Fluency - Friend or foe for learners? An examination of whether the perceptual fluency effect occurs for complex text material

B. Devlin-Glass

School of Psychology, Building 49, The Australian National University, Acton ACT 0200, Canberra, Australia

The perceptual fluency effect refers to the finding that people base their judgments about future memory performance on the ease with which material is comprehended during learning. Under some conditions, an illusion of comprehension may occur, leading to an overestimate of ability to recollect studied information. This experiment tested 95 participants to examine whether the perceptual fluency effect can lead to an illusion of knowing about complex text material. In Experiment 1, 80 participants studied six text extracts. Fluency was manipulated via inclusion of familiar but only tangentially-related material aimed at fostering a level of ease which might lead to an illusion of knowing. There were three presentation conditions, manipulated between-subjects: In the control condition, extraneous material was absent, and participants studied only baseline texts; in the separated condition, extraneous material was presented in a paragraph preceding the baseline texts; in the integrated condition, extraneous material was incorporated into baseline texts. Prior to study, participants provided ease-of-learning judgments (EOLs), estimating how easy they believed the material would be to learn. Following study, participants provided judgments of learning (JOLs) either immediately or following a 24-hour delay, estimating how they thought they would perform on subsequent tests of the material. Participants then took comprehension tests to measure their accuracy. Experiment 2 was identical but used a within-subjects design: 15 participants studied six texts, two from each of the three fluency conditions. Results showed that the presence of extraneous information misled learners about their level of understanding, with poor calibration between JOLs and accuracy in both experiments, but particularly where the extraneous information was interleaved or integrated with the basic text. Further analyses revealed that participants not exposed to extraneous information were able to answer more high-level inference questions than those who were exposed to such material. These results have implications for educators, particularly in the design of psychologically-appropriate coursework for students, and in instructing students to recognise textual cues which can mislead metamnemonic monitoring.

When studying for exams, many students experience difficulty predicting their likely performance (Benjamin, Bjork & Schwartz, 1998). Students often get frustrated: They believe they have studied effectively and expect to perform well, make a judgment that their memory of the material is sufficient to cease revision, and are disappointed when they perform more poorly than anticipated on the subsequent test (Shaddock & Carroll, 1997). Laboratory studies corroborate these observations: Evidence suggests most students stop self-paced study too soon, and consequently attain test results which are both inadequate *and* inconsistent with initial predictions about their likely performance (Mazzoni, Cornoldi & Marchitelli, 1990). Evidently, then, the reading an individual takes of his or her capacity to recall studied material can be as

important as his or her actual comprehension of that information (Benjamin & Bjork, 1996).

If learners are to perform effectively in testing situations, they need a range of monitoring strategies to assess their cognitive *competence*, in addition to the cognitive strategies used to facilitate learning and task completion (Flavell, 1979). These monitoring strategies—termed ‘metacognitive skills’—include activities such as planning and regulating one’s cognitive processes during learning, setting goals and selecting study strategies, and monitoring comprehension (Brown, Bransford, Ferrara & Campione, 1983). Arguably the most important metacognitive skill, however, is making accurate mnemonic judgments about the objective state of one’s memory (King, 1991). Because metamnemonic judgments about learning and future

performance are linked to the amount of cognitive effort a student will employ when studying a particular topic, the accuracy of such predictions is inextricably related to a student's likelihood of success on a subsequent test (Nelson & Narens, 1992). Ultimately, the effective monitoring of memory is critical for achieving optimal control of subsequent study, affecting the strategic regulation of learning and remembering processes and, therefore, memory performance itself (Koriat & Goldsmith, 1994). Accordingly, understanding the variables which influence effective memory monitoring is of prime importance in designing psychologically-appropriate coursework and study strategies for students (Benjamin, 2003). If a relationship exists between metamnemonic monitoring and later recall, then improving students' ability to assess their mnemonic capacity should enhance students' actual performance.

This research aims to examine whether text-based cues can influence and mislead metamnemonic monitoring, and consequently affect learning and performance. It will investigate whether perceptual fluency—or the familiarity or ease with which text can be comprehended during learning—can lead students to misdiagnose what they know about complex text. Hitherto, fluency's pervasive misleading influence has largely been studied using list-learning and cue-pair associates paradigms. Investigation of whether the perceptual fluency effect persists for complex text material will be carried out using a methodology developed specifically for this study. The research seeks to answer the question: does the presence of familiar extraneous material *aid* learners in helping them comprehend, or does it *mislead* them by inducing an illusory belief about comprehension?

