ELSEVIER

Acta Psychologica 98 (1998) 183–210

acta psychologica

Conceptually driven encoding episodes create perceptual misattributions

Michael E.J. Masson *, Judy I. Caldwell

Department of Psychology, University of Victoria, P.O. Box 3050, Victoria, BC, Canada V8W 3P5

Abstract

Processing fluency caused by prior encoding of a word is shown to increase duration judgments about that word and to decrease brightness contrast judgments about its mask when the word is presented in a masked word identification task. These effects occurred following an encoding task that involved visual perception of the words (reading aloud) and a task that provided no direct visual experience (generation from a semantic cue). Analysis of judgments conditionalized on correct or failed identification of target words indicated that judgments were powerfully affected by successful identification. Subjective estimates of the proportion of targets that were previously studied suggested that awareness of prior occurrence followed as an attribution based on fluent word identification, rather than acting as a causal agent for identification or altered perceptual judgments. We conclude that prior perceptual and conceptual encoding episodes can contribute to fluent processing of target words on a subsequent masked word identification task and that, regardless of its source, this fluency is experienced in a generic form that is susceptible to attribution to various causes, including prior experience (creating a sense of recollection) and current stimulus conditions.

1. Introduction

Attributional processes, long an integral concept in social psychology, have been accepted as an important concept in cognition, due in large part to the work of Jacoby and his colleagues (e.g., Jacoby and Brooks, 1984; Jacoby and Dallas, 1981; Jacoby et al., 1989b; Kelley and Jacoby, 1990). In the target article for this special

* Corresponding author. Tel.: +1 250 721-7536; fax: +1 250 721-8929; e-mail: mmasson@uvic.ca.

0001-6918/98/$19.00
PII S 0 0 0 1 - 6 9 1 8 (9 8) 0 0 0 4 2 - 5

issue, Kelley and Jacoby (1998) review, among other things, the idea that subjective experience is derived from attributional processes that follow experienced variations in performance. An important source of variation in task performance is prior experience. In word identification tasks, for example, prior encoding episodes can change the fluency with which a word is subsequently identified. An experience of fluent processing of a word can give rise to a feeling of familiarity or remembering through a process of attribution, in which fluency is attributed to prior occurrence.

Early evidence for the relation between the fluent processing of a stimulus and a tendency to classify that stimulus as an item that had been encountered previously was correlational in nature. That is, items that were more fluently identified or read during a recognition memory test were more likely to be classified as old. This result generally holds both for studied and for nonstudied items. Masson (1984) found an effect of this kind with typographically transformed sentences that were read aloud then subjected to a recognition memory decision, and Johnston et al. (1985) obtained a similar result with single word or nonword targets that were obscured by a visual mask during the identification and recognition test.

The idea that remembering can arise from attributional processes has gained even stronger support from a series of ingenious demonstrations in which a false sense of prior occurrence, or an illusion of remembering, is generated by directly manipulating the quality of a test stimulus. For example, Jacoby and Whitehouse (1989) used masked repetition primes, presented immediately prior to target words on a recognition memory test, to increase the likelihood that subjects would classify a target item as old. This effect was found even for new target words that had not been studied earlier. The masked prime apparently increased the fluency with which subjects encoded the subsequent target word, leading subjects to attribute the fluency to prior occurrence. When the primes were presented for a longer duration, enabling subjects to identify them, subjects no longer showed the memory illusion. Instead, the fluency was correctly attributed to the presentation of the prime rather than to presentation of the target in the study phase.

Similarly, Whittlesea et al. (1990) created an illusion of remembering by manipulating the perceptual quality of the stimuli. In those experiments, subjects were shown word lists using rapid serial visual presentation (RSVP). Each list was followed by a probe word and the task was to decide whether the probe word appeared in the list. Each probe word was covered by a mask consisting of a random dot pattern that took on one of two possible density levels. The variation in density was not apparent to subjects. When a lower density was used for the mask, the probe word was more easily seen and subjects were more likely to claim that it had appeared in the list than when a higher density mask was used. In a second experiment, subjects were made aware of the manipulation of mask density. Under these conditions the effect of density on recognition decisions was eliminated, leading Whittlesea et al. to conclude that awareness of the mask manipulation caused subjects to attribute variation in fluency of processing the probe to mask characteristics rather than to prior presentation of the probe. In another demonstration, Lindsay and Kelley (1996) presented recognition probes as word fragments that had to be completed prior to making a recognition decision. The fragments varied in their relative ease of completion. It was found that targets whose fragments were more easily completed were more likely to be classified as "old" in the recognition decision.

2. Influence of prior experience on stimulus judgments

The logic of creating illusions of remembering from variations in characteristics of stimulus presentation can be reversed to create illusions of perceptual experience arising from prior exposure to stimuli. In this case, processing fluency is generated by prior experience, but is mistakenly attributed to current processing circumstances or a special stimulus feature. Behind this strategy is the Whittlesea et al. (1990) suggestion that the attribution of fluency depends upon the subject's goals. If processing is fluent and subjects are instructed to make memory decisions, then the fluency will tend to feel like remembering, regardless of its source. If subjects are instructed to judge some characteristic of the stimulus, however, then processing fluency can feel like effortless perception, even if this fluency arises from previous experience.

Whittlesea et al. (1990) demonstrated that the fluency associated with prior exposure can be incorrectly attributed to characteristics of the current task. Once again, they presented a list of words in RSVP and then a probe word that was covered by a mask. A word in the RSVP stream was also covered by a mask, and the subject's task was to decide whether the mask covering the probe word was lighter than mask presented earlier in the list. On half of the trials, the probe word was a repetition of a word in the list and when it was a repetition, the repeated word in the list was covered by the mask. The results indicated that when the word was repeated, subjects were more likely to judge the probe's mask as less noisy than when the word was not repeated.

A number of other findings support the conclusion that prior exposure to a stimulus can influence judgments about its characteristics. Witherspoon and Allan (1985) found that the estimated duration of briefly presented target words was elevated for words that had been studied earlier, and Jacoby et al. (1988) showed that prior auditory presentation of sentences led to lower ratings of auditory noise when those sentences were later played against a background of white noise.

Not only can judgments of physical attributes be affected by experienced processing fluency, but classification judgments can be affected as well. First, Whittlesea (1993) showed that ratings of a probe word's relatedness in meaning to an earlier item presented in an RSVP list was affected by varying the density of the visual mask covering the probe. A lower density mask induced greater processing fluency, and subjects were more likely to classify the probe as semantically related to a word that had appeared in the list. Second, ratings of an item's pleasantness are enhanced by prior exposure (e.g., Kunst-Wilson and Zajonc, 1980; Mandler et al., 1987; Seamon et al., 1984; Whittlesea, 1993; Zajonc, 1980). Third, Jacoby and his colleagues (Jacoby et al., 1989a; Jacoby et al., 1989c) found that the fluency associated with prior exposure to nonfamous names can lead subjects to classify those names as famous. This attribution does not involve ascribing fluency to prior exposure to the names in the context of the experiment, but rather to some nonexistent, preexperimental experience with the names. Finally, judged difficulty of solving an anagram is reduced by prior exposure to the anagram's solution word (Kelley and Jacoby, 1996).

2.1. Fluency without recollection

There is substantial evidence that the experience of fluency and its attribution to a characteristic of the stimulus can occur in the absence of conscious recollection of prior exposure. Kunst-Wilson and Zajonc (1980), Mandler et al. (1987), and Seamon et al. (1984) found that prior exposure enhanced preference ratings for items even though recognition memory for those items was at chance. Jacoby et al. (1989a, c) ruled out conscious recollection as a factor in producing enhanced fame judgments by informing subjects that all names they had read in a study phase were nonfamous. If subjects recollected reading a name in the study phase of the experiment, they could immediately conclude that it was nonfamous. By dividing attention or imposing a delay between the study phase and the test phase, Jacoby and his colleagues were able to reduce conscious recollection. Under these conditions, the processing fluency subjects experienced when shown previously studied names often was misattributed to fame; subjects were unable to use conscious recollection to block that misattribution. The effect of prior exposure on fame judgments was prevented by allowing subjects to use full attention or by testing them immediately after the study phase. These conditions enabled reliable conscious recollection of prior occurrence, leading subjects to classify nonfamous names correctly.

Further evidence that processing fluency can affect stimulus judgments without conscious recollection of prior presentation was provided by Paller et al. (1991). They tested amnesic subjects and age-matched control subjects in a duration judgment task. Like Witherspoon and Allan (1985), subjects judged the exposure duration of briefly presented target words. Duration judgments were longer for words that had been read earlier, and this effect was equally strong for amnesic and control subjects, even though amnesic subjects were severely impaired on a recognition memory test of the target words.

