak of a perception of discrepancy, rather than discrepancy).

The occurrence of those illusions is valuable in establishing how the perception of discrepancy causes feelings of familiarity through the development of expectations, evaluation of outcomes relative to those expectations, and attribution of a perceived disparity to a plausible source. However, the occurrence of such illusions does not mean that the perception of discrepancy ordinarily causes illusory recognition; it did so in those studies because the situation was rigged to produce erroneous evaluations, using a procedure (the pause) that people would rarely have encountered in their daily lives and so would not know how to discount. In fact, we suspect that the perception of discrepancy ordinarily causes appropriate feelings of remembering. For example, in a standard recognition experiment, items shown earlier are more easily processed on a subsequent occasion (repetition priming; e.g., Scarborough, Cortese, & Scarborough, 1977). That facilitation causes the fluency of processing to be greater than could normatively be expected for that item; the perception of a discrepancy between expectation and outcome thus serves as a useful basis for discriminating between old and new test items. The perception of discrepancy has similarly beneficial effects in more natural settings, as can be illustrated with the example of the butcher on the bus (Mandler, 1980) discussed earlier.

The Perceptions of Discrepancy and Integrality

As reported in Experiment 1A, when complete sentences having high- and low-constraint stems were presented in study without a pause, we observed approximately equal hits but fewer false alarms for low-constraint sentences as opposed to high-constraint sentences. That same pattern was also observed by Whittlesea (2002b, 2004) and Whittlesea and Koriat (2006). They concluded that this pattern of effect comes about in a different way, through the development of definite expectations rather than through the indefinite expectations responsible for the perception of discrepancy.[5] The idea is that exposure to a sentence in study causes the person to integrate or unitize its parts by thinking about the theme or schema of the sentence as a whole. This schema can serve as an expectation about how subsequent processing of an event should unfold. Low- versus high-constraint stems cause expectations dif-

---

[5] Because of their differential origins in indefinite versus definite expectations, the perception of discrepancy typically causes inconclusive feelings of remembering (a generic feeling of familiarity, unaccompanied by an ability to identify the specific source of that feeling), whereas the perception of integrality is typically accompanied by more categorical and well-defined cognitions about the past. However, unlike two-process accounts of remembering (e.g., Atkinson & Juola, 1973; Gardiner & Conway, 1999; Jacoby, 1991; Mandler, 1980; Tulving, 1985), which assume that familiarity and recall of context are directly linked to different forms of representation or process, the SCAPE framework (Selective Construction And Preservation of Experience; Whittlesea, 1997) assumes that either subjective experience can be caused by either perception, depending on the context within which they are interpreted. See Bodner and Lindsay (2003) for a similar perspective.

fering in specificity, such that the meaning of schemas formed from sentences with low-constraint stems is dominated by the meaning of the target word, whereas the meaning of schemas formed from high-constraint sentences is less specific to that particular word and relatively more determined by the meaning of the stem. Thus, for example, the high-constraint sentences "After the accident he was covered in *GLASS*" and "After the accident he was covered in *BLOOD*" have much the same meaning (hurt, damage), whereas the low-constraint sentences "On the corner of the table there was a bit of *GLASS*" suggests a minor mishap but "On the corner of the table there was a bit of *BLOOD*" suggests something more frightening, such as murder. One effect of this difference is that high-constraint stems cause more reports of the last word in a recall test (although many of these reports are inaccurate; Whittlesea, 2002b); another is that the subjects are less accurate in predicting their later ability to remember low-constraint sentences, even though those sentences later cause more accurate remembering (Whittlesea & Koriat, 2006).

These definite expectations are aroused by re-presentation of sentences in test. Sentences that are identical to ones seen earlier validate these expectations, causing a perception of integrality (wholeness, or unitariness). That perception in turn produces a higher rate of hits than occurs for the same target words presented either alone or in sentences, when complete sentences are not presented in study. The difference in false alarms rates occurs because of the difference in specificity of schemas formed from low- and high-constraint sentences: Because the latter are less specific to the particular target word, a different word can sometimes be substituted (e.g., *BLOOD* for *GLASS*) without the subject realizing it, so long as the overall meaning of the sentence is preserved.