INVESTIGATING METAMNEMONIC MONITORING

To date, judgments of learning (JOLs) have been the index most commonly used to investigate metamnemonic monitoring. JOLs are metacognitive judgments in which one predicts the likelihood of remembering an item on a criterion test (see, for example, Nelson & Narens, 1992). Several extraneous variables influence JOL accuracy. Research indicates that delaying JOLs for 30 seconds after study increases their accuracy relative to JOLs made immediately after study (Hertzog et al., 2003); this effect persists even when JOLs are delayed for 24-hours (Dunlosky & Nelson, 1994; Shaddock & Carroll, 1997). When JOL is delayed,

the information used to make such estimates involves a target retrieval attempt that is highly diagnostic of their subsequent target retrieval attempt at the time of test (Finn & Metcalfe, 2007; Koriat & Ma'ayan, 2005). However, there is less agreement about what information people use to make immediate JOLs. Information in short-term memory (Rawson & Dunlosky, 1992), an ease of processing evaluation based on ease of encoding (Begg, Duft, Lalonde, Melnick & Sanvito, 1989), or normative difficulty of the items (Koriat, 1997) have been proposed as candidates.

Most research favours the view that JOLs are heuristics which are inferential in nature (e.g. Koriat, 1997), and that they are based on a variety of cues pertaining to the task, and on an individual's belief about how these cues relate to subsequent test performance (Matvey, Dunlosky & Guttentag, 2001). Some cues are information-based, mediated by deliberate, analytic inferences that rely on retrieved memories. Other cues are experience-based, relying on a variety of internal, mnemonic cues which operate below full consciousness to influence and shape subjective experience (Koriat, 1997). One high impact cue is the ease or fluency of processing of a presented item (Begg et al., 1989). Because the inferential view holds that metamnemonic judgments are not based directly on memory traces, it predicts that JOLs may be systematically erroneous in some cases. Hence, it envisages situations in which judgments of memorability and later performance may be correlated negatively (Schwartz et al., 1997). Indeed, evidence supporting this view comes from observations documenting dissociations between subjective and objective indices of memory performance, such as poor JOL and accuracy calibration (Koriat & Ma'ayan, 2005).

The current study focuses on the notion of perceptual fluency as a mnemonic cue underlying metamnemonic judgments (Kelley & Jacoby, 1998; Koriat & Bjork, 2005). Perceptual fluency—the experience of fluent or facile processing—gives rise to predictions of future performance that are somewhat inflated (Benjamin et al., 1998). In some circumstances, using perceptual fluency as a heuristic for predicting future retention is sensible because more well-known material will tend to be recallable in the future. However, it is sometimes a poor predictor: For instance, Benjamin et al. (1998) have shown that more familiar studied material is processed with greater ease or speed. Learners use this speed as a heuristic to predict that the material is more likely to be recallable on a future test.

Consequently, they devote less study time to learning such material, misjudging their understanding of the relevant material (Koriat & Bjork, 2006). The same is true of the bias arising from word frequency (Begg et al., 1989; Greene & Thapar, 1994).

This body of research suggests that individuals predict mnemonic success for items that are easiest to process in the manner demanded by the task. Overcoming the influence of perceptual fluency requires a shift to an analytic mode of reasoning, and a correct attribution of the feeling of fluency to its cause (e.g. speed of processing). The drawback of using such heuristic judgments is that their correctness arises from associations that are correlational rather than causal. Specifically, differences in perceptual fluency do not *cause* items to be differentially memorable, though students studying for an exam may believe so. In fact, students' claims that they have studied sufficiently for exams should be treated with scepticism; the design of lecture and textbook materials which are initially easy to understand may actually hinder student learning. There are reliable findings suggesting that perceptual factors that *impede* fluency such as increasing difficulty or cognitive effort (Butler, Marsh, Goode & Roediger, 2006; Alter, Oppenheimer, Epley and Eyre, 2007) improve memory accuracy, by activating more analytic forms of reasoning during initial processing. Conceptual factors, too, such as varying the coherence of prose so as to increase readers' processing effort (Harten, 1999) improves the accuracy of assessment of their text comprehension.

The current studies investigate the generalisation of these findings to more naturalistic material: text extracts that might be used in the classroom. Currently, it is unclear whether the perceptual fluency effect persists with the complex text material that is routinely used in educational settings. Empirical studies supporting the fluency effect's existence have been readily applied to education, notwithstanding hefty criticisms launched over the ecological validity of the research paradigms employed (Koriat & Goldsmith, 1994). Principally, this criticism pertains to the use of list-learning and paired-associate tasks to investigate perceptual fluency, using simple, common, and impoverished material compared to text-based materials (Robinson, Johnson & Robertson, 2000). Generalisation to educational contexts requires the use of appropriate materials.