2.2. Common ground with indirect tests of memory

The finding that fluency arising from prior exposure can be misattributed to some other source in the absence of conscious recollection of that prior exposure provides an important link with studies of direct and indirect tests of memory. Performance on indirect tests of memory, such as word stem completion and masked word identification tasks, is affected by unconscious influences of memory. Although there is evidence that indirect tests of memory might also be influenced by conscious recollection (e.g., Jacoby et al., 1993), it seems probable that the processing fluency associated with the misattribution effects reviewed above is the signature of the same unconscious influence of memory that underlies enhanced performance on indirect tests of memory. On this view, one would expect factors that constrain performance on indirect tests of memory to exert similar control over the processing fluency that leads to misattribution effects.

One of the most powerful constraints on the effect of prior encoding experience on indirect tests of memory is the similarity between processing operations applied during encoding and operations required by the indirect test (e.g., Blaxton, 1989; Jacoby, 1983; Roediger and Blaxton, 1987; Weldon, 1991). Indirect tests of memory that are presumed to involve primarily perceptual processes, such as word stem completion and masked word identification, have typically shown repetition priming effects when prior encoding episodes involve perceptual processing (reading target words) but when the prior encoding episode invokes conceptual rather than perceptual processing (generation of a target word from a semantic cue), these effects are reduced or eliminated (Jacoby, 1983; Toth et al., 1994; Weldon, 1991).

An exception to this pattern has been reported by Masson and MacLeod (1992) in the case of the masked word identification task. In their experiments, various kinds of semantic generation cues used in a study phase led to enhanced masked word identification that was similar to the amount of enhancement produced by reading clearly presented target words during the study phase. These results strongly suggest that direct perceptual experience is not required for the masked word identification task to yield a substantial repetition priming effect. MacLeod and Masson (1997) subsequently verified that the word fragment completion task is unlike the masked word identification task in this regard. For that task, generation from semantic cues did not produce as much enhancement as reading the target words. They proposed that in the masked word identification task, the availability of the entire target word, albeit for a brief time, enabled the recruitment of prior study episodes that were conceptually related to the target. In the word fragment completion task, and probably the word stem completion task as well, presentation of an incomplete word is less likely to make contact with prior episodes that involved conceptual processing because the available graphemic information does not provide adequate constraints to specify a particular conceptual interpretation of the test stimulus.

On the account advocated by Masson and MacLeod (1992, 1997) and MacLeod and Masson (1997), word identification processes engaged by the masked word identification task involve both conceptual and perceptual knowledge. Prior episodes involving a target word in either of these two domains of processing can be recruited by a brief presentation of the target word and can contribute to fluent processing or identification of that word.

The claim that earlier conceptual processing can enhance the identification of a target word presented under difficult perceptual conditions suggests a certain plasticity in the processing fluency that arises from prior exposure. Indeed, there is evidence for just this kind of plasticity in misattribution paradigms. Whittlesea (1993) had subjects judge whether a probe word presented after an RSVP list of words had appeared in that list. When tested, however, the probe word was presented in a sentence context that was either predictive of the word's identity or nonpredictive (e.g., The anxious student wrote a TEST vs. Later in the afternoon she took a TEST). The conceptually driven fluency created by the predictive sentence context increased the probability of claiming that the probe had appeared in the RSVP word list. Luo (1993) obtained a similar result, in which a positive recognition memory decision was more likely if the subject performed a processing task on the target (e.g., providing a rhyming word or typing the word's letters in reverse order) just prior to making the decision. Thus, subjects appear unable to distinguish reliably between fluency generated by prior episodic experience, by current perceptual experience, or by current conceptual experience. Attributions flowing from each of these classes of experience appear to be malleable and directed by the subject's current goals.

2.3. Generic processing fluency

The malleability of attributions regarding experienced processing fluency is reminiscent of the demonstration by Schachter and Singer (1962) and its review by Kelley and Jacoby (1998) of varying attributions arising from a generic physiological activation created by an injection of adrenaline and directed by cues in the social environment. By analogy, we suggest that a generic sense of processing fluency can arise either from prior encoding episodes or conditions of current processing of a target. We propose that generic processing fluency can be attributed to prior occurrence, to confidence in the target's identity (the basic repetition priming effect), or to aspects of the stimulus presentation, depending on the subject's goals (see also Kelley and Jacoby, 1990; Whittlesea, 1997).

The experiments reported here test the proposition that prior perceptual and conceptual processing episodes can have similar effects on attributions of fluency, even when those attributions are directed toward immediate perceptual experience. The general procedure we used was similar to that used by Witherspoon and Allan (1985). In the encoding phase of our experiments, subjects were asked to read words aloud (a perceptual encoding experience) or to generate them from semantic cues (a conceptual encoding experience). Subjects then were presented with a series of training trials involving filler items in which a word was presented briefly and followed by a pattern mask. Target duration (Experiments 1 and 2) or brightness contrast of the mask (Experiment 3) was varied and the task was to judge this varying attribute of the display. Finally, a series of test trials similar to the training trials was presented, including critical items that had been presented in one of the encoding phase tasks or that were new.

We reasoned that if a conceptually driven encoding task is capable of producing fluent processing on the masked word identification task, as claimed by Masson and MacLeod (1992, 1997) and MacLeod and Masson (1997), and if that fluency is experienced as qualitatively similar to the fluency produced by a prior perceptual encoding of a target word, then the generate and read encoding tasks should have similar effects on stimulus judgments. Specifically, we predicted that judgments of target duration should be elevated among studied words and that judgments of pattern mask contrast should be reduced for studied words, regardless of whether the study experience involved the generate or the read task. These perceptual illusions are assumed to derive from a generic processing fluency that can arise from either prior perceptual or prior conceptual experience with a target (e.g., Luo, 1993; Whittlesea, 1993).

3. Experiment 1

In the encoding phase of Experiment 1, subjects read critical words or generated them from a cue that provided semantic information along with the first letter of the target. A training phase then followed in which filler words were presented for one of three brief durations, followed by a mask. Subjects were required to judge the duration of each word, classifying the duration as short, medium, or long. The training phase moved directly into a test phase in which a mixture of critical and filler words were presented at one of the two shorter durations used during training trials. Subjects were not informed of this transition and presumably continued to believe that three different durations were in effect. Critical items used in the test phase had either been encoded in one of the two encoding tasks (generate or read) or were new (i.e., had not been presented in the encoding phase).

The longest of the three durations was not used during the test phase. We expected that without continuing experience with that duration, subjects might be more likely to misclassify a shorter duration presentation into the longest duration category. It was misclassification of this kind in which we were most interested. The influence of memory for a prior encoding episode was expected to induce subjects to experience an exposure duration involving a repeated target as longer than the actual duration. By excluding the longest duration from the test phase, we hoped to avoid exposing subjects to the contrast between apparent and truly long durations. It was thought that experience of this kind might lead subjects to be more cautious in their evaluation of their experience, thereby reducing the effect we were trying to produce.

Earlier work using similar methods has shown that prior exposure can create perceptual misattributions of this kind, but those demonstrations have been restricted to encoding tasks that involved direct perceptual experience with target items (e.g., Jacoby et al., 1988; Witherspoon and Allan, 1985). The critical question here was whether an encoding task that did not involve such experience (generation from a semantic cue) would produce a similar result.

3.1. Method

3.1.1. Subjects

Thirty undergraduate students at the University of Victoria participated in the experiment for extra credit in an introductory psychology course. Their ages ranged from 18 to 29 years, with a median age of 18 years. All subjects were native speakers of English.

3.1.2. Materials and design

The materials consisted of 90 critical words ranging from four to eleven letters in length. A generation cue was created for each critical word. This cue consisted of a short phrase and the first letter of the target word. For example, the cue for the target word *arrow* was *an archer shoots a bow and – a*. The 90 critical items were divided into three blocks of 30 and each block was assigned to one of three encoding conditions, read, generate, or new. The assignment of blocks to conditions was counterbalanced across subjects. Eight additional items were included as practice items in the encoding phase of the experiment. There were also an additional 81 filler words used in the experiment, ranging in length from four to eight letters. Forty-five of these words were used in a training phase in which subjects were given experience with three different exposure durations in the masked word identification task. Six filler items were used as practice items at the beginning of the subsequent test phase, and the remaining 30 fillers were used in that test phase.

A 2×3 repeated measures design was used, with encoding condition (read, generate, and new) and presentation duration at test (short or long) as factors. Half of the 30 critical items assigned to each encoding condition were tested in each of the two exposure durations used in the test phase.