The insertion of a pause between stems and target words in the study sentences of Experiment 1A caused a twist in this process. When the study stems were of high constraint, the subjects developed an indefinite expectation on reading the stem. That indefinite expectation coupled with the pause produced a perception of discrepancy, a sense of surprise that caused the subject to integrate the specific meaning of the target word more effectively into the developing schema of the sentence. That is, the definite expectation that was operative at test was modified by an indefinite expectation occurring on line while reading the sentence on the first occasion. That made the schema more effective as an expectation on a subsequent occasion, resulting in the observation of higher hits and lower false alarms than when no pause had occurred on the original occasion. The low-constraint stems did not cause this indefinite expectation so that nothing different happened when a pause was inserted compared with when it was not.

Although complex, these results are not really surprising. People remember extended events better than single, isolated occurrences and remember surprising events better than events that although consisting of novel combinations of familiar units can be assimilated easily by existing knowledge structures. What is interesting about the current pattern of data is the interaction between two important aspects of remembering, the subjective experience that the subject has on an original encounter with a stimulus and a later encounter and the quality of the representation that is encoded on the first encounter and that controls processing of the second.

SCAPE Versus Other Accounts

The experiments in this article were based on the assumptions of the SCAPE framework (Selective Construction And Preservation of Experience; Whittlesea, 1997; Whittlesea & Leboe, 2000). This account attempts to present an integrated understanding of the objective properties of people's experience and performance as well as the phenomenology that accompanies that processing. By this account there are two basic functions of mind: production and evaluation. The former results from the interaction of the stimulus with the task, context, and cued traces and results in cognitive and motoric performance. The latter consists of assessment of the quality and content of that performance in the context of salient aspects of the task and intuitive theories of cause and effect; it produces phenomenology, the person's subjective reactions, attitudes, or feelings. By this account, an act of remembering is essentially a two-part event: the target word, serving as a cue, causes some behavior to unfold: a coming-to-mind of the context of a previous experience or perhaps merely enhanced fluency of identification. The person then has to assess the significance of these mental events: How likely are they to be caused by actual previous experience; how likely are they to be caused by some other factor? The answer to this question determines the person's feeling about their current processing: The true source affects the evidence available for this decision, but not the decision process.

In these experiments, we observed that presenting a pause before the terminal word of a study session increased hits in a recognition test but also increased false alarms in a once-versus-twice test. This effect of the pause stands as a challenge to many accounts of learning and remembering. Two common accounts of these processes include signal detection (e.g., Donaldson, 1996; Glanzer, Kim, & Adams, 1998; Miller & Wolford, 1999; Wickens & Hirshman, 2000; Wixted & Stretch, 2000) and global matching approaches, which postulate that test cues activate many memory traces, depending on the degree of similarity between the cue and trace(s); a criterion is used to decide old–new decisions (e.g., Eich, 1982; Gillund & Shiffrin, 1984; Humphreys, Bain, & Pike, 1989; Murdock, 1982) or apparent frequency (Hintzman, 1988; Murdock, Smith, & Bai, 2000; Shiffrin, 2003; Shiffrin, Ratcliff, & Clark, 1990; Shiffrin & Steyvers, 1997; see also Dougherty, Gettys, & Ogden, 1999).

Signal detection and global matching accounts assert that there is a direct correspondence between production and phenomenology: The strength of the representation, or the strength of the echo from the representation as currently cued, determines the strength of the person's feeling of remembering. In principle, signal detection and global matching accounts could explain the various results of our studies by arguing that inserting the pause led to better encoding of words, consequently leading to either greater strength of the trace (according to signal detection accounts) or greater match between the representation of the study and test experiences (according to global matching accounts). These assumptions could explain both the greater discriminability between new and old words observed in Experiment 1A and the illusion of double presentation observed in Experiments 2 and 3. However, as demonstrated by Whittlesea and Williams (2001b), neither of these accounts can provide an adequate account of the variety of effects that are observed when a pause is presented during test trials. More germane to the present purpose, neither of these accounts has a

motivated reason to predict that inserting a pause would have such an effect on encoding. Their assumptions about criterion-setting and trace-matching are designed to explain the act of recognition: They do not offer any insights into varieties of performance in the initial act of the encoding.

The most plausible explanation consistent with signal detection and global matching accounts theories is that presenting a word after a pause isolated it from the stem, thereby increasing the amount of attention the subject pays to it. However, there is no obvious theoretical reason to make this assumption. Further, although that could explain the illusion observed in Experiment 2B, in which once-presented words followed their stems with a pause inserted or not, it cannot explain the similar illusion reported in Experiment 2A, in which once-presented words were shown either in isolation or following a stem and a pause. Presumably, if the effect in Experiment 2B were based on psychological isolation of the word, then the actual physical isolation of the words in Experiment 2A should have created as many false alarms as those presented with a stem and a pause in Experiment 2A.

In contrast to signal detection and global matching accounts, the SCAPE framework provides a concrete psychological mechanism whereby the pause has its effect in Experiment 1A, namely the experience of surprise, which attracts attention and consequential better encoding. Experiment 1B buttresses that explanation by demonstrating that the effect of the pause in decreasing false alarms for high-constraint items disappears when the subject generates candidate terminations and hence is unlikely to be surprised by the termination presented experimentally. More important, SCAPE's explanation of the effect of the pause in study items is derived from its explanation of the same manipulation in test items. In both cases, we assume that the proximal mechanism underlying the effect of the pause is surprise. In test items, this surprise causes a need to justify, resulting in an attribution to a plausible source (in this case, the past, leading to a feeling of familiarity), which results in both greater hits and false alarms. In study items, the same experience of surprise causes greater attention to the item, producing richer encoding, which results in greater hits but lower false alarms in a subsequent recognition test and results in higher false alarms in a subsequent once-versus-twice test: that is, the SCAPE account of pause effect. (Note that SCAPE has no problem with the difference in Experiments 2A and 2B because both contain a stem and pause condition which creates surprise, leading to the observed conclusion.)

We do not argue that signal detection and global matching accounts are not useful in making predictions about the coming-to-mind of ideas. They do make accurate claims about the source of mental content, which is dependent on prior experiences, or objective familiarity. However, they have no motivated mechanism for dealing with the effect of the pause during the study experience. That is because the effect of the pause appears to be mediated by the quality of the experience, rather than the frequency of experiences of the stimulus. We do not argue that such accounts cannot be modified to accommodate our observations; but we do argue that to do so, they would have to be substantially modified to take into account not only objective characteristics of the experience but also the subjective reactions that people experience in processing those characteristics.

Subjective Quality and the Availability Heuristic

The experiments presented here demonstrate that experiencing a perception of discrepancy in the original event need not inevitably increase later discrimination, as it did in Experiment 1A; rather, in this case, experiencing that perception in the original encounter can also bias subjects to having the belief that target words had been encountered twice. The principle governing the effect of a perception of discrepancy on later processing seems clear: It enhances processing of the relationship between the target stimulus and its context in the original event. That comes in handy during a basic old–new recognition test but is of little help in a once-versus-twice decision, whereby other characteristics of the stimulus, such as vividness, clarity, and ease of processing are more useful. On the assumption that multiple prior presentations augment such characteristics, influencing the subjective quality of the trace, the subjective quality can sensibly be used to infer multiple prior experiences. By enhancing encoding of the item–context relationship, the perception of discrepancy causes false claims about once-presented target words because the subjective quality of the trace is used as a heuristic to infer multiple prior experiences.

Because the aim of our article was to outline the specific perceptions involved in creating the memory trace (i.e., our paradigm is essentially a learning one), many readers may question the difference between what we are proposing occurs during the once-versus-twice rating and what occurs during other frequency ratings. More specifically, the processes that we outlined occurred during the test phase of the once-versus-twice judgment task (Experiments 2 and 3) may appear to be similar to the use of an *availability heuristic* (Tversky & Kahneman, 1973), which could involve the same process as what we are proposing. The availability heuristic is an account for the observation of a tendency for people to overestimate objective frequency of emotionally charged or vivid events. The mechanism for availability could be objective familiarity (Bearden & Wallsten, 2004; Dougherty, Gettys, & Ogden, 1999; Hintzman, 1988), ease of retrieval (e.g., Schwarz, et. al., 1991), or simply an estimate based on the number of instances retrieved (Tversky & Kahneman, 1973).

We believe that the availability heuristic process stems from the use of the subjective quality of the trace, cued by the prior experience of the learning process (and the corresponding perceptions that were involved), and relies on the cognitive processes of production and evaluation. Thus, our account of the availability heuristic is distinct from recent accounts of availability that adequately explain the production function but fail to account for the evaluative component (for a notable exception, see Dougherty & Franco-Watkins, 2003, for a discussion on the role of source monitoring framework in frequency judgments). We argue that the foundation of the availability heuristic, too, is the perception of discrepancy.

We suggest, however, that the specific use of the subjective quality as a heuristic is distinct from the availability heuristic because of the specific task that subjects are required to carry out. The former involves a once-versus-twice task, the latter involves a frequency task. In the experiments presented here, subjects were not required to consciously count the frequencies to conclude multiple prior experiences, and we do not believe they did so spontaneously (otherwise we would not observe the illusion in

Experiments 2 and 3). Thus, they arrived at the conclusion of multiple prior experiences solely on the basis of the subjective quality of the trace.

Conclusion

The perception of discrepancy is probably a fairly rare experience. Certainly, the occurrence of feelings of familiarity, which is thought to be a consequence of that perception, is fairly rare: Most people do not experience a strong subjective feeling of having prior experience more often than once a day or week. That perception may be of particular importance in understanding certain puzzling phenomena that have to do with surprise at the time of encoding, such as flashbulb memory, or surprise at the time of remembering, such as the effects of being reminded of something by someone else, after a series of failures. It may also be important in understanding effects of novelty and surprise in aesthetic or emotional reactions. However, probably the greatest importance of understanding how that perception affects remembering decisions is its usefulness in revealing the interplay between cognitive and metacognitive processing, between the encoding and reproduction of mental contents and the phenomenology that both directs and follows from those acts.

References

Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 583–612). New York: Academic Press.

Bearden, J. N., & Wallsten, T. S. (2004). MINERVA-DM and subadditive frequency judgments. *Journal of Behavioral Decision Making, 17,* 349–363.

Bernstein, D. M., Whittlesea, B. W. A., & Loftus, E. F. (2002). Increasing confidence in remote autobiographical memory and general knowledge: Extensions of the revelation effect. *Memory & Cognition, 30,* 432–438.

Bodner, G. E., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory and Language, 48,* 563–580.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24,* 523–533.

Dougherty, M. R. P., & Franco-Watkins, A. M. (2003). Reducing bias in frequency judgment by improving source monitoring. *Acta Psychologica, 113,* 23–44.

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. (1999). MINERVA-DM: A memory process model for judgments of likelihood. *Psychological Review, 106,* 180–209.

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review, 89,* 627–661.

Gardiner, J. M., & Conway, M. A. (1999). Levels of awareness and varieties of experience. In B. H. Challis & B. M. Velichkovsky (Eds.), *Stratification in cognition and consciousness: Advances in consciousness research* (pp. 237–254). Amsterdam, the Netherlands: John Benjamins.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Glanzer, M., Kim, K., & Adams, J. (1998). Response distribution as an explanation of the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 633–644.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple trace model. *Psychological Review, 95,* 528–551.

Hintzman, D. L. (2004). Judgment of frequency versus recognition confidence: Repetition and recursive reminding. *Memory & Cognition, 32,* 336–350.

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review, 96,* 208–233.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513–541.

Kronlund, A., & Bernstein, D. M. (2006). Unscrambling words increases brand name recognition and preference. *Applied Cognitive Psychology, 20,* 681–687.

Kronlund, A., & Whittlesea, B. W. A. (2005). Seeing double: Levels of processing can cause false memory. *Canadian Journal of Experimental Psychology, 59,* 11–16.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252–271.

Miller, M. B., & Wolford, G. L. (1999). Theoretical commentary: The role of criterion shift in false memory. *Psychological Review, 106,* 398–405.

Murdock, B., Smith, D., & Bai, J. (2000). Judgments of frequency and recency in a distributed memory model. *Journal of Mathematical Psychology, 45,* 564–602.

Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 1–17.

Schwarz, N., Bless, H., Strack, F. Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61,* 195–202.

Shiffrin, R. M. (2003). Modeling memory and perception. *Cognitive Science, 27,* 341–378.

Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 179–195.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM-retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 592–604.

Tulving, E. (1985). How many memory systems are there? *American Psychologist, 40,* 385–398.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5,* 207–232.

Whittlesea, B. W. A. (1997). Production, evaluation and preservation of experience: Constructive processing in remembering and performance tasks. In D. L. Medin (Ed.) *The psychology of learning and motivation* (Vol. 37, pp. 211–264). New York: Academic Press.

Whittlesea, B. W. A. (2002a). False memory and the discrepancy attribution hypothesis: The prototype familiarity illusion. *Journal of Experimental Psychology: General, 131,* 96–115.

Whittlesea, B. W. A. (2002b). Two routes to remembering (and another to remembering NOT). *Journal of Experimental Psychology: General, 131,* 325–348.

Whittlesea, B. W. A. (2004). The perception of integrality: Remembering through the validation of expectation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 891–908.

Whittlesea, B. W. A., & Koriat, A. (2006). *Remembering under the influence of expectations: Subjective monitoring of the effects of learning.* Manuscript submitted for publication.

Whittlesea, B. W. A., Kronlund, A., Joordens, S., & Hockley, W. E. (2005). *The mirror effect in perspective: The source of feelings of familiarity.* Manuscript submitted for publication.

Whittlesea, B. W. A., & Leboe, J. P. (2000). The heuristic basis of

remembering and classification: Fluency, generation, and resemblance. *Journal of Experimental Psychology: General, 129,* 84–106.

Whittlesea, B. W. A., & Leboe, J. P. (2003). Two fluency heuristics (and how to tell them apart). *Journal of Memory and Language, 49,* 62–79.

Whittlesea, B. W. A., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? The unexpected basis of feelings of familiarity. *Acta Psychologica, 98,* 141–166.

Whittlesea, B. W. A., & Williams, L. D. (2000). The source of feelings of familiarity: The discrepancy attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 547–565.

Whittlesea, B. W. A., & Williams, L. D. (2001a). The discrepancy attribution hypothesis: I. The heuristic basis of feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 3–13.

Whittlesea, B. W. A., & Williams, L. D. (2001b). The discrepancy attribution hypothesis: II. Expectation, uncertainty, surprise and feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 14–33.

Wickens, T. D., & Hirshman, E. (2000). False memories and statistical design theory: Comment on Miller and Wolford (1999) and Roediger and McDermott (1999). *Psychological Review, 107,* 377–383.

Wixted, J. T., & Stretch, V. (2000). The case against a criterion-shift account of false memory. *Psychological Review, 107,* 368–376.

Zacks, R. T., & Hasher, L. (2002). Frequency processing: A twenty five year perspective. In P. Sedlmeier and T. Betsch (Eds.), *Frequency processing and cognition* (pp. 21–36). New York: Oxford University Press.