To date, the most common methodology psychologists have used to manipulate fluency within text is to degrade the grammatical structure, logical coherence, or physical presentation of the material. For instance, Hertzog et al. (2003), across four experiments, had participants read texts, predict their performance for each one, and take tests. Fluency was manipulated by having participants read texts that varied in coherence; the authors altered either: the causal relatedness across sentence pairs, the structure of sentences within paragraphs, or the structure of paragraphs by deleting letters from some words. Results showed that prediction magnitudes increased as coherence increased, suggesting that predictions were based on processing ease and fluency. Similarly, Harten's (1999) study manipulated the logical coherence of history texts, and found that JOL prediction magnitudes increased as the text's lucidity increased. Memory performance, however, was not consistent with the memory predictions in either of these experiments. In Harten's study, degrading textual coherence activated a more analytic form of reasoning and improved metamnemonic monitoring.

The approach adopted in the current study investigates the effect of perceptual fluency caused by a different kind of facility of processing from the structural ease studies described above. It is based on the notion that pre-existing knowledge, i.e. familiar material such as tangentially-related examples and anecdotes, can also lead to a misleading fluency about one's understanding of newly learned material. When people have to learn new technical information, of some degree of difficulty, extracted from high-quality tertiary-level textbooks, the presence of irrelevant but familiar material may lead them to believe they *understand* the key technical concepts in the text. Does old but irrelevant information lead to greater miscalibration of judgments and performance on newly learned material than the absence of such irrelevant information? Further, if such miscalibration occurs, is it more likely to occur when the irrelevant material is embedded within the new information, or when it is presented *en bloc* separately from it?

We expect, then, that if perceptual fluency caused by familiar but irrelevant material occurs, there will be poorer calibration between JOLs and subsequent memory accuracy for the exposed conditions, relative to a condition in which participants are not exposed to such material. Furthermore, when material can be more easily discounted, as when the irrelevant material is separated from the new

material, calibration may be less impaired than when the material is integrated with the to-be-learned information.

We also examine the effect of delayed or immediate JOLs. Consistent with previous research, we expect that when JOLs are made *immediately* after study, miscalibration will be more marked than when JOLs are provided after a 24-hour delay.

Finally, we also look at the depth of understanding of the new material, by comparing memory accuracy for more superficial aspects of the text, compared to deeper, more inferential understanding. Conditions that lead to fluency and thus to memory illusions may affect only deeper aspects of understanding, leaving memory intact where verbatim information is assessed.

EXPERIMENT 1

Method

Participants

Eighty English-speaking individuals (47 female; 33 male) took part in the experiment. Over half were undergraduate psychology students from The Australian National University who participated for course credit; the remaining 37 were recruited via advertisement on university community notice boards. Due to the textual complexity of experimental materials, the results of non-native English speakers were excluded from final analysis. Mean age was 23.3 years ($SD=7.506$), with a minimum age of 17 years, and a maximum of 61 years. 60 (75%) were currently studying at a tertiary level, 15 (18.8%) had completed a tertiary undergraduate degree, and 5 (6.2%) had completed only secondary level studies.

Materials

A set of six technical difficulty- and length-matched 'baseline' texts was compiled, adapted from introductory-level tertiary textbooks. The texts covered a range of different topics, including euthanasia, drug abuse, terrorism, nuclear energy, tsunamis, and climate change. Six corresponding comprehension tests were created, containing 10

questions each; tests consisted of 4 questions which could be answered verbatim by remembering words in the text; 4 comprehension questions whose answers could be found in the text, but which required a level of understanding of the material; and 2 high-level inference questions whose answers were not directly stated but required deductions to be drawn from the text. Participants were randomly allocated to one of three experimental conditions – control, integrated, and separated. The control group studied *only* base texts; the separated and integrated group studied texts which included the whole of the base texts plus extraneous information. This information consisted of easy-to-read familiar material, designed to elicit the perceptual fluency effect, gathered from magazines and popular websites, and was identical across experimental conditions. It included well-known topical examples of the text's issues, which would not, however, aid understanding of the text that was to be learned. For instance, in the text on Nuclear Energy and Radiation, the following anecdotal material was inserted: *Stories like 'Sadako and her Thousand Paper Cranes' show that the bomb's radiation still causes leukaemia almost 60 years after the bomb was dropped.* In the integrated condition, the extraneous material appeared interleaved appropriately among the sentences in the base text. In the separated condition, the extraneous material was presented in an isolated paragraph which preceded the base text.

Design and Procedure

This experiment was a 3 (perceptual fluency condition) x 2 (immediate or delayed JOL) between-subjects design. Participants were tested individually or in small groups. Prior to the experiment, participants were instructed that they were engaging in an experiment about students' learning strategies, which would involve studying six texts, making judgments about their likely performance on a subsequent test of the materials, and completing associated comprehension tests. Participants were informed that they could be tested on any of the material presented, to ensure that those in the separated condition did not discount or ignore the extraneous material. Text order was counterbalanced across conditions, to eliminate text or fatigue effects.

Condition		Immediate JOL		Delayed JOL	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Control	EOL	58.14	19.35	62.55	12.35
	JOL	63.93	15.08	52.77	10.69
	Accuracy	68.15	9.61	52.62	8.03
	Calibration	-4.23	14.70	-0.15	7.27
Separated	EOL	59.12	16.70	55.06	11.64
	JOL	60.90	10.12	37.24	18.81
	Accuracy	45.10	15.18	34.62	13.54
	Calibration	15.79	13.39	2.63	10.46
Integrated	EOL	60.51	6.48	63.29	16.62
	JOL	63.91	9.83	52.74	17.15
	Accuracy	54.04	10.56	34.29	14.08
	Calibration	9.87	8.83	18.45	13.76

Table 1: Mean percentage and standard deviations for EOLs, JOLs, accuracy and calibration scores for the three perceptual fluency conditions, calculated in both the immediate and delayed JOL conditions

Pilot testing revealed ten minutes of allocated study time was sufficient for participants to achieve approximately 50-60% accuracy on the subsequent test; hence, across conditions, participants were given ten minutes to study the base texts. Because of the extra material to be read in the integrated and separated conditions, these participants were given an additional three minutes' study duration, for a total of 13 minutes' study time.

Participants were allowed 30 seconds to scan each text, presented on a separate page of a booklet, before making an ease of learning judgment (EOL) on a scale of 0-100%, indicating how easy they believed the material would be to learn. They then studied the first text in the condition to which they had been assigned. Following this, half the participants (Immediate JOL group) provided an immediate judgment of learning (JOL), indicating on a scale of 0-100% how well they thought they would perform on an upcoming test of the material. After studying and providing EOLs and JOLs for three texts, participants completed the three associated comprehension tests, presented on separate pieces of paper with separate headings. Thereafter, they repeated the process with the remaining three texts. A maximum of ten minutes was allowed to complete each test. The remaining half of participants (Delayed JOL group) studied and made EOLs for the six texts consecutively, returning 24-hours later to provide delayed JOLs and complete comprehension tests.

Results

Completed tests were graded for accuracy according to a standard marking rubric; half marks

were awarded for answers which envisaged a clear understanding but were nonetheless incomplete.

A between-groups 2 x 3 factor *ANOVA* was conducted to explore the impact of perceptual fluency condition and retention interval (immediate or delayed JOL) on average EOL and JOL estimates, accuracy, and calibration.

Judgments of Learning: There was a statistically significant main effect for both perceptual fluency condition [$F(2, 77)=3.819, p=.026$, partial $\eta^2=.94$] and retention interval [$F(1, 79)=23.77, p=.000$, partial $\eta^2=.243$] on JOLs. Post-hoc comparisons indicated that the control group's mean JOL ($M=58.35, SD=14.03$) was significantly greater than the separated group's ($M=49.07, SD=19.09$), whilst the integrated group's mean JOL ($M=58.33, SD=14.83$) did not differ significantly from either the control or separated groups. Across the three fluency conditions, JOLs were greater when provided immediately ($M=62.91$) than when provided after a 24-hour delay ($M=47.58$).

Accuracy: Similarly, a statistically significant main effect for perceptual fluency condition [$F(2, 77)=22.041, p=.000$, partial $\eta^2=.373$] and retention interval [$F(1, 79)=32.093, p=.000$, partial $\eta^2=.303$] on accuracy scores was found. As might be expected, accuracy was poorer when 24 hours had elapsed ($M=40.51$) than when participants were tested within the same experimental session ($M=55.76$). Post-hoc tests revealed that mean accuracy for the control group ($M=60.39, SD=11.75$) was significantly greater than the separated group ($M=39.86, SD=15.07$) and the integrated group ($M=44.17, SD=15.81$).

Calibration: Of greater importance were the calibration scores. Calibration was calculated by subtracting accuracy scores from JOL estimates: perfect calibration was reflected by a score of zero, overconfidence by a positive score, and underconfidence by a negative score.

There was a statistically significant effect of fluency condition on calibration scores [$F(2, 77)=13.633$, $p=.000$, partial $h^2=.269$]. Post-hoc tests indicated that mean calibration for the control group ($M=2.04$, $SD=11.59$) was significantly better (closer to zero) than the separated group ($M=9.21$, $SD=13.55$) and the integrated group ($M=14.16$, $SD=12.14$), supporting our expectations.

A further analysis examined the difference in accuracy among the different types of test questions: verbatim, comprehension and inferential. There was a statistically significant difference for the proportion of high-level inference questions answered correctly across both the immediate and delayed JOL conditions [$F(2, 77)=19.588$, $p=.000$, partial $h^2=.34$]. Post-hoc comparisons indicated that the mean percentage of high-level inference questions answered by the control group ($M=59.96$) was significantly greater than in both the separated ($M=34.81$) and integrated groups ($M=43.00$, $SD=14.60$). The separated and integrated groups did not differ.

On verbatim questions, the control group ($M=66.21$, $SD=10.94$) also performed better than both the separated ($M=44.62$, $SD=15.86$) and integrated groups ($M=52.12$, $SD=11.96$) [$F(2, 77)=19.193$, $p=.000$, partial $h^2=.33$], which did not differ from each other.

Discussion

The results of this study failed to confirm the hypothesis that students exposed to tangentially-related information would exhibit higher JOL estimates relative to students *not* exposed to such material. No significant difference was found between the three fluency manipulations for mean JOLs, either in the immediate or delayed JOL conditions. This result contradicts findings from list-learning and paired-associates tasks, which generally indicate that participants exposed to perceptually fluent information make more inflated metamnemonic estimates than control subjects. One explanation for this finding is Vernon and Usher's (2003) anchoring hypothesis, which suggests that the

magnitude of metamnemonic judgments did not differ across the fluency manipulations because *initial* JOL estimates were used as an anchor point for *subsequent* judgments. Specifically, once participants provided an initial estimate—typically at an intermediate point on the scale between 0-100%—they used it as an anchor point, and adjusted their JOLs around this central position when making successive metamnemonic assessments (Serra & Dunlosky, 2005).

Importantly, the results of this study supported the hypothesis that students exposed to tangentially-related information would exhibit impaired calibration scores relative to students not exposed to such material. When participants provided JOLs immediately after study, those in the separated and integrated conditions exhibited poorer calibration (overconfidence) relative to the control group. This result demonstrates that the perceptual fluency effect *does* occur for complex text material, and that fluency is in fact *foe*, not *friend*, to learners. The outcome also strengthens previous evidence documenting the fluency effect in list-learning and paired-associates paradigms.

When JOLs were delayed by 24-hours, calibration improved markedly: although accuracy scores were naturally lower, there was a corresponding drop in JOL magnitude. Only individuals in the *integrated* condition exhibited significant overconfidence; by contrast, the control and separated conditions both exhibited near perfect calibration. The misleading influence of perceptual fluency effect persists only when the extraneous information is embedded within the to-be-learned text. Why should participants be misled in both integrated and separated conditions when JOLs are provided immediately after study, but misled only in the integrated condition when they are provided after a 24-hour delay? One possibility is that a 24-hour delay allowed subjects to 'filter out' superfluous information when *separated* from the baseline text, whilst the subtle, persistent nature of the *integrated* extraneous material made it more difficult to disregard. What subjects remember could well be a text that is inferred to be more coherent in retrospect, leading to overconfident JOLs (Harten, 1999).

The finding that the control group correctly answered significantly more verbatim *and* high-level inference questions than either the separated or integrated conditions implies that the linking of new, technical information into a familiar context is only helpful and enhancing to memory when the familiar

Perceptual Fluency Condition		Experiment 2: Delayed JOL		Experiment 1: Delayed JOL	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Control	EOL	59.03	21.24	62.55	12.35
	JOL	48.00	20.75	52.77	10.69
	Accuracy	50.50	10.23	52.62	8.03
	Calibration	-2.50	15.53	-0.15	7.27
Separated	EOL	52.33	17.61	55.06	11.64
	JOL	46.67	19.70	37.24	18.81
	Accuracy	38.00	13.07	34.62	13.54
	Calibration	8.67	16.98	2.63	10.46
Integrated	EOL	57.40	18.67	63.29	16.62
	JOL	50.33	20.37	52.74	17.15
	Accuracy	38.50	8.33	34.29	14.08
	Calibration	11.83	16.92	18.45	13.76

Table 2: Mean percentage and standard deviations for average EOLs, JOLs, accuracy and calibration scores across the three perceptual fluency manipulations after a 24-hour delay

context is *relevant* and helps students *understand* the material at hand (Shaddock & Carroll, 1997). If the examples are not helpful in comprehension of new material they serve only as a vehicle to mislead the learner with respect to the degree of his or her comprehension. Genuine “deep” learning of a text means that individuals can infer new facts, and use the information in novel ways (Kintsch, 1994). Control subjects in both the immediate and delayed conditions accurately assessed what they knew; they were able to both infer and remember the surface forms of the texts. Those misled by fluency were not only misled about their deep understanding of the texts but had not even achieved sufficient memory for the superficial aspects.

EXPERIMENT 2

Research indicates that within-subjects designs are more sensitive to differences in metamemory measures than between-subjects designs (Shaddock & Carroll, 1997). Nelson (see Carroll and Nelson, 1993) used a metaphor to describe the different encoding strategies participants in the two designs might use: within-groups subjects ‘turn the microscope to low power’ to focus on large, salient differences between conditions, whilst between-groups subjects ‘turn the microscope to high power’ so smaller differences between items become more salient than larger, contextual ones. In Experiment 1, no differences in JOL between conditions were evident, even though miscalibration occurred in the experimental groups. Differences in the influence on JOLs of perceptual fluency may be magnified when people have an opportunity to learn under all three conditions of Experiment 1. Thus Experiment 2 seeks to confirm the miscalibration found in that

experiment, but also investigates whether the differences in memory monitoring judgements can be detected in the more sensitive paradigm in which an individual can make judgments under all three conditions.

Method

Participants

Fifteen English-speaking individuals who had not taken part in Experiment 1 participated (8 female; 7 male). All participants completed pre-tests measuring age, level of education, gender, and first language; results of non-native English speakers were excluded from final analysis. Mean age was 27.3 years ($SD=9.208$), with a minimum age of 19 years, and a maximum of 57 years. Six (40%) were currently studying at a tertiary level, 9 (60%) had completed a tertiary undergraduate degree.

Design and Procedure

This experiment employed a within-subjects design: each participant studied two texts under each of the three fluency conditions of Experiment 1: control, integrated and separated. Because Experiment 1 results indicated that delayed JOLs yielded better calibration rates than immediate JOLs in two of the three conditions, Experiment 2 was a delayed-only JOL experiment.

Participants were tested individually or in small groups, and were provided with initial instructions identical to Experiment 1. Text order and condition were counterbalanced, eliminating text or fatigue effects. As in Experiment 1, participants were given ten minutes to study each baseline text, and an additional 3 minutes to study the additional material, in the integrated and separated conditions.

Participants were given each text, allowed 30 seconds to ‘cast their eyes’ over the materials, and then instructed to make an EOL, before studying the text for the appropriate duration depending on the condition presented. This process was repeated for six texts; two from each of the control, integrated and separate conditions. Participants returned 24 hours later to provide JOLs and complete comprehension tests. Participants were allowed a maximum of ten minutes to complete the tests, in which they were denied access to the studied materials.

Results

One-way repeated measures ANOVAs were conducted on EOLs, JOLs, accuracy and calibration. The mean and standard deviations for EOL, JOL, accuracy and calibration are presented in the left half of Table 2, with corresponding means for Experiment 1 provided for comparison.

JOL: There was no statistically significant difference between the three conditions in mean JOLs.

Accuracy: The control condition’s mean accuracy score ($M=50.50$, $SD=10.23$) was greater than both the separated ($M=38.00$, $SD=13.07$) and integrated conditions ($M=38.50$, $SD=8.33$). [Wilks’ $\Lambda=.30$, $F(2, 13)=15.141$, $p=.000$, partial $\eta^2=.70$].

Calibration: As in Experiment 1, when participants studied the control text, they exhibited near perfect calibration after a 24-hour delay, with slight underconfidence. However, calibration scores showed that overconfidence was exhibited by participants in *both* the separated and integrated conditions after a 24-hour delay. F statistics confirmed a significant effect at the $p<.001$ level for fluency condition on mean calibration scores, [Wilks’ $\Lambda=.352$, $F(2, 13)=11.951$, $p=.001$, partial $\eta^2=.65$]. This latter finding differs from Experiment 1 results, which suggested that after 24-hours, poor calibration persisted, and overconfidence even *increased*, for the integrated group, whilst the separated group’s calibration improved to near perfect levels.

Discussion

Contrary to expectations, Experiment 2 showed that the memory monitoring measures did not differ among the three conditions even when a more sensitive within-subjects design was used. When

participants made judgments under conditions of perceptual fluency that were designed to mislead them into believing they had understood the texts, they failed to make judgments of higher magnitude than those given to the base texts. Thus, fluency apparently does not increase the metamemorial judgements above the rate of texts that do not contain fluency-inducing material. The magnitude of these judgments was also comparable to those of Experiment 1. As in Experiment 1, it appears that the use of an anchor is employed in both the within and between subject designs: it appears that participants in Experiment 2 made EOL and JOL estimates by choosing an initial anchor point on the scale between 0-100%, and based all subsequent judgments around that value (Serra & Dunlosky, 2005). Given that, in both experiments, judgments were made on the basis of whole texts, it may be that the within subject design is more sensitive to conditions only when individuals make item by item judgments.

As in Experiment 1, accuracy scores in the two experimental conditions were poorer than for the control, leading to significant miscalibration. Surprisingly, however, Experiment 2 participants exhibited significant negative calibration in *both* the separated and integrated fluency conditions after a 24-hour delay. This suggests that participants in the within-subjects design were affected differently by the 24-hour delay than those in the between-subjects design. Perhaps participants in the between-subjects design were able to disregard the irrelevant information in the separated condition over the 24-hour delay, whilst those in the within-subjects design were unable to engage in such a filtering process, because, having been exposed to all three fluency conditions, they were not as certain about which information would be redundant for testing.

An alternative explanation for the finding of generally poor accuracy in the experimental conditions in both experiments is that there was more material to remember overall in these two conditions than in the control conditions. The addition of extra study time for the two experimental conditions was designed to rule out any simple explanation for poorer accuracy based on sheer memory load. Nor did participants appear to perceive any extra memory demands compared to the control texts when they made their EOL judgments prior to study. These judgements would presumably have taken into account any readily perceived differences in the lengths of text, along with differences in comprehensibility.

GENERAL DISCUSSION

This research demonstrates that the perceptual fluency effect *does* occur for complex, real-world text material, implying that fluency, when it consists of familiar but irrelevant material, is *foe*—not *friend*—to learners who are attempting to understand new technical material. Students in real-life learning situations may be unable effectively to filter out such irrelevant information; instead, the fluency engendered by the extraneous material misleads them to believe that they have mastered the new information to the same extent as when such extraneous information is absent. They are misled because their memory accuracy is significantly poorer than when only the new information is presented.

We find support for earlier research (Koriat, Bjork, Sheffer & Bar, 2004; Benjamin & Bjork, 1996; Benjamin, Bjork & Schwartz, 1998) that ease of access effected by certain variables leads learners to rely on the implicit heuristic that what is easy to process *now* will be well recalled *later*. The variable manipulated here is familiarity, operationalised as already known irrelevant material that is either embedded in, or that precedes, complex new to-be-learned material. Because this study suggests that fluency's effects persist for 24-hours, and that the presence of this irrelevant material reduces a learner's ability to answer correctly *both* verbatim and inference-style questions, it is important that educators pay particular attention to the *type* of

examples they use in their course material. Although much research remains to be done concerning beneficial or detrimental elements of fluency in real-life contexts, it is clear that the inclusion of some material such as that used here may lead to an illusion of knowing. Further research might investigate, for instance, whether other extraneous material, such as the use of jokes and irrelevant anecdotes during lectures, devices designed to capture attention and to engage students, have the same effect. Clearly, the inclusion of *relevant* examples, which induce deeper elaborative understanding, may lead to a feeling of fluency that is entirely justified and which leads to accurate memory outcomes.

Other avenues for future research concern a potentially educable component of metacognition. Providing students with a rudimentary understanding of how fluency can hinder memory has the potential to prevent the kinds of generalisations that mislead metamnemonic predictions (Benjamin, Bjork & Hirshman, 1998). Whilst research overcoming illusions of knowing is in its infancy, training learners to identify misleading metamnemonic cues in list-learning and paired-associates paradigms can prompt them to develop and use metacognitive skills to assess their knowledge more accurately (Renner & Renner, 2001). This research should now be extended to cover real-world text-based material, particularly since training for these skills has practical value in the classroom (Koriat & Bjork, 2006).