3.1.3. Procedure

Subjects were tested individually in the presence of an experimenter. A Macintosh II computer with two monochrome monitors was used to present the stimuli and record responses. Software was written to synchronize presentation of stimuli with the monitor's raster scan cycle of 15 ms, allowing precise control over stimulus presentation duration in 15-ms increments. One monitor presented the instructions and stimuli to the subjects and the second monitor was used to present information about the correct response on each trial to the experimenter so that coding of the correctness of the subject's oral responses could be done as each trial was conducted. A button box was also used by the subject for the purpose of making rating responses.

The experiment began with the presentation of a general set of instructions, followed by the instructions for the encoding phase of the experiment. The encoding phase began with the presentation of eight randomly ordered practice items (four read and four generate trials). The practice trials were followed immediately by a randomly ordered presentation of 30 critical read and 30 critical generate trials. For the read trials, the critical word was presented in lowercase letters at the center of the subject's monitor and remained on the screen until the subject read it aloud. In the generate condition, the cue for the critical word was presented at the center of the subject's monitor and the subject was to say aloud the word that fit the cue. If the subject provided the wrong word or was unable to respond to the cue, the experimenter told the subject the correct word.

After the encoding phase, the instructions for the training and test phases were displayed on the subject's monitor. For this part of the experiment, each trial involved the brief presentation of a target word followed by a pattern mask. The subject's task was to classify the word's duration as short, medium, or long by pressing one of three keys on the button box. The training phase consisted of 45 trials involving filler words. Fifteen filler words were presented in each of three different durations: 30, 60, and 90 ms. The presentation of these trials was randomly ordered. The purpose of the training phase was to acquaint the subjects with the three different exposure durations and to give them practice at classifying these durations as short, medium, or long.

Test trials were presented immediately after the training trials. From the subject's perspective, the test trials were simply a continuation of the training trials. The test trials began with a set of six practice items, followed by a randomly ordered presentation of 90 critical trials (30 read, 30 generate, and 30 new) and 30 filler trials. Three of the practice items were displayed for 30 ms and three for 60 ms. Similarly, half of the critical trials in each encoding condition and half of the filler items were presented at each of these two durations. Filler trials were included so that half of the items presented at each duration would be old (read or generate) and half would be new (new critical words or filler words). Although only the two shorter durations were used in the test phase, subjects were not made aware of this fact and they continued to respond using the three available duration categories.

Each trial in the training and test phases of the experiment began with the subject pressing a button to initiate the trial. After this button press, a fixation cross appeared at the center of the monitor for 255 ms, followed by a blank screen for 255 ms. The word was then presented at the center of the monitor for the appropriate duration and was followed immediately by a 14-character ampersand mask, which remained on the screen until the subject responded. Subjects indicated their response by pressing one of three buttons on the button box which corresponded to short, medium, and long duration options. Subjects were given a short rest break every 26 trials.

3.2. Results

In the study phase, the mean proportion of targets in the generate encoding condition that were correctly generated was 0.98. Given the high success rate in generating targets assigned to the generate encoding condition and the fact that subjects were informed of the correct target when they failed to generate it, all generate items were included in the analyses reported here, regardless of whether the subject correctly generated it during the encoding phase. This approach was taken in all of the experiments reported here.

The performance measure of interest in Experiment 1 was the average duration rating given to critical words in the test phase, on a scale from one to three. Rating values of 1, 2, and 3 were assigned to the short, medium, and long categories, respectively. For each subject, the average duration rating for critical words was computed separately for each of the three encoding conditions. The mean ratings are shown in Table 1.

The ratings were analyzed using a repeated measures analysis of variance (ANOVA) with encoding condition (read, generate, and new) and exposure duration (30

Table 1
Mean duration ratings as a function of prior encoding and actual exposure duration in experiment 1

Exposure duration (ms)	Encoding condition		
	Generate	Read	New
30	1.55	1.60	1.46
	(0.38)	(0.33)	(0.30)
60	2.27	2.25	2.25
	(0.27)	(0.26)	(0.24)

Standard deviations are shown in parentheses.

and 60 ms) as factors. The type I error rate used in this and all other analyses reported here was 0.05. The ANOVA revealed a reliable main effect of target duration, indicating that the longer exposure duration condition received higher duration ratings than the shorter duration (2.26 vs. 1.54), $F(1{,}29) = 85.68$, MSE = 0.274. In addition, there was a reliable effect of encoding condition, $F(2{,}58) = 3.55$, MSE = 0.026, supporting the prediction that a prior encoding experience would lead to higher exposure duration judgments (1.91, 1.93, and 1.85, for the generate, read, and new conditions, respectively). The interaction between these two factors was also reliable, $F(2{,}58) = 3.24$, MSE = 0.023, indicating that the effect of prior experience with a target item depended on the exposure duration used in the test phase.

Separate ANOVAs were conducted for the 30 and 60-ms conditions to examine the effect of prior encoding on duration judgments under each of the exposure durations. Although there was no effect of prior encoding in the 60-ms condition, $F < 1$, there was a reliable effect in the 30-ms condition, $F(2{,}58) = 5.24$, MSE = 0.031. Pairwise comparisons of means in the 30-ms condition, using the MSE from the omnibus test of those means, indicated that both the read and the generate encoding conditions differed from the new condition, $F(1{,}58) = 10.13$, and $F(1{,}58) = 4.43$, respectively. The read and generate conditions did not reliably differ from one another, $F < 1.2$.

3.3. Discussion

The most important result of Experiment 1 is the finding that not only a perceptual encoding task (read), but also a conceptually driven encoding task (generate), led to an apparent perceptual illusion in which the exposure duration of studied words appeared to be longer than for new words. This finding suggests that in the masked word identification task, words encoded either by read or generate encoding tasks produce a form of processing fluency that can be misinterpreted in the same way.

The misattribution effect was observed only when targets were presented for 30 ms, but not when a 60-ms exposure duration was used. It might be supposed that the longer duration condition did not produce the effect because in the test phase subjects did not experience any targets at the longest duration used during training (90 ms) and therefore rarely classified any presentations into the longest duration category. The pattern of classification responses contradict this notion: 34.5% of trials for which the actual duration was 60 ms were categorized as being of the longest duration. Thus, it appears that subjects continued to hold the impression that all three exposure durations were used during the test phase, even though only the shorter two durations were in effect.

Instead of reluctance to classify durations into the longest category, it may be that targets presented in the 60-ms duration condition were too easily identified, regardless of whether or not they had been presented in the study phase. In that case, subjects would not have experienced differential fluency across encoding conditions and no effect on duration ratings would be expected. Alternatively, the fact that subjects were not required to identify the target words might have played a role in the failure to find a prior encoding effect in the 60-ms condition. If subjects did not consistently attempt to identify the targets, they may not have experienced the fluency that is assumed to drive the rating effect. We note, however, that Witherspoon and Allan (1985) did not find a difference in duration rating effects as a function of whether subjects were instructed to attempt to identify the targets. Moreover, it is not clear why the absence of such an instruction would affect ratings only in the 60-ms condition and not in the 30-ms condition, both of which were experienced by the same subjects. We are unable to specify, then, exactly why the effect of prior encoding in Experiment 1 did not occur in both duration conditions.

4. Experiment 2

In Experiment 2, we attempted to produce a more robust misattribution effect by making three changes to the method used in Experiment 1. First, we required subjects to identify a target word prior to making a duration rating in the masked identification task. We reasoned that the requirement to identify target words would ensure that subjects would have frequent opportunities to experience the processing fluency associated with the target words, thereby setting the stage for the misattribution effect. In Experiment 1, it is possible that subjects did not systematically attempt to identify the target words and therefore were not consistently affected by differential processing fluency of studied and nonstudied words.

Second, we adopted an approach similar to that used by Witherspoon and Allan (1985) in setting up exposure durations for the training and test phases. Three exposure durations were used in the training phase: 15, 45, and 75 ms. In the test phase, all critical items were presented using a single exposure duration set midway between two adjacent durations used during training. Filler items were presented at the other duration that had been used in training. This approach created an ambiguous duration in the test phase: 30 ms for half of the subjects (midway between 15 and 45 ms) and 60 ms for the other half of the subjects (midway between 45 and 75 ms). For each critical item, then, the ambiguity of the exposure duration would have to be resolved by classifying the duration into a category that in training was either longer or shorter than the actual test duration. It was expected that this problem would often be resolved by rating the duration for studied targets as longer than the actual duration and by rating the duration for new targets as shorter than the actual duration, creating the misattribution effect.

Finally, the critical items used in Experiment 2 were low-frequency words. Earlier studies have shown that low-frequency words produce larger effects of prior exposure than do high-frequency words (e.g., Jacoby and Dallas, 1981; Scarborough et al., 1977). It was expected that by using words of low-frequency, larger effects of prior presentation would be experienced by subjects and therefore a more substantial influence of prior presentation on duration ratings would be found than if higher-frequency words were used.

In Experiment 2 we also had subjects respond to a questionnaire at the end of the test phase. Of particular interest was the subjects' assessment of the proportion of items in the test phase that they believed had been presented in the study phase of the experiment. We were interested in that estimate because it might be argued that the effect of prior presentation on duration ratings comes about through conscious recollection of studied words. In particular, one might suppose that conscious recollection of a word's occurrence in the study phase would contribute to its successful identification on the masked word identification test and also influence the duration rating. On this view, the causal chain begins with conscious recollection and ends with an elevated duration rating. We postulated a different causal chain, grounded on the proposal that conscious recollection can be constructed as an attribution following fluent processing (e.g., Jacoby and Dallas, 1981; Jacoby and Whitehouse, 1989; Johnston et al., 1985; Whittlesea, 1993, 1997). By that account, subjects would experience fluent processing of a studied word when presented in the masked word identification task, leading to a greater probability of identification and an attribution of fluency to longer exposure duration. The increased fluency might also be accompanied by an attribution of prior occurrence.

These two versions of the sequence of events involving processing fluency, duration rating, and conscious recollection lead to different predictions regarding subjects' estimates of the proportion of targets that were studied. If conscious recollection drives both identification and the duration rating, subjects' estimates of the proportion of targets that were studied should be close to, or somewhat lower than, the actual proportion. An underestimate would occur because not all studied items would be recollected and there is no reason to expect the rate of false recollection of new items to be higher than the rate of failed recollection of studied items. The net effect should be estimates that are close to or perhaps below the actual proportion of old items.

If conscious recollection surfaces as an attribution following experienced fluency, which drives both identification and duration ratings, a different outcome would be expected. In particular, the processing fluency created by the longer exposure duration might often be misattributed to prior occurrence, leading to false recollection of new items. This attribution process would leave subjects with the erroneous impression that a substantial number of new items had been studied earlier, leading to a systematic overestimate of the proportion of targets that had been studied.

4.1. Method

4.1.1. Subjects

Thirty subjects participated in Experiment 2 and were taken from the same pool as those in Experiment 1. Their ages ranged from 18 to 37 years, with a median age of 19 years. All subjects were native speakers of English. Fifteen subjects were randomly assigned to each of the two exposure duration conditions.

4.1.2. Materials and design

The critical set of materials used in Experiment 2 consisted of 75 low-frequency words (1 to 10 occurrences per million, Kučera and Francis, 1967). These words were four to six letters in length. As in Experiment 1, a generation cue was created for each critical word, consisting of a short phrase and the first letter of the target word. The 75 critical words were divided into three blocks of 25 items. One block of 25 items was assigned to each of the three encoding conditions, read, generate, or new. The assignment of blocks to conditions was counterbalanced across subjects. Eight additional items, four read and four generate, were selected for practice trials in the encoding phase. An additional 126 words, four to six letters in length were selected. Forty-five of these words were used in the training phase of the experiment, six served as practice items at the beginning of the test phase, and 75 were used as nonstudied filler items in the test phase.

The design was a 3×2 mixed factorial, with encoding condition (read, generate, and new) as a within-subject factor and critical exposure duration at test (30 and 60 ms) a between-subject factor.

4.1.3. Procedure

The equipment and general procedure was the same as that used in Experiment 1, so only those details that differ from the first experiment will be described. The three durations used in the training phase of this experiment were 15, 45, and 75 ms, with 15 items displayed at each of these three durations.

Two durations were used for items presented in the test phase. For the subjects assigned to the 30-ms condition, one duration was midway between the two shorter durations used in the training phase (i.e., 30 ms) and the other duration was the longest one used in training (75 ms). All 75 critical items were presented for 30 ms. There were also 25 fillers presented for 30 ms, creating a mixture of 50 studied and 50 nonstudied (25 new and 25 fillers) items. An additional 50 filler items were displayed for 75 ms. Thus, one third of the test trials used the longest duration, maintaining the same proportion as experienced during training trials. For the subjects in the 60-ms condition, one duration was the shortest used during training (15 ms) and the other was midway between the two longer training durations (i.e. 60 ms). The 75 critical items and 25 filler items were presented for 60 ms, and a further 50 fillers were displayed for 15 ms.

There were three additional procedural differences from Experiment 1. First, the mask was the same length as the target word, rather than a fixed length of 14 characters. Second, in the training and test phases, subjects were asked to identify the target word before making their duration judgments. Finally, at the end of the test phase, subjects were asked to explain the basis for their duration judgments and were asked to estimate the proportion of words in the duration judgment task that had been presented in the first part of the experiment (the encoding phase).

4.2. Results and discussion

The mean proportion of targets in the generate condition that were correctly generated by subjects during the encoding phase was 0.96.

4.2.1. Target identification

The mean proportions of critical targets that were correctly identified in the test phase are shown in Table 2. An ANOVA with encoding condition and exposure duration as factors indicated that there was a reliable main effect of exposure duration, $F(1,28) = 39.26$, $MSE = 0.072$, with greater accuracy in the 60-ms duration condition than in the 30-ms duration condition (0.93 vs. 0.57). The encoding task effect was also reliable, $F(2,56) = 12.08$, $MSE = 0.007$, indicating that prior encoding increased the probability of correct identification on the masked word identification task. The interaction between these two factors was reliable, $F(2,56) = 5.12$, $MSE = 0.007$, due to a ceiling effect on performance in the 60-ms duration condition.

Separate ANOVAs conducted for the 30- and 60-ms conditions supported this conclusion. In the 60-ms condition, there was no effect of encoding condition, $F < 1.3$, whereas in the 30-ms condition, the encoding task effect was reliable, $F(2,28) = 12.20$, $MSE = 0.009$. Pairwise comparisons in the 30-ms condition, using the MSE from the omnibus test of those means, indicated that both the read and the generate encoding conditions differed from the new condition, $F(1,28) = 23.04$, and $F(1,28) = 11.61$, respectively. The read and generate conditions did not differ reliably from one another, $F(1,28) = 1.94$, $p > 0.15$.

4.2.2. Duration ratings

Mean duration ratings for items in each encoding condition were computed as in Experiment 1. The mean ratings within each exposure duration condition are shown in Table 2 as a function of prior encoding task. Ratings are shown for all targets and conditionalized on correct or incorrect identification of the target. Separate analyses were conducted for all targets and for the conditionalized data.

An ANOVA was conducted on the ratings based on all targets with encoding task (read, generate, and new) and duration at test (30 and 60 ms) as factors. This analysis revealed a reliable main effect of exposure duration, demonstrating higher duration ratings in the 60-ms condition than in the 30-ms condition (2.22 vs. 1.53),

Table 2
Mean proportion of targets correctly identified and mean duration ratings as a function of prior encoding and exposure duration in experiment 2

Exposure duration and encoding task	Identification	Duration rating		
		All items	Correct ident.	Incorr. ident.
30 ms				
Generate	0.60 (0.22)	1.58 (0.33)	1.88 (0.33)	1.24 (0.23)
Read	0.64 (0.21)	1.56 (0.28)	1.78 (0.32)	1.25 (0.30)
New	0.48 (0.24)	1.44 (0.23)	1.76 (0.25)	1.22 (0.18)
60 ms				
Generate	0.94 (0.06)	2.25 (0.22)	2.27 (0.22)	1.65 (0.50)
Read	0.94 (0.05)	2.24 (0.18)	2.26 (0.19)	2.07 (0.44)
New	0.91 (0.11)	2.18 (0.17)	2.20 (0.17)	1.85 (0.39)

Standard deviations are shown in parentheses. Correct ident. ratings are based on items that were correctly identified; Incorr. ident. ratings are based on items that were not correctly identified.

$F(1,28) = 72.29$, MSE = 0.150. The main effect of encoding condition was also reliable, indicating that duration ratings were affected by prior exposure to the words (1.92, 1.90, and 1.81, for generate, read, and new, respectively), $F(2,56) = 7.78$, MSE = 0.012. The interaction between these two factors was not reliable, suggesting that the pattern of duration ratings was similar across the two duration conditions, $F < 1$.

Collapsing across exposure duration, pairwise comparisons using the MSE from the omnibus test of encoding task indicated reliably higher mean ratings in both the read and the generate conditions than in the new condition, $F(1,56) = 9.99$, and $F(1,56) = 13.14$, respectively. There was no difference between read and generate encoding conditions, $F < 1$. These results are consistent with the predictions of higher duration ratings for studied words than for new words and of similar effects for the read and generate encoding conditions.

4.2.3. *Conditionalized duration ratings*

The fact that subjects attempted to identify the targets permitted us to examine duration ratings conditionalized on correct or on incorrect identification. The mean conditionalized ratings are shown in Table 2. Because of the high proportion of correctly identified targets achieved by subjects in the 60-ms condition, there were substantial missing data for duration ratings conditionalized on incorrect identification. Only eight subjects contributed data to all three of the encoding task conditions on this measure. Moreover, the mean ratings produced by these subjects were based on a small number of observations. Therefore, duration ratings conditionalized on incorrect identification in the 60-ms condition were not analyzed further because of concerns about the reliability of the data. Instead, the data from the 30-ms condition were analyzed in an ANOVA with encoding task and identification correctness as factors. The major question addressed by this ANOVA was whether identification success was significantly associated with higher duration ratings. The rating data conditionalized on correct identification in the 60-ms condition were analyzed using a pair of orthogonal contrasts to test the hypothesis that even among identified targets, subjects would experience greater fluency for studied targets.

The ANOVA applied to the 30-ms duration condition excluded the data from one subject who failed to identify any items in the new condition and therefore had missing data in the cell defined by correctly identified new items. This ANOVA revealed only a main effect of identification outcome, as duration ratings were higher following correct identification than following incorrect identification (1.80 vs. 1.24), $F(1,13) = 44.25$, MSE = 0.152. Neither the main effect of encoding task, $F(2,26) = 2.14$, MSE = 0.017, $p > 0.10$, nor the interaction, $F < 1$, were significant. This result clearly demonstrates that subjects were inclined to rate the target word's duration as longer if they successfully identified it. The lack of a reliable effect of encoding task means that once the outcome of the identification attempt is factored out, no reliable variance is explained by prior presentation.

For the 60-ms condition, ratings conditionalized on correct identification included the vast majority of critical trials because correct identification occurred so frequently (0.93 of the trials). This set of conditionalized data, therefore, was relatively stable and offered an opportunity to test the possibility that subjects were sensitive to differences in the fluency with which identification was achieved. It was expected that identification of studied targets, either from the generate or read condition, would be more fluent than identification of new targets. This difference in fluency might be realized as a difference in duration ratings, even though only correctly identified targets are considered. This hypothesis was tested using a pair of orthogonal contrasts. In the first contrast, duration ratings for the generate and read conditions were compared and were found not to differ reliably, $F < 1$. In the second contrast, to maximize power, the mean duration rating averaged across the generate and read conditions was compared to the mean rating in the new condition. This contrast showed that the average rating of the two studied conditions was significantly higher than the rating for the new condition (2.27 vs. 2.20), $F(1,14) = 6.27$, $\text{MSE} = 0.005$. The latter result is consistent with the claim that subjects' ratings were affected by subtle differences in the ease with which targets were identified, not just the success or failure of that identification.

4.2.4. Subjective report

When asked at the end of the test phase to explain the basis for their duration judgments, all but one subject stated that their duration ratings were based on how easily they could see the word. One subject in the 60-ms condition reported a strategy that involved mental counting to determine the duration.

4.2.5. Estimated proportion of studied items

To provide evidence that the higher duration ratings for old words were the product of fluency rather than of conscious recollection, we examined subjects' estimates of the proportion of items that were old. These estimates were compared to estimates we computed, based on items the subjects correctly identified. This comparison was made by constructing a scatter plot in which each subject's estimate of the proportion of items that were old was plotted against the proportion of identified items that were old. The resulting scatter plot can be see in Fig. 1.

The actual proportion of items that were old were computed for each subject by finding the proportion of correctly identified items in the test phase (critical and filler items) that were old. This conditionalized proportion was used because we assumed that the subjects' estimates of the proportion of items that were old would be determined primarily by their experience with correctly identified items. Because we did not keep track of identification accuracy during the training phase (in which all items were new), our computation of the proportion of identified items that were old was based only on targets presented in the test phase. This constraint on our computation means that the values on the horizontal axis in Fig. 1 somewhat overestimate the true proportion of identified items that were old. This point is noteworthy because the scatter plot in Fig. 1 clearly shows a strong tendency for subjects to overestimate the proportion of items that were old. If items from the training phase were to be taken into account, the degree of subjects' overestimation would be even greater than that seen in Fig. 1. We can be confident, therefore, that subjects generally overestimated the proportion of targets that were old.

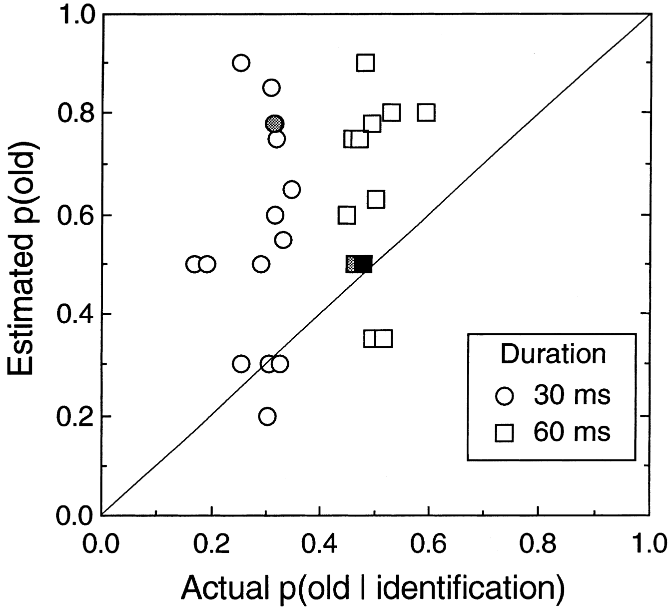

Fig. 1. Subjects' estimates of proportion of targets that were old (previously studied) plotted against actual proportion of correctly identified targets in the test phase that were old in Experiment 2. Gray and black symbols indicate that two or three subjects, respectively, had identical or nearly identical points in the plot. The diagonal line represents perfect estimation.

Although the distribution of estimates provided by subjects was similar in the two duration conditions (median estimates were 0.55 and 0.60 in the 30- and 60-ms conditions, respectively), the two distributions are separated in the scatter plot in Fig. 1 because of differences in the actual proportion of identified items that were old. For subjects in the 60-ms condition, a greater proportion of identified items were old because the old items were tested at the longer duration (60 ms), whereas in the 30-ms condition, the old items were tested at the shorter duration and were therefore less likely to be identified. Despite this systematic difference between the two groups in the actual proportion of old items among those that were identified, there was no apparent systematic difference in the subjects' estimates of proportion of items that were old. The absence of a systematic difference in subjective estimates suggests that subjects' sense of having encountered target items previously in the experiment was strongly affected by something other than actual prior presentation.

The strong trend toward overestimation of the proportion of old items is consistent with the proposition that subjects experienced conscious recollection as an attribution of processing fluency to prior occurrence. Processing fluency could be created by having experienced an item in the study phase, by using a longer exposure duration in the test phase, or by some idiosyncratic aspect of the target that made it relatively easy to identify. It appears that subjects were liable to make the mistake of attributing occasions of fluent processing of new items to prior occurrence, thereby overestimating the proportion of targets that were old. Had subjects used conscious recollection, driven by factors other than attribution of processing fluency, as a basis for identification and duration ratings, we should have observed either a systematic underestimation of the proportion of items that were old, or no trend for either over- or underestimation. An underestimation would have occurred in case not all old items were recollected.

The systematic overestimation of proportion of items that were old, coupled with the Paller et al. (1991) result that the effect of prior exposure on duration ratings is preserved among amnesic subjects, suggests that the effect of the study phase experience on subsequent duration ratings was not caused by conscious recollection of prior occurrence of old target items. Rather, recollection appears to have resulted from an attribution that followed successful identification of the target and an assessment of the ease or fluency of that identification.

5. Experiment 3

In the final experiment, the objective was to replicate and extend the finding that both read and generate encoding tasks lead to similar experiences of processing fluency during subsequent presentation of target words. Rather than manipulating exposure duration of the target words, a different perceptual quality was varied: the brightness contrast of the string of ampersands that served as the mask on masked word identification trials. A related manipulation was used by Whittlesea (1993), Whittlesea et al. (1990) and by Merikle and Reingold (1991), who varied the density of a random dot mask that was superimposed upon a target word. These studies showed that variations in the density of the mask affected subject's judgments about the target word and that prior presentation of the target word affected judgments about the density of the mask. We expected that similar effects would be apparent when the brightness contrast of the mask characters was varied.

In Experiment 3, subjects began with a study phase identical to that used in Experiment 2, followed by a training and a test phase similar to those used in Experiment 2. During masked word identification trials in the training phase, the brightness contrast of the mask was set to one of three possible values. With the lowest contrast value, the mask was least effective at obscuring the target word, whereas the mask was most effective when its contrast value was at the highest setting. The perceptibility of target words, then, was manipulated by altering the appearance of the mask.

In the test phase, only two contrast levels were used, the lowest training contrast value and a value selected from a point between the highest and the middle values used during training. By using a contrast value in the test phase that was positioned between two of the training values, we expected to create ambiguity regarding the appropriate classification of a mask's contrast value, just as the use of an intermediate duration value in Experiment 2 created ambiguity with respect to classifying exposure durations. Studied words, regardless of whether they were presented in the read or the generate task, were expected to be more easily identified than new words, leading subjects to classify the mask as having a lower contrast value.

At the end of the experiment, subjects were asked to explain the basis for their contrast judgments and to estimate the proportion of target words that were old, just as they had done in Experiment 2. Fluent identification of words, sometimes created by prior study and sometimes created by using a low contrast mask, was expected to be attributed in many cases to prior study, thereby elevating subjects' estimates of the proportion of items that were old beyond the actual proportion. This pattern of overestimation can be taken as evidence that conscious recollection arises from an attribution that follows experienced fluency, rather than constituting an event that enables target identification or produces fluency.

5.1. Method

5.1.1. Subjects

Twenty-seven subjects participated in Experiment 3. These subjects were drawn from the same pool as those in the earlier experiments. Their ages ranged from 18 to 28 years, with a median age of 19 years. All subjects used in this experiment were native speakers of English.

5.1.2. Materials

The materials were identical to those used in Experiment 2.

5.1.3. Procedure

The general procedure used in this experiment was the same as that used in the previous experiments, except that rather than varying the exposure duration of target words in the training and test phases, the brightness contrast of the mask was manipulated and subjects were required to make a judgment about the mask.

The study phase was conducted in the same way as in Experiment 2. The training phase was similar to that used in Experiment 2, except that all target words were presented for 15 ms and followed by a mask, consisting of a row of ampersands, which was in view for 105 ms. A shorter target exposure duration was used here than in Experiments 1 and 2 (15 vs. 30 ms) because the brightness contrast of the mask was varied, making word identification easier under conditions of low contrast. A shorter target exposure duration was used to keep the identification task challenging. The mask characters were presented at one of three possible contrast values, as defined by the gray scale values on the Macintosh computer monitor. The values were 235, 190, and 160, where 255 corresponds to no contrast (invisible) and 0 corresponds to full black on white contrast. Fifteen trials were presented in random order using each contrast value. Subjects were told that the mask would vary in its effectiveness at obscuring the target word. They were instructed first to attempt to identify the target word, then to categorize the effectiveness of the mask as weak, medium, or strong by pressing one of three buttons on a button box.

The test trials began immediately after the training trials with no apparent change in procedure. For the test trials, however, only two contrast values were used, although subjects continued to make their responses based on the original three categories. The two contrast values used in the test phase were the lowest level of contrast used in training trials (235) and an intermediate contrast value that fell between the medium and highest contrast values used in training (170). The test phase began with a randomly ordered presentation of six practice items; three used the lowest contrast value and three used the intermediate value. The 75 critical trials and 75 filler trials then followed in random order. For the critical trials and for 25 filler trials, the mask was presented at the intermediate contrast value. For the other 50 filler trials, the mask contrast value was equal to the lowest value used in training. At the end of the test phase, subjects were asked to explain how they made their decisions about the effectiveness of the masks and to estimate the proportion of target words that had been presented in the study phase.

5.2. Results and discussion

The mean proportion of words correctly generated in the encoding phase was 0.95.

5.2.1. Target identification

The proportion of words correctly identified in the test phase is shown in Table 3 as a function of encoding condition. An ANOVA revealed a reliable effect of encoding condition, $F(2,52) = 22.14$, $\text{MSE} = 0.011$. Pairwise comparisons using the MSE from the omnibus test of means indicated that more items were correctly identified in each of the study conditions than in the new condition, $F(1,52) = 16.33$, and $F(1,52) = 43.55$, for the generate and read conditions, respectively. In addition, identification accuracy was higher in the read condition than in the generate condition, $F(1,52) = 6.54$.

These effects of encoding task on masked word identification are somewhat different from the results we obtained in Experiment 2 and from results reported in earlier studies by Masson and MacLeod (1992). In particular, for generation tasks like that used in Experiment 3, we typically have found comparable performance following generate and read encoding tasks on masked word identification. An exception to this pattern has been reported by Weldon (1991) and more recently by MacLeod and Masson (1997). In each of those two cases, three or four different encoding tasks were presented in blocked format to subjects during a study phase. MacLeod and Masson speculated that this format of presentation might have consequences for the way in which the encoding tasks are performed, thereby leading to differential

Table 3
Mean proportion of targets correctly identified and mean mask ratings as a function of prior encoding in experiment 3

Encoding task	Identification	Mask rating		
		All items	Correct ident.	Incorr. ident.
Generate	0.53 (0.22)	2.27 (0.33)	2.06 (0.34)	2.48 (0.46)
Read	0.60 (0.20)	2.27 (0.32)	2.16 (0.32)	2.46 (0.43)
New	0.41 (0.19)	2.36 (0.30)	2.14 (0.30)	2.50 (0.36)

Standard deviations are shown in parentheses.

performance on the masked identification task. In Experiment 3, however, only two encoding tasks were used and these were presented in the mixed list format typical of earlier experiments in which equal performance has been found with generate and read encoding tasks.

Another possibility is that differences in exposure duration of targets in the masked word identification task might account for the different pattern of results found here. In Experiment 3, we used an exposure duration of only 15 ms, rather than the 30-ms duration used in earlier studies. It could be that with a briefer exposure duration, presentation of the target might have been less likely to make contact with relevant conceptual knowledge, thereby reducing the probability of recruiting prior episodes involving conceptual processing. Finally, only low-frequency target words were used in Experiments 2 and 3, whereas Masson and MacLeod (1992), MacLeod and Masson (1997) and Weldon (1991) used words representing a wider range of frequencies. Differences between read and generate encoding tasks might be more substantial when low-frequency words are used. Although Experiment 2 did not reveal a reliable difference between read and generate items on masked word identification, the power of that experiment to detect a difference between read and generate of the size found in Experiment 3 was only 0.53.

5.2.2. Mask ratings

Mean ratings of the effectiveness of the visual mask are shown in Table 3. An ANOVA revealed a reliable difference among the three means, demonstrating that prior exposure to words influenced the mask rating, $F(2,52) = 5.50$, $\mathrm{MSE} = 0.013$. Pairwise comparisons using the MSE for the omnibus test of means showed that there was no reliable difference in the mean mask ratings associated with the generate and read conditions, $F < 1$, and that both of these conditions produced lower ratings of mask effectiveness than the new condition, $F(1,52) = 8.52$, and $F(1,52) = 7.98$, for the generate and read conditions, respectively.

This pattern of results replicates the findings in Experiments 1 and 2, in which generate and read encoding tasks had a similar effect on judgments about the perceptual quality of masked word presentation. Unlike Experiments 1 and 2, however, subjects in Experiment 3 were rating the effectiveness of the masking stimulus, rather than the duration of the target display. Therefore, prior study produced numerically lower ratings relative to the new condition in Experiment 3.

It is interesting that the generate and read encoding conditions had comparable effects on mask ratings in Experiments 2 and 3, given the finding that target identification success was reliably greater in the read condition in Experiment 3 and a similar but not reliable difference was found in Experiment 2. One might interpret this differential effect of encoding task as a dissociation between masked word identification and evaluation of mask effectiveness. Although that is an interesting possibility, the evidence available from the current experiments does not provide strong support for it. In particular, although Experiments 2 and 3 revealed a reliable effect of prior exposure on duration and mask ratings, those experiments had very little power to detect a difference in ratings between the read and generate conditions. Assuming an effect size proportional to that found in the masked word identification task in

Experiment 3 (i.e., 37% of the difference between read and new conditions), the estimated power of the duration rating task in Experiment 2 to detect a read–generate difference was 0.59. The estimated power of the mask ratings in Experiment 3 to detect a read–generate difference was only 0.19. Thus, the rating tasks appear to be less sensitive to encoding task differences than is the masked word identification task.

5.2.3. Conditionalized mask ratings

As in Experiment 2, a conditional analysis was conducted in which mean mask ratings were computed separately for those trials on which subjects correctly identified the target and for those trials on which identification failed. These means exclude the data from one subject who was unable to identify any of the critical target words. The mean conditionalized mask ratings are shown in Table 3. An ANOVA with encoding task and identification correctness (correct vs. incorrect) as factors revealed a reliable effect of identification correctness, $F(1,25) = 25.03$, MSE = 0.204. Mask effectiveness ratings were lower for targets that were correctly identified than for targets that were not correctly identified (2.12 vs. 2.48). There was no reliable effect of encoding task nor an interaction between identification correctness and encoding task, $Fs < 1.3$. The lack of an effect of encoding task in this analysis indicates that when identification success is partialled out, there is little systematic variation in ratings due to prior study. A similar result was observed in Experiment 2 in the 30-ms duration condition, although subjects in the 60-ms condition in that experiment showed an effect of prior study even when only identified targets were considered. As discussed in Experiment 2, there probably is not sufficient power to detect a reliable effect of prior study on the rating tasks when substantial proportions of trials are excluded from the computation of subject means by conditionalizing on identification success. The present analysis clearly shows, however, that subjects' ratings of mask effectiveness were strongly influenced by their ability to identify the target word.

5.2.4. Subjective report

At the end of the test phase, subjects attempted to describe the basis of their contrast judgments. Twenty-one of the subjects claimed that their judgments were based on whether or how easily they could see the word. Two subjects each reported that they relied on the intensity of the mask, the "darkness" of the target's letters (which did not physically vary), or the duration of the target word (which also did not vary). These reports generally support the conclusion that ratings were determined in large part by what subjects experienced in their attempt to identify the target word.

5.2.5. Estimated proportion of studied items

As in Experiment 2, we asked each subject to estimate the proportion of old items in the test phase of the experiment. We then plotted these estimates against the proportion of correctly identified targets in the test phase that were old. This scatter plot is shown in Fig. 2. As in Experiment 2, all but a few subjects overestimated the proportion of old items. Moreover, the degree of overestimation shown in Fig. 2 probably is less than the true amount of overestimation, given how we defined the actual proportion of old items (see Experiment 2 for further discussion). The overestima-

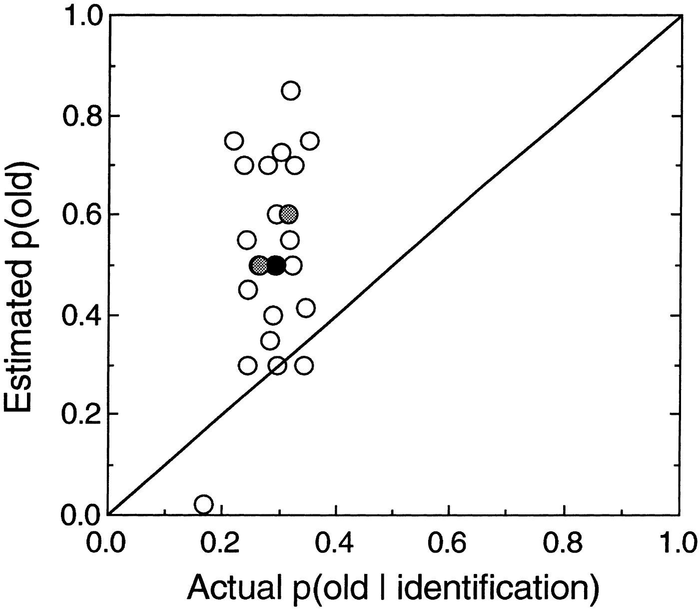

Fig. 2. Subjects' estimates of proportion of targets that were old (previously studied) plotted against actual proportion of correctly identified targets in the test phase that were old in Experiment 3. Gray and black symbols indicate that two or three subjects, respectively, had identical or nearly identical points in the plot. The diagonal line represents perfect estimation.

tion of proportion of old targets supports the view that processing fluency affects judgments about prior experience at the time an item is identified. By using a mask with low contrast, targets are rendered easier to identify and the resulting experience of processing fluency seems to be attributed to prior study even though the target had not been previously studied. This proposed sequence of events runs counter to the notion that identification and perceptual ratings (e.g., mask effectiveness) are mediated by or are sequelae of conscious recollection. Were genuine conscious recollection responsible for target identification and mask ratings, subjects' estimates of the proportion of targets that were old should not have shown a pattern of overestimation.

6. General discussion

The reported experiments examined whether processing fluency associated with prior exposure to a word would be misattributed to a perceptual aspect of the stimulus display, even when the source of processing fluency was a prior conceptually driven encoding episode. In Experiments 1 and 2, judged duration of a target word presented in a masked word identification task was increased for targets that earlier were presented either in a generate or a read task. These two encoding experiences produced similar changes in judged duration of target words, relative to target words that had not been encoded earlier. Similarly, in Experiment 3, where the brightness contrast of the mask was varied in the masked word identification task, judged effectiveness of the mask at camouflaging the target word was reduced for previously studied targets. Again, this effect was similar for both the generate and the read encoding tasks.

Three sources of evidence indicated that the influence of prior exposure operated through subjects' attempts to identify the masked target words. First, the effect of prior exposure was less robust when subjects were not instructed to identify the target words (Experiment 1). This point must be qualified, however, by other differences between Experiment 1 and the later experiments (i.e., the critical exposure durations used in Experiment 1 were the same as durations used during the training phase, rather than new durations placed between trained durations; target words varied in frequency in Experiment 1 but were low-frequency in the later experiments). Second, when identification attempts were required, it was found that successfully identified targets received longer duration ratings and lower mask effectiveness ratings than targets that were not correctly identified, regardless of prior exposure. Finally, post-experiment questioning of subjects indicated that a large majority of subjects used ease or success of target identification as a heuristic in judging duration or mask effectiveness. Thus, the fluency with which targets could be identified played a substantial role in determining judgments about stimulus characteristics.

The influence of identification success on judgments might lead one to suspect that identification or stimulus judgments were determined either by a strategy whereby subjects systematically gave higher duration ratings to words they successfully identified, or by conscious recollection of the earlier presentation of target words. Evidence that subjects did not rely entirely on a systematic strategy comes from subjects in the 60-ms condition of Experiment 2. These subjects nearly always correctly identified the target word, yet they showed a difference between studied and new words in their duration ratings. Even when we considered only words that these subjects correctly identified, ratings were reliably higher for items that had been studied. These results indicate that subjects were responding not only to the coarse difference between identified and unidentified targets, but perhaps also to differences in the fluency with which identification was achieved.

On the question of conscious recollection as a basis for the duration and mask effectiveness ratings, we believe that this possibility is unlikely for a number of reasons. First, the sensitivity of ratings to identification success cut across all target encoding conditions, including the new condition. Moreover, generate and read encoding tasks produced similar ratings of target duration and mask effectiveness, despite the fact that conscious recollection of items generated from semantic cues like those used here is much more likely than recollection of items in the read condition (Masson and MacLeod, 1992; MacLeod and Masson, 1997). Finally, subjects' estimates of the proportion of targets that had been studied earlier typically were substantially higher than the actual proportion of items that were old, as shown in Figs. 1 and 2. This pattern of overestimation is not what would be expected if conscious recollection determined successful target identification. Rather, subjects ought to have recollected only some of the studied items, producing something of an underestimate of the proportion of items that were old (as when hit rates on a recognition memory test are below ceiling). The strong tendency toward overestimation suggests instead that subjects experienced a sense of familiarity with targets as a result of fluent processing that was enabled by favorable stimulus conditions (i.e., longer target duration or lower contrast mask). By attributing this fluency to prior study, subjects would mistakenly experience many of the filler or new critical items as having been previously studied.

It is particularly important that the experiments reported here demonstrated that both conceptually driven and data-driven encoding episodes led subjects to experience enhanced fluency in the masked word identification task. That finding suggests that these two different sources of fluency induce qualitatively similar experiences, inasmuch as both led to elevated judgments of exposure duration and to reduced judgments of mask contrast. Two significant implications stem from this finding.

First, these results are inconsistent with the view that the masked word identification task should be construed as purely data-driven and susceptible to repetition priming following only those encoding episodes that involve perceptual processing (e.g., Jacoby, 1983; Toth et al., 1994; Weldon, 1991). Rather these studies are consistent with the view that brief presentation of a masked word can make contact with memory for prior encoding episodes involving either perceptual or conceptual processing (see MacLeod and Masson, 1997; Masson and MacLeod (1992, 1997), for a detailed account of this view). A wider consideration of this issue suggests that greater caution should be exercised when one is tempted to classify a test of memory as purely or even primarily data-driven or conceptually driven. In our view, this concern is as valid as concern regarding the possibility that direct and indirect tests of memory are not process pure with respect to conscious and unconscious influences of memory (e.g., Jacoby, 1991; Jacoby et al., 1993; Kelley and Jacoby, 1998).

It might be argued that the differential fluency experienced by subjects when attempting to identify studied versus nonstudied words in the masked word identification task might not derive from the same processes that determine identification performance. The basis for this argument is that Experiment 3, and to some extent Experiment 2, produced a dissociation between masked word identification and fluency ratings; the read encoding task led to better identification performance than the generate task, whereas the two tasks yielded equal duration or mask ratings. There was even a suggestion that the generate task had a greater influence on ratings than the read task, when those ratings were conditionalized on correct identification, although those effects were not reliable. A dissociation between masked word identification and ratings, however, does not necessarily mean that the two measures are determined by different processes (e.g., Dunn and Kirsner, 1988).

Instead, the dissociation may have arisen because prior conceptual processing has a larger influence on attributional processes than does prior perceptual processing. Thus, even though fewer generate items were correctly identified, the processing fluency associated with those items might have had a larger impact on ratings of perceptual features of the target display than the fluency associated with read items. This possibility is consistent with the finding reported by Whittlesea (1993) that conceptual processing manipulations had stronger effects on attributions involving judgments of familiarity than did perceptual processing manipulations. Finally, the reliability of the dissociation remains in doubt because of the relatively low power of the ratings measure to detect a difference between the read and generate encoding conditions. A full examination of the question of a possible dissociation between masked identification performance and ratings of perceptual features of target events must await further research.

The second significant implication of our finding that conceptual and perceptual encoding episodes lead to similar experiences of processing fluency is that these experiences may be generic in nature. This suggestion is not entirely novel. Whittlesea (1993) obtained evidence for a generic influence of processing fluency by showing that decisions about the semantic relatedness of present and past events can be influenced by either perceptual or conceptual aspects of current processing. Like the generalized heightening of activation produced by adrenaline in the Schachter and Singer (1962) study, a sense of processing fluency appears, at least in its initial stages, to be amorphous, awaiting the influence of internal goals or external cues to determine a source to which that sensation ought to be attributed. It remains to be seen how wide a variety of origins of processing fluency are interchangeable in this way.

Acknowledgements

This research was supported by a research grant to Michael Masson and by a postgraduate scholarship to Judy Caldwell, both from the Natural Sciences and Engineering Research Council of Canada. We are grateful to Graham Brown, Chris Gill, Carrie Hicks, Alyssa Megale, and Allison Patton for their assistance in conducting the experiments, and to two anonymous reviewers for insightful comments on an earlier version of this paper.

References

Blaxton, T.A., 1989. Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. Journal of Experimental Psychology: Learning, Memory, and Cognition 15, 657–668.

Dunn, J.C., Kirsner, K., 1988. Discovering functionally independent mental processes: The principle of reversed association. Psychological Review 95, 91–101.

Jacoby, L.L., 1983. Remembering the data: Analyzing interactive processes in reading. Journal of Verbal Learning and Verbal Behavior 22, 485–508.

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

Jacoby, L.L., Allan, L.G., Collins, J.C., Larwill, L.K., 1988. Memory influences subjective experience: Noise judgments. Journal of Experimental Psychology: Learning, Memory, and Cognition 14, 240–247.

Jacoby, L.L., Brooks, L.R., 1984. Nonanalytic cognition: Memory, perception, and concept learning. In: Bower, G.H. (Ed.), The psychology of Learning and Motivation, vol. 18. Academic Press, New York, pp. 1–47.

Jacoby, L.L., Dallas, M., 1981. On the relationship between autobiographical memory and perceptual learning. Journal of Experimental Psychology: General 110, 306–340.
Jacoby, L.L., Kelley, C.M., Brown, J., Jasechko, J., 1989a. Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. Journal of Personality and Social Psychology 56, 326–338.
Jacoby, L.L., Kelley, C.M., Dywan, J., 1989b. Memory attributions. In: Roediger III, H.L., Craik, F.I.M. (Eds.), Varieties of Memory and Consciousness: Essays in Honor of Endel Tulving. Erlbaum, Hillsdale, NJ, pp. 391–422.
Jacoby, L.L., Toth, J.P., Yonelinas, A.P., 1993. Separating conscious and unconscious influences of memory. Journal of Experimental Psychology: General 122, 139–154.
Jacoby, L.L., Whitehouse, K., 1989. An illusion of memory: False recognition influenced by unconscious perception. Journal of Experimental Psychology: General 118, 126–135.
Jacoby, L.L., Woloshyn, V., Kelley, C.M., 1989c. Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. Journal of Experimental Psychology: General 118, 115–125.
Johnston, W.A., Dark, V.J., Jacoby, L.L., 1985. Perceptual fluency and recognition judgments. Journal of Experimental Psychology: Learning, Memory, and Cognition 11, 3–11.
Kelley, C.M., Jacoby, L.L., 1990. The construction of subjective experience: Memory attributions. Mind and Language 5 (1), 49–68.
Kelley, C.M., Jacoby, L.L., 1996. Adult egocentrism: Subjective experience versus analytic bases for judgment. Journal of Memory and Language 35, 157–175.
Kelley, C.M., Jacoby, L.L., 1998. Subjective reports and process dissociation: Fluency, knowing, and feeling. Acta Psychologica 98, 7–20.
Kučera, J., Francis, W.N., 1967. Computational analysis of present day American English. Brown Univ. Press, Providence, RI.
Kunst-Wilson, W.R., Zajonc, R.B., 1980. Affective discrimination of stimuli that cannot be recognized. Science 207, 557–558.
Lindsay, D.S., Kelley, C.M., 1996. Creating illusions of familiarity in a cued recall remember/know paradigm. Journal of Memory and Language 35, 197–211.
Luo, C.R., 1993. Enhanced feeling of recognition: Effects of identifying and manipulating test items on recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition 19, 405–413.
MacLeod, C.M., Masson, M.E.J., 1997. Priming patterns are different in masked word identification and word fragment completion. Journal of Memory and Language 36, 461–483.
Mandler, G., Nakamura, Y., Van Zandt, B.J.S., 1987. Nonspecific effects of exposure on stimuli that cannot be recognized. Journal of Experimental Psychology: Learning, Memory, and Cognition 13, 646–648.
Masson, M.E.J., 1984. Memory for the surface structure of sentences: Remembering with and without awareness. Journal of Verbal Learning and Verbal Behavior 23, 579–592.
Masson, M.E.J., MacLeod, C.M., 1992. Reenacting the route to interpretation: Enhanced perceptual identification without prior perception. Journal of Experimental Psychology: General 121, 145–176.
Masson, M.E.J., MacLeod, C.M., 1997. Episodic enhancement of processing fluency. In: Medin, D.L. (Ed.), The Psychology of Learning and Motivation, vol. 37. Academic Press, San Diego, CA, pp. 155–210.
Merikle, P.M., Reingold, E.M., 1991. Comparing direct (explicit) and indirect (implicit) measures to study unconscious memory. Journal of Experimental Psychology: Learning, Memory, and Cognition 17, 224–233.
Paller, K.A., Mayes, A.R., McDermott, M., Pickering, A.D., Meudell, P.R., 1991. Indirect measures of memory in a duration-judgement task are normal in amnesic patients. Neuropsychologia 29, 1007–1018.
Roediger III, H.L., Blaxton, T.A., 1987. Effects of varying modality, surface features, and retention interval on priming in word fragment completion. Memory Cognition 15, 379–388.
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.

Schachter, S., Singer, J., 1962. Cognitive, social, and physiological determinants of emotional state. Psychological Review 69, 379–399.

Seamon, J.G., Marsh, R.L., Brody, N., 1984. Critical importance of exposure duration for affective discrimination of stimuli that are not recognized. Journal of Experimental Psychology: Learning, Memory, and Cognition 10, 465–469.

Toth, J.P., Reingold, E.M., Jacoby, L.L., 1994. Toward a redefinition of implicit memory: Process dissociations following elaborative processing and self-generation. Journal of Experimental Psychology: Learning, Memory, and Cognition 20, 290–303.

Weldon, M.S., 1991. Mechanisms underlying priming on perceptual tests. Journal of Experimental Psychology: Learning, Memory, and Cognition 17, 526–541.

Whittlesea, B.W.A., 1993. Illusions of familiarity. Journal of Experimental Psychology: Learning, Memory, and Cognition 19, 1235–1253.

Whittlesea, B.W.A., 1997. Production, evaluation, and preservation of experiences: Constructive processing in remembering and performance tasks. In: Medin, D.L. (Ed.), The Psychology of Learning and Motivation, vol. 37. Academic Press, San Diego, CA, pp. 211–264.

Whittlesea, B.W.A., Jacoby, L.L., Girard, K.A., 1990. Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. Journal of Memory and Language 29, 716–732.

Witherspoon, D., Allan, L.G., 1985. The effect of a prior presentation on temporal judgments in a perceptual identification task. Memory Cognition 13, 101–111.

Zajonc, R.B., 1980. Feeling and thinking: Preferences need no inferences. American Psychologist 35, 151–175.