Abstract  Ordinarily, deeper levels of processing in a study session increase the accuracy of later remembering. We modified the standard levels-of-processing procedure by presenting items either once or twice in the study phase, each item being the subject of a semantic, phonemic, or graphemic question. At test, the subjects judged the frequency with which each word had occurred in the study phase. Deeper processing during encoding increased accuracy in judging twice-presented items. However, it also caused an illusion of repetition for items presented only once. The result underlines the importance of thinking of remembering as a process of evaluation and inference, rather than simple retrieval.

In 1978, Lee Brooks announced a Grand Idea: that although memory is responsible for the incredible diversity of human performance, the actual mechanism underlying that complexity might be very simple. Rather than precomputing the abstract essences of all of the various concepts and categories that people must deal with, he suggested that memory encodes each separate stimulus experience; those experiences, cued narrowly or broadly by future contexts, could act singly or in parallel to control behaviour. This encoding of instances rather than abstracting relationships proved to be a very powerful idea: It could not only explain people’s sensitivity to regular or typical aspects of past experience (e.g., implicit learning of grammars; cf. Vokey & Brooks, 1992), but also suggested a way in which people could make new discriminations “on the fly,” in dealing with unanticipated decisions (e.g., Kahanman & Miller, 1986).

One implication of this idea is that, rather than having precomputed answers to important questions, people are often in the position of making what sense they can of those aspects of the past that are cued by current stimulus conditions. Applied to the problem of memory for particular events rather than general concepts, this idea leads one to think of remembering as a problem-solving function, involving evaluation, interpretation and attribution, rather than simple retrieval (an idea originally proposed by Bartlett, 1932). This aspect of memory has been explored in a number of paradigms, including studies of source monitoring (e.g., Johnson, Hashtroudi, & Lindsay, 1993), the fluency heuristic (e.g., Jacoby & Whitehouse, 1989), the misinformation effect (e.g., Zaragoza & Lane, 1994), and autobiographical memory (e.g., Lindsay & Read, 1994). These investigations all demonstrate cases in which claims of remembering are modulated by factors affecting the quality of processing, independent of the retrieval of memory contents. Such evidence appears to suggest that inferential processes are an essential part of the act of remembering.

However, one might question the generality of that claim. For example, the classic studies that introduced the idea of transfer-appropriate processing show that remembering improves to the extent that activities performed in test overlap with those conducted in the stimulus encounter (e.g., Kolers, 1976; Morris, Bransford, & Franks, 1977; Tulving & Thompson, 1973). Such observations clearly demonstrate the crucial role of retrieval cues in remembering; they do not appear to require additional assumptions about inference and attribution. That might suggest that such processes may only be induced in special cases, biasing the normal course of remembering. Alternatively, however, the separate influence of those processes may simply not be observable in those studies, because manipulation of retrieval simultaneously affects a major component of the quality of processing. To investigate that issue, we performed a re-examination of one of the most celebrated effects that helped to bring about the cognitive revolution, namely, levels of processing (henceforth LOP).

The LOP idea was initially proposed by Craik and Lockhart (1972; see Craik, 2002 for a fascinating retrospective on the history of the effect). The effect itself was clearly seen in an experiment conducted by Craik and Tulving (1975). Those authors presented words in a study phase, accompanied by one of three types of questions, about the graphemic, phonemic or semantic properties of the word. The famous result was that the
“shallow” task led to poor later recognition (less than 20%) whereas the “deep” task resulted in excellent recognition (about 80%). Publication of that work aroused a controversy about whether the framework was circular (defining depth in terms of outcome rather than some independent standard), whether depth is really different from other ways of making processing more elaborate, and whether the effect of deeper processing occurs through creating a more effective knowledge structure that facilitates reconstruction of the original encoding (recapitulation of the encoding process; cf. Kolers, 1973, 1976) or whether instead it permits an experience that has been retrieved to be distinguished more easily from other experiences of the same word (e.g., Moscovitch & Craik, 1976). However, despite these residual issues, the field as a whole agreed that the effect was informative and needed to be included in every textbook on cognition or memory (at last count, the Craik & Lockhart article has been cited 2657 times).

The conclusion that LOP assists remembering was reached under procedures motivated by the “memory quantity” approach identified by Koriat and Goldsmith (1996; Koriat, Goldsmith, & Pansky, 2000). In this approach, the chief concern is to discover how much information can be stored in memory and effectively recovered, and under what conditions. In keeping with that approach, Craik and Tulving (1975) gave their subjects a list of words to study under various conditions that could affect later remembering. The words themselves served as interchangeable units that could be used to index amount of remembering. What mattered was not what words were remembered, but the conditions that caused more of them to be remembered.

Following a recognition test involving the 60 studied items and 120 new items, the investigators counted the number of old items in each condition that the subjects checked off. Those counts were used to calculate the proportion of each training list that was recovered in the test (the “input-bound” approach identified by Koriat and Goldsmith). Significantly, false alarms were not reported in that article (as they also were not in many contemporary articles). By the “memory quantity” approach, false alarms are not of much interest, because, being claims about items not shown in the past, they by definition could not reveal anything about how prior experience controls retrieval. No attempt was made to examine the subjective quality of reports; an item that was ticked off in the recognition test was assumed to be remembered in the same way as each other item.

To us, what is most important about this summary of LOP is that it does not seem to require a concern with the phenomenology of remembering. What could it add to the conclusion that elaboration is important for successful remembering, to insist on examining the metacognitive or phenomenological activities taking place during the test session? Given that the subjects said “old” or “new” to each test item, they presumably felt that they were old or new, so that the phenomenology lined up with the performance: Elaboration may lead to greater success in identifying old items through creating a difference in subjective experience, but because that subjectivity apparently has no independence from performance, it may not seem necessary to consider it in understanding the effects of elaboration on remembrance. Our current investigation was designed to demonstrate the importance of considering inferential processes along with retrieval even in cases in which their influence is not obvious.

In the standard LOP study, half of the items presented in the test have been shown once in the study session and the others have not; all have been seen many times previously, before the experiment. The challenge for the subjects in such a study is thus to discriminate between items that have been presented within the experimental context in addition to many other contexts from those that have not been seen within the study session. The effect of manipulating LOP has also been examined with frequency judgments. For example, Maki and Ostby (1987) required subjects to discriminate between items presented once, twice or four times in the study session. They observed that deeper levels of processing increased the accuracy of discrimination and the effect of elaboration is important for the subjects in such a study is thus to discriminate between items presented within the experimental context in addition to many other contexts from those that have not been seen within the study session. The effect of manipulating LOP has also been examined with frequency judgments. For example, Maki and Ostby (1987) required subjects to discriminate between items presented once, twice or four times in the study session. They observed that deeper levels of processing increased the accuracy of discrimination between less and more frequent words.

However, in that investigation, the subjects were required to perform a “levels” task on every study trial, so that repeated words were repeatedly processed at a higher level. This repetition of the induction task means that the LOP effect could occur through superior encoding (caused either by creating multiple distinct traces or by allowing the subject to recognize repetitions as such during the study phase), rather than through an inferential process performed in the test. Moreover, the dependent variable was the mean judged frequency of words presented various numbers of times. Because that measure is based on both correct and incorrect estimates of frequency, it is difficult to examine the effects of the levels manipulation on those outcomes separately.
only once, a “yes” answer was a false alarm. We examined the effects of the LOP manipulation on each category of answer. Second, subjects were required to answer an LOP question only once for each word. If a word was shown twice, it was presented once to be read aloud and once as the subject of an LOP question; if presented only once, the subjects were asked to answer an LOP question. Thus, in contrast to the Maki and Ostby (1987) procedure, later reporting that a word had occurred twice could not be based on multiple experiences of answering the LOP question, but rather would have to be based on remembering a second presentation of the word without a question. It was thus of interest to examine whether different levels of processing in one experience of a word would influence claims about a second putative experience of the same word that would have occurred without an LOP question. That is, this procedure allowed us to examine directly the effects of the levels manipulation on inferential processes leading to claims of remembering.

Method

Subjects

Twenty-two Simon Fraser University students participated for course credit.

Procedures

We selected 120 words from the MRC Psycholinguistic Database (http://www.psy.uwa.edu.au/MRCDatabase/uwa_mrc.htm). The words were five to eight letters in length and had moderate familiarity and concreteness (ratings of between 350 to 400, and 400 to 600, respectively). We generated six questions for each word, so that any word could be asked about at any level (semantic, phonemic or graphemic) with either a “yes” or “no” answer expected.

In the study phase, 60 words were presented only once and 60 twice. On one occurrence of twice-presented words (either the first or second, at random), the subject was signaled to read it aloud. All other presentations of words were accompanied by an LOP question (20 at each level in both frequency conditions). Within each of the six conditions presenting a question (one vs. two presentations crossed with three levels), 10 items were shown with a “yes”-answered question and 10 with a “no”-answered question. No question was asked twice. Subject to these constraints, words (and the appropriate questions) were assigned at random to conditions, and rerandomized for each subject.

Because half the words were presented twice, there were 180 study trials. Words in this phase were presented in a random order, rerandomized for each subject. During the test phase, subjects were again shown all of the words shown in training (in a freshly random-ized order) and instructed to indicate whether they had seen the target once or twice in the study phase by pressing a button on a button box. There were 120 trials in this test. A 3 (Levels) x 2 (Once/Twice Presented) x 2 (Yes/No Answer Expected) repeated measures ANOVA was carried out; only effects of interest are reported. An alpha level of .05 is assumed throughout.

Results and Discussion

The data are presented in Table 1. Error rates in answering LOP questions were very low (less than 2% for most subjects). Overall, the subjects achieved 39% hits and 14% false alarms, $F(1,21) = 70.63, MSE = .06, p < .001$. They clearly treated the decision as anchored at one presentation, responding “yes” only if they thought they had some evidence of an extra presentation.

We observed two effects associated with the classical demonstrations of LOP (Craik & Tulving, 1975). First, of course, was an increase in the number of hits with increasing depth of processing, $F(2,21) = 19.30, MSE = .02, p < .001$. Each increase in depth resulted in about a 10% increase in hits, such that graphemic processing produced about 50% hits and semantic processing about 30% hits. Greater depth also led to greater differences between hits and false alarms, such that the differences were about 22%, 25%, and 30% for graphemic, phonemic, and semantic questions, respectively, $F(2, 21) = 3.81, MSE = .01, p = .030$. Second, we also observed a tendency to claim words to have been presented twice more often when the correct answer to the question asked in the training was “yes” rather than “no,” although that effect was of only marginal reliability, $F(1, 21) = 3.92, MSE = .02, p = .061$.

However, we also observed that increasing depth of processing led to increasing rates of false alarms, $F(2,21) = 18.43, MSE = .02, p < .001$. Each increase in depth resulted in about 6% more false alarms, such that graphemic processing produced about 8% false alarms and semantic processing about 20% false alarms. That is, in discrimination between one versus two presentations, manipulation of LOP between shallowest and deepest did almost as much harm (12%) as did good (20%). (The same effects are observable in SDT mea-
sures of discrimination ($d_L$) and bias ($C_L$), as shown in Table 2. Increasing levels of processing from graphemic to semantic reliably increased discrimination, $F(2,21) = 5.67$, $MSE = .67$, $p = .007$, but also reliably led to a more liberal bias, $F(1.21) = 5.74$, $MSE = 1.35$, $p = .006$.

Thus the same factor that increases accurate report of double presentation also promoted false report of double presentation. This effect could not have been observed in the original LOP experiments (e.g., Craik & Tulving, 1975), because items presented in test had been either shown or not shown in the study session, and in the latter case, of course, the subjects could not experience different levels of processing. That factor thus could not bias the later remembering decision for those items, so that LOP could not influence false alarms.

We suspect that some readers will regard the use of one versus two presentations as a special case, unrepresentative of the “real” remembering problem, which is “did this word occur at all?” However, although the “once-versus-twice” paradigm is unusual in studies of remembering, it is quite common in studies of perception (e.g., the phenomenon of “repetition blindness”: cf. Kanwisher, 1987; Whittlesea & Podrouzek, 1995). Moreover, once-versus-twice decisions occur naturally in real life: Did I lock the door the second time I left? Did I take both doses of medication, at lunch and dinner? Did I meet with Sam once or twice last week? and so on. We thus regard our paradigm as a fair examination of the LOP effect: That LOP has some negative consequences for remembering under these circumstances is simply part of the larger picture of memory.

Clearly, greater depth did improve the accuracy of remembering later: It probably did so (at least in part) by increasing the distinctiveness of the study experience or by assisting the subjects in regenerating parts of their earlier experience, as Craik and Lockhart (1972) suggested. However, the results also reveal that the effect of the levels manipulation is not simply to allow the person to know more accurately their past experience, but also to cause an illusion of having seen a word twice. That additional complexity can be explained, we believe, only through considering the inferential function of memory.

If the effect of LOP was simply to assist the person to retrieve distinctive information about a prior experience of a word (process it in a way that is diagnostic of having processed it before in a specific way), it could not have had the effect of increasing claims that items presented only once were shown twice in our study. That is because every word shown in our test was also shown once in training with a levels question. By an account based only on retrieval, when a word was shown twice, the subjects would remember the isolated presentation with some probability and also the LOP presentation with some other probability; deeper processing on the latter occasions would increase the likelihood of remembering that occurrence, thus increasing the probability of accurately reporting two occurrences. But when a word was presented only once, with an LOP question, the likelihood of remembering that event could not combine with remembrance of an isolated presentation to cause a claim of double presentation. Deeper levels of processing could only make it more likely that the person would remember that single presentation, which could only inform the person that that word was seen once. If that was all there were to the effect of LOP, we would not have observed an effect on false alarms.¹

We therefore reject the idea that the effect of LOP was simply to increase retrieval of earlier occurrences. Instead, we suggest that it also changed the subjective quality of the experience, and that people use that subjective quality, rather than just retrieval, to make remembering decisions. Deep processing did not just make the words more likely to be retrieved; in addition, it caused them to feel special. The subjects used this phenomenology heuristically, attributing a rich subjective quality to multiple prior exposures. That is, we argue that the illusion of repetition caused by deeper levels of encoding reveals two separate functions of memory: the production of mental events cued by current stimuli and prepared by prior experiences, and the evaluation of the quality with which that production occurs, resulting in subjective experience (cf. Whittlesea, 2002).

¹ The effect of LOP on accurate claims of “presented twice” may have been to some extent caused by people noticing repetition more often during the training on seeing the isolated presentation, when the question asked on the other occasion was deep rather than shallow. Remembering later that one had seen a word in isolation could be used effectively to make a claim of “presented twice,” because that only happened when a word was presented twice. However, that does not affect the interpretation of the LOP effect on claims of “twice” for once-presented items. Subjects could not use a “recall to reject” strategy (Rotello & Heit, 2000), because in that case isolated presentations were not shown.
We suggest that the same two functions are engaged in the standard LOP study, in which only half of the test words are exposed in training. However, in that case, the subjective quality caused by deep encoding is redundant with the retrieval: Because both have the same effect, of making the person claim to recognize a word that actually was shown, it is not possible to see their separate contributions to the process. That does not mean that one can afford to ignore the importance of the two separate functions; instead, it means that, to understand the complexity of memory, we should seek phenomena in which those functions can be independently manipulated.

There is a growing literature on false positives in remembering, including errors caused by enhancing fluency of test items (Higham & Vokey, 2000; Jacoby & Whitehouse, 1989; Joordens & Merkle, 1992), presenting lists of related words in study (Roediger & McDermott, 1995), and having subjects solve anagrams in test (Westerman & Greene, 1996). The same effect is observed in attempts to increase the amount of memory retrieval using techniques such as hypnosis (Loftus, 2000), guided imagery (Wade, Garry, Read, & Lindsay, 2002), and the cognitive interview (Roberts & Higham, 2002). In each of these phenomena, some principle that assists accurate remembering under some circumstances causes false memory under others. These errors in remembering do not mean that memory cannot be trusted. Instead, they mean that remembering is a fundamentally inferential and attributional function, in performing which people can only make the best sense that they can of the quality of their current experience. It appears that the most general conclusion that one can reach about remembering accuracy is: “It all depends.”

Antonia Kronlund and Bruce W. A. Whittlesea, Simon Fraser University. This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to the second author.

Correspondence concerning this article should be addressed to Bruce W. A. Whittlesea, Department of Psychology, Simon Fraser University, Burnaby, British Columbia V5A 1S6 (E-mail: bruce_whittlesea@sfu.ca).

References
Le présent article célèbre la carrière de Lee Brooks en examinant un aspect implicite de son idée primordiale, à savoir que plutôt que de posséder des réponses préétablies à des questions importantes, les gens sont souvent contraints d’interpréter les aspects du passé évoqués par des stimulus actuels (Brooks, 1978). S’agissant du problème du souvenir d’événements particuliers plutôt que de concepts généraux, l’idée porte à croire que la remémoration est une fonction de résolution de problèmes, faisant intervenir évaluations, interprétations et attributions, plutôt qu’un simple processus de récupération. Nous avons examiné la question en adaptant la procédure normalisée fondée sur les niveaux de traitement par la présentation, à une ou deux reprises, d’articles particuliers à l’étape de l’étude, chaque article ayant été l’objet d’une question sémantique, phonémique ou graphémique. À l’étape du test, les sujets ont jugé de la fréquence de la présentation de chaque article à l’étape de l’étude. Un traitement approfondi au cours de l’encodage a donné lieu à un nombre accru de jugements précis concernant les articles présentés deux fois. Par contre, il a aussi engendré l’illusion que des articles présentés une fois avaient été montrés à deux reprises. Il suit que le facteur qui accroît la précision des déclarations de double présentation hausse également les fausses déclarations de double présentation.

Les résultats obtenus font valoir l’importance de considérer la remémoration comme un processus d’évaluation et d’inférence plutôt qu’un simple mécanisme de récupération. En clair, l’approfondissement du traitement s’est traduit par une précision accrue de la remémoration ultérieure, phénomène qui était probablement attribuable (du moins en partie) à l’intensification du caractère distinctif de l’expérience de l’étude ou par l’aide dont bénéficiaient les sujets pour ranimer des aspects de l’expérience antérieure, comme l’ont proposé Craik et Lockhart (1972). Par ailleurs, les résultats montrent également que la manipulation des niveaux n’a pas pour effet seulement de permettre à une personne de connaître avec précision son expérience, elle provoque également l’illusion qu’un sujet a perçu un mot deux fois. À notre avis, seule la prise en compte de la fonction inférentielle de la mémoire peut expliquer la complexité accrue dont nous faisons état ci-dessus.