Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Memory and Language 52 (2005) 478–492
Journal of Memory and Language
www.elsevier.com/locate/jml

The effects of encoding fluency and retrieval fluency on judgments of learning ☆

Asher Koriat *, Hilit Ma'ayan

Department of Psychology, University of Haifa, Mount Carmel, Haifa 31905, Israel

Received 16 August 2004, revision received 16 December 2004
Available online 12 February 2005

Abstract

This study investigated the heuristic bases of judgments of learning (JOLs). JOLs were elicited either immediately after study or after a shorter or longer delay. In Experiment 1, the effects of encoding fluency (inferred from self-paced study time) on both JOLs and recall decreased with JOL delay, whereas those of retrieval fluency (inferred from the success and latency of pre-JOL retrieval) increased. In this experiment, JOLs (as well as recall) decreased with increasing study time, presumably under the heuristic that items requiring more time to study are less likely to be recalled. In contrast, in Experiment 2, in which study time was experimentally manipulated, JOLs as well as recall actually increased with study time. In both experiments JOLs increased with retrieval fluency. The results demonstrate that JOLs are based on the flexible and adaptive utilization of different mnemonic cues according to their relative validity in predicting memory performance.

Keywords: Metacognition; Judgments of learning; Delayed JOLs; Encoding fluency; Retrieval fluency; Heuristics

There has been a great deal of research in recent years on the bases of metacognitive judgments such as the judgments of learning (JOL) that are made during the study of new materials, the feelings of knowing (FOK) that sometimes accompany the failure to retrieve a solicited target from memory, and the subjective confidence in one's selected or produced answers (see Koriat & Levy-Sadot, 1999; Metcalfe, 2000). Early proposals subscribed to the direct-access view according to which such judgments are based on the direct monitoring of memory traces (e.g., Arbuckle & Cuddy, 1969; Cohen, Sandier, & Keglevich, 1991; Hart, 1965). In recent years, however, an alternative, cue-utilization view has been gaining impetus, according to which metacognitive judgments are inferential in nature, relying on a variety of internal, mnemonic cues that have some degree of validity in predicting one's performance. Much of the evidence in support of this view comes from observations documenting dissociations between subjective and objective indices of memory performance (e.g., Benjamin, Bjork, & Schwartz, 1998; Koriat, 1995). Such observations not only raise doubts about the central

☆ This research was carried out at the Institute of Information Processing and Decision Making, University of Haifa. It was supported by a grant from the German Federal Ministry of Education and Research (BMBF) within the framework of German-Israeli Project Cooperation (DIP). We are grateful to Rinat Gil for her help in conducting the experiments, and to Limor Sheffer for her advice in the statistical analyses.

* Corresponding author. Fax: +972 4 8249431.
E-mail address: akoriat@research.haifa.ac.il (A. Koriat).

0749-596X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2005.01.001

implication of the direct-access view, that metacognitive judgment are inherently accurate, but also disclose some of the mnemonic cues that may lead one's metacognitive judgments astray.

Several mnemonic cues have been mentioned in the literature as possible determinants of JOLs, FOK, and subjective confidence. These cues include the ease or fluency of processing of a presented item (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Koriat, Bjork, Sheffer, & Bar, 2004), the familiarity of the cue that serves to probe memory (Metcalfe, Schwartz, & Joaquim, 1993; Reder & Schunn, 1996; Son & Metcalfe, in press; Vernon & Usher, 2003), the accessibility of partial information about the solicited memory target (Dunlosky & Nelson, 1992; Koriat, 1993; Nelson, Narens, & Dunlosky, 2004) and the ease with which information comes to mind (Benjamin & Bjork, 1996; Kelley & Lindsay, 1993; Robinson, Johnson, & Herndon, 1997; Zakay & Tuvia, 1998).

Are all these cues alternative manifestations of one general dimension that may be termed "processing fluency?" Indeed, processing fluency figures as a central theoretical construct in the conceptualization of Jacoby and his associates as a determinant of different forms of subjective experience (Kelley & Rhodes, 2002). According to Jacoby and his associates, processing fluency can give rise to different types of phenomenal experience depending on whether it is attributed to the past (familiarity) or to particular attributes of the stimulus such as visual brightness, auditory clarity, duration, and the like.