REFERENCES

- Alter, A.L., Oppenheimer, D.M., Epley, N., & Eyre, R.N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, 136, 569-576.
- Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, 28, 610-632.
- Benjamin, A.S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition*, 31, 297-305.
- Benjamin, A.S., & Bjork, R.A. (1996). Retrieval fluency as a metacognitive index. In L. Reder (Ed.), *Implicit Memory and Metacognition* (pp. 309-338).
- Benjamin, A.S., Bjork, R.A., & Hirshman, E. (1998). Predicting the future and reconstructing the past: A Bayesian characterisation of the utility of subjective fluency. *Acta Psychologica*, 98, 267-290.
- Benjamin, A.S., Bjork, R.A., & Schwartz, B.L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, 127, 55-68.
- Brown, A.L., Bransford, J.D., Ferrara, R.A., & Campione, J.C. (1983). Learning, remembering and understanding. In J.H. Flavell & E.H. Markman (Eds.), *Handbook of child psychology: Cognitive development*. New York: Wiley.
- Butler, A.C., Marsh, E.J., Goode, M.K., & Roediger, H.L. (2006). When additional multiple-choice lures aid versus

- hinder later memory. *Applied Cognitive Psychology*, 20, 941-956.
- Carroll, M., & Nelson, T.O. (1993). Effect of overlearning on the feeling of knowing is more detectable in within-subject than in between-subject designs. *American Journal of Psychology*, 106, 227-235.
- Dunlosky, J., & Nelson, T.O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when JOLs occur? *Journal of Memory and Language*, 33, 545-565.
- Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 238-244.
- Flavell, J.H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental enquiry. *American Psychologist*, 34, 906-911.
- Greene, R.L. & Thapar, A. (1994). Mirror effect in frequency discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 946-952.
- Harten, A.C. (1999). *An investigation of calibration of comprehension: Text processing variables that affect college students' evaluation of their comprehension*. Unpublished doctoral dissertation, University of Texas, Austin, USA.
- Hertzog, C., Dunlosky, J., Robinson, A.E., & Kidder, D.P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 22-34.
- Kelley, C.M., & Jacoby, L.L. (1998). Subjective reports and process dissociation: Fluency, knowing, and feeling. *Acta Psychologica*, 98, 127-140.
- King, A. (1991). Improving lecture comprehension: Effects of a metacognitive strategy. *Applied Cognitive Psychology*, 5, 331-346.
- Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, 49, 294-303.
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilisation approach to judgments of learning. *Journal of Experimental Psychology: General*, 126, 349-370.
- Koriat, A., & Bjork, R.A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 187-194.
- Koriat, A., & Bjork, R.A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 1133-1145.
- Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General*, 123, 297-315.
- Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language*, 52, 478-492.
- Koriat, A., Bjork, R.A., Sheffer, L., & Bar, S.K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, 133, 643-656.
- Matvey, G., Dunlosky, J., & Guttentag, R. (2001). Fluency of retrieval at study affects judgments of learning (JOLs): An analytic or nonanalytic basis for JOLs? *Memory & Cognition*, 29, 222-233.
- Mazzoni, G., Cornoldi, C., & Marchitelli, G. (1990). Do memorability ratings affect study time allocation? *Memory & Cognition*, 18, 196-204.
- Nelson, T.O., & Narens, L. (1992). Metamemory: A theoretical framework and new findings. In T.O. Nelson (Ed.), *Metacognition: Core Readings* (pp. 50-58). Sydney: Allyn & Bacon.
- Rawson, K.A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 69-80.
- Renner, K.H., & Renner, M.J. (2001). But I thought I knew that: Using confidence estimation as a debiasing technique to improve classroom performance. *Applied Cognitive Psychology*, 15, 23-32.
- Robinson, M.D., Johnson, J.T., & Robertson, D.A. (2000). Process versus content in eyewitness metamemory monitoring. *Journal of Experimental Psychology: Applied*, 6, 207-221.
- Serra, M.J., & Dunlosky, J. (2005). Does retrieval fluency contribute to the underconfidence-with-practice effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 1258-1266.
- Shaddock, A., & Carroll, M. (1997). Influences on metamemory judgments. *Australian Journal of Psychology*, 49, 21-27.
- Vernon, D., & Usher, M. (2003). Dynamics of metacognitive judgments: Pre- and postretrieval

mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 339-346