In contrast, the position taken in this study is that different types of mnemonic cues might have distinct effects on subjective experience and on metacognitive judgments (see Schwarz, 2004; Whittlesea & Leboe, 2003). This position is in the spirit of the two-stage model of JOLs advanced recently by Son and Metcalfe (in press). In their study, participants learned a list of paired associates and then made JOLs after attempting to retrieve the target. Reaction time for making JOLs increased with decreasing JOLs, suggesting that learners relied on retrieval fluency as a cue for JOLs. When learners made JOLs without attempting to retrieve the target first, however, an inverted U function was found, suggesting that JOLs could also be based on pre-retrieval cues, possibly the familiarity of the stimulus that is used to probe memory.

In Experiment 1 we focus on the distinction between two types of cues that are assumed to affect JOLs—encoding fluency and retrieval fluency. Encoding fluency refers to the ease with which to-be-remembered items are mastered during study, whereas retrieval fluency refers to the ease with which they come to mind (see Benjamin & Bjork, 1996). We examine the hypothesis that these two types of cues make independent contributions to JOLs, and their relative dominance in affecting JOLs may differ depending on the phase at which JOLs are elicited. In Experiment 2 we examine the possibility that the same cue that is interpreted as reflecting encoding fluency can have the opposite effects on JOLs when it is interpreted differently. We now outline the procedure that was used in Experiment 1 and describe how encoding fluency and retrieval fluency were operationally defined.

The procedure of Experiment 1 is very similar to the Pre-judgment Recall And Monitoring (PRAM) methodology described recently by Nelson et al. (2004). That methodology was developed in order to allow a partition of the overall accuracy of JOLs into several components, and to examine an explanation of the delayed-JOL effect in which JOLs are more accurate when made at some delay after study than when made immediately after study (Dunlosky & Nelson, 1992, 1994; Nelson & Dunlosky, 1991). The procedure used in Experiment 1, in contrast, was primarily intended to separate between the contribution of encoding fluency and that of retrieval fluency, and therefore, as will be described below, involved a self-paced paradigm: Participants studied a list of paired-associates with the instruction to spend as much time as they needed on each item, and the time spent studying each item was measured. JOLs were solicited either immediately after study (Immediate, or I), a short time after study (Delay 1, D1), or a longer time after study (Delay 2, D2). For each of the three conditions, a pre-JOL recall test preceded immediately the solicitation of JOLs: Only the cue member of the pair in question was presented, and participants were required to recall the target. The JOL probe then appeared as soon as a response was produced or after 8 s had elapsed, requiring participants to judge the probability that they would recall the target at a final cued recall test to take place after the conclusion of the study phase. The latency for producing a pre-JOL recall was also measured.

We hypothesized that when JOLs are solicited immediately following study, they are influenced primarily by ease of learning. As JOL elicitation is delayed, a shift occurs towards greater reliance on retrieval fluency in making JOLs. The rationale for this prediction is that immediately after learning, the most accessible mnemonic cue concerns the effort invested in committing the item to memory. In contrast, a short while after, the memory for the effort invested in studying the item fades away and the most accessible mnemonic cue concerns the effort and success involved in attempting to retrieve the target.

The operationalization of pre-JOL retrieval fluency is relatively straightforward: Two measures will be used, one is accessibility, that is whether a target was accessed (regardless of its correctness), and the second is retrieval latency when a target was produced. The operationalization of encoding fluency, in contrast, requires some explanation. As a measure of the encoding fluency of an item we used the amount of self-paced study time spent in studying that item. This measure is based on evidence obtained by Koriat, Ma'ayan, and Nussinson (2005), suggesting that under typical self-paced conditions the amount of time invested in each item is data-driven, determined by the item itself in a bottom-up fashion. Hence study time can be used as a rough index of encoding fluency. Focusing on immediate JOLs, Koriat et al. presented evidence suggesting that learners also use study time (or study effort) as a cue for JOLs under the memorizing effort heuristic according to which the more effort is invested in studying an item the *less* likely it is to be recalled. They proposed that it is study time that affects JOLs rather than vice versa (cf. Nelson & Leonesio, 1988). Thus, our first prediction is that both encoding fluency and retrieval fluency affect JOLs. However, whereas the contribution of encoding fluency to JOLs is strong for immediate JOLs and decreases with JOL delay, the contribution of retrieval fluency to JOLs is small for immediate JOLs and increases as the elicitation of JOLs is delayed.

The second prediction concerns the validity of encoding fluency and retrieval fluency in predicting subsequent recall. We expect encoding fluency to have some predictive validity such that the more time is invested in an item the *less* likely it is to be later recalled. Retrieval fluency, on the other hand, should correlate positively with final recall and its predictive validity should increase with JOL delay. This is because with increased delay the conditions in which JOLs are solicited (and cued by the stimulus alone) approximate better the eventual criterion test (see Dunlosky & Nelson, 1997). Hence, the accuracy of JOLs in predicting subsequent recall should be mediated primarily by encoding fluency when JOLs are solicited immediately after study, but it should be mediated primarily by retrieval fluency when JOLs are delayed.

Experiment 1

Method

Participants

Twenty-seven Hebrew-speaking undergraduates (18 women and 9 men) participated in the experiment for course credit.

Materials

A list of 90 Hebrew word pairs was compiled, representing a wide range of degree of relatedness between the members of the pairs. For 45 pairs, the likelihood of the stimulus word eliciting the response word as a first associate (associative strength) was greater than zero according to Hebrew word-association norms (Rubinsten, Henik, Faran, & Drori, 2005). For these pairs, associative strength ranged from 0.98 to 94.12% (mean 14.36%). The remaining 45 pairs were selected such that the two members were unrelated: The stimulus words were taken from the norms, and each was paired with a response word that was not listed in these norms as a response. Effort was made to avoid obvious links between words that belonged to different pairs.

Apparatus and procedure

The experiment was controlled by a PC. In the study phase, participants were presented with the 90 paired-associates one after the other, and were instructed to study each pair so that they would be able to recall the second word when prompted by the first. Participants were told that when they had studied the pair well enough they should press the left mouse button. They were instructed that their success would be evaluated primarily by their ability to recall correctly as many words as possible, but that they should make an effort to spend as little time as possible in studying the entire list.

On each trial, the pair appeared at the center of the screen until the participants pressed the left mouse button key, at which time the pair disappeared from the screen. A pre-JOL recall was tested at one of three delays: Immediate (I), short delay (D1) or long delay (D2) with a third of the pairs assigned to each delay condition. For the I condition, the stimulus word was shown again immediately after the participant had pressed the left mouse button. For the D1 pairs, that word appeared equally often after 5, 7, or 9 study trials, whereas for the D2 pairs it appeared equally often after 20, 25 or 30 study trials. On average, the temporal interval between end of study and the prompt for pre-JOL recall was about 1, 60, and 215 s, for the I, D1, and D2 conditions, respectively. For all three delay conditions, participants were asked to recall the response word corresponding to the presented stimulus word, and their retrieval latency was measured by the computer using voice activation. As soon as a word was produced or after 8 s had elapsed, the following statement was presented: "Chances to recall: 0–100%." Participants were required to assess the chances that they would recall the second word at test when presented with the first word. The recall and JOL responses were typed by the experimenter.

The order of presentation of the stimuli and the assignment of each pair to its specific delay condition was determined randomly with the constraint that related and unrelated pairs were equally distributed among the three delay conditions. This random arrangement was constrained by the fact that the last 5 pairs were all in the I condition, and the D2 items could appear for study only during the first 70 trials. The assignment of pairs to the three delay conditions was counterbalanced across each group of three participants.

Following the study phase, participants were given a filler task that lasted for 1.5 min (counting backwards at intervals of 3, starting from a specified 3-digit number).

Table 1
Mean pre-JOL recall, JOLs, and final recall for the three delay conditions (Experiment 1)

Delay condition	Pre-JOL recall	JOL	Final recall
I	99.25	63.49	62.17
D1	77.33	63.36	70.23
D2	57.82	51.66	60.03

The final cued-recall test then followed: Each of the 90 stimulus words was presented for up to 8 s, and participants had to say aloud the response word within the 8 s allotted. The experimenter scored the response, and 1 s thereafter a beep was sounded and the next stimulus word was presented.

Results

Mean pre-JOL recall, JOLs, and final recall as a function of JOL delay

As might have been expected (see also Nelson et al., 2004), pre-JOL recall decreased monotonically with delay (see Table 1), $F(2,52) = 77.04$, $MSE = 153.57$, $p < .0001$. Recall was practically perfect at immediate testing (there were only two instances in which a participant failed to recall the target), but declined steeply with delay. Interestingly, this decline was not paralleled by a drop either in JOLs or in final recall. For final recall there was a trend for a curvilinear relationship with delay, so that the highest recall performance was actually at the intermediate delay (D1). A one-way ANOVA for the effects of delay on final recall yielded $F(2,52) = 9.83$, $MSE = 79.76$, $p < .001$. JOLs, in contrast, decreased somewhat with delay $F(2,52) = 17.87$, $MSE = 69.72$, $p < .0001$. [1] Note also the curious observation that for the D2 condition there was actually little drop from pre-JOL recall to final recall [in fact, there was a significant improvement, $t(26) = 3.23$, $p < .005$]. Across all delay conditions, JOLs and final recall averaged 59.59 and 64.14%, respectively, suggesting a certain degree of underconfidence, $t(26) = 2.13$, $p < .05$

The effects of encoding fluency on JOLs

Across all items participants spent an average of 7086 ms studying each item. Assuming that study time reflects encoding fluency or memorizing effort (Koriat et al., 2005), we should expect JOLs to decrease with increasing study time, particularly for the I condition. In that condition, the within-participant Pearson correlation between JOLs and study times averaged −.35, $t(26) = 5.96$, $p < .0001$: The more time invested in the study of an item the lower were JOLs for that item. This result was reliable: The correlation was negative for 22 out of the 27 participants, $p < .001$ by a binomial test.

As expected, the negative correlation between study time and JOLs dropped monotonously with delay, averaging −.19 and −.10, for the D1 and D2 conditions, respectively. Each of these correlations was significant (or near-significant), $t(26) = 4.35$, $p < .001$, and $t(26) = 2.01$, $p < .06$, respectively. However, a one way ANOVA comparing the correlations for the three levels of delay yielded $F(2,52) = 8.77$, $MSE = 0.05$, $p < .001$. The correlation for the I condition differed significantly from that of the D1 and the D2 conditions, $t(26) = 2.83$, $p < .01$, and $t(26) = 3.72$, $p < .001$, respectively. The D1–D2 comparison yielded $t(26) = 1.62$, $p < .12$. Thus, encoding fluency made a significant contribution to JOLs but this contribution decreased with increasing interval between the study phase and the elicitation of JOLs.

The effects of retrieval fluency on JOLs

Two indices were used to tap pre-JOL retrieval fluency. The first, which will be termed accessibility, is whether any answer was produced. Following Koriat's (1993) accessibility model, we assume that the accessibility of an answer contributes to the enhancement of JOLs regardless of whether that answer is right or wrong. Accessibility, then, was scored dichotomously such that an item received 100 when a response was produced and 0 when no response was produced. The second index is retrieval latency, which was measured when a response was produced; this is assumed to reflect the ease with which the solicited target comes to mind (see Kelley & Lindsay, 1993).

Focusing first on accessibility, this index was virtually perfect for the I condition, averaging 99.75, but dropped to 81.31 for the D1 condition, and to 66.29 for the D2 condition. These values actually represent the mean percentage of trials on which a response was produced. The output-bound accuracy of recall (see Koriat & Goldsmith, 1996) amounted to 99.49, 89.03, and 85.70% for the I, D1, and D2 conditions, respectively.

What was the contribution of accessibility to JOLs? A within-person gamma correlation between JOLs and accessibility averaged .41, .86, and .88, for the I, D1, and D2 conditions, respectively. The mean correlation for the I condition was based only on two participants, but the means for the D1 and D2 correlations were based on 25 and 27 participants, respectively. Consistent with predictions (see also Nelson et al., 2004), these latter two correlations were positive and highly significant, $t(24) = 18.74$, $p < .0001$, and $t(26) = 17.37$, $p < .0001$, respectively. Across the D1 and D2 conditions JOLs averaged 71.42% when a response was reported compared with 13.69% when no response was produced.

We turn next to the second index—retrieval latency. We shall use the term "commission latency" to refer to

[1] This trend may have simply stemmed from the uneven distribution of the I, D1, and D2 items across input serial positions. However, see Experiment 2.

the latency of producing a response. Because this index is available for only a selected subset of the items, we will also use a second index in which retrieval latency is set at 8000 ms for all trials in which no response was produced within the time allotted (8 s).[2] Unless specified otherwise, it is in this sense that "retrieval latency" will be used henceforth. The advantage of this index is that it allows us to compare the effects of encoding fluency and retrieval fluency across the entire set of items.

Commission latency increased with delay, averaging 957, 2183, and 2637 ms, for the I, D1, and D2 conditions, respectively, $F(2, 52) = 199.24$, $MSE = 101955.83$, $p < .0001$. The difference between the D1 and D2 latencies yielded $t(26) = 5.38$, $p < .0001$.

With regard to retrieval latency (with 8000 ms for omission responses), this index too increased with delay, averaging 974, 3236, and 4400 ms for the I, D1, and D2 conditions, respectively. Pearson correlations between retrieval latency and JOLs averaged −.11, −.69, and −.75, for the I, D1, and D2 conditions, respectively. These correlations were all negative and significant, $t(26) = 2.66$, $p < .05$, for the I condition, $t(26) = 15.33$, $p < .0001$, for the D1 condition, and $t(26) = 15.82$, $p < .0001$, for the D2 condition. Note that now the correlation for the I condition can be taken more seriously than it was the case for the accessibility index, and it would seem that retrieval fluency does make some contribution to JOLs even in the immediate condition. The correlation, however, increased substantially with delay: A one-way ANOVA for the effects of delay yielded $F(2, 52) = 90.54$, $MSE = 0.037$, $p < .0001$. A comparison between the I and D1 conditions yielded $t(26) = 10.96$, $p < .0001$, whereas that between the D1 and D2 conditions yielded $t(26) = 1.62$, $p < .12$. These results are consistent with the hypotheses that JOLs increase with increasing ease of accessing the target at the time of making JOLs, and that the contribution of retrieval fluency to JOLs increases the longer the interval between the study phase and the elicitation of JOLs.

Comparing the effects of encoding fluency and retrieval fluency on JOLs

To contrast more directly the effects of encoding fluency and retrieval fluency on JOLs, we plotted in Fig. 1 the mean Pearson correlations between JOLs and study time, on the one hand, and between JOLs and retrieval latency, on the other hand, as a function of delay. A two-way ANOVA on the means depicted in Fig. 1 indicated that the JOL-latency correlations were higher on average (mean −.52) than the JOL-study time correlations (mean −.21), $F(1, 26) = 62.75$, $MSE = 0.060$, $p < .0001$. However, the interaction was also highly significant, $F(2, 52) = 60.12$, $MSE = 0.051$, $p < .0001$. Particularly impressive is the crossover interaction obtained in comparing the I condition with the D1 and D2 conditions. For the I condition, JOLs were more strongly correlated with study time than with retrieval latency, $t(26) = 3.25$, $p < .005$. For the D1 and D2 conditions, in contrast, JOLs were more strongly correlated with retrieval latency than with study time, $t(26) = 8.74$, $p < .0001$, for D1, and $t(26) = 11.39$, $p < .0001$, for D2. The analyses just reported were repeated separately for the related and unrelated paired associates, and the same general pattern as in Fig. 1 was found for each of the two classes.

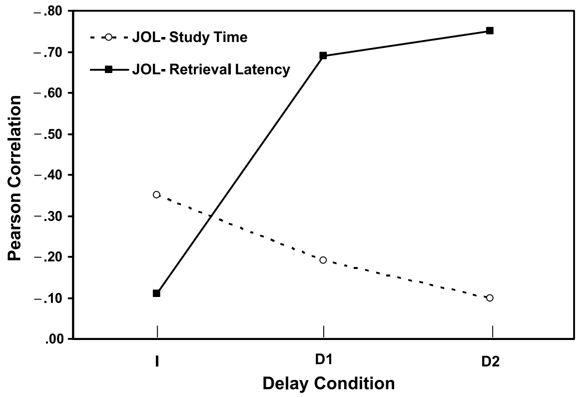

Fig. 1. Mean within-participant Pearson correlations between JOLs and study time, and between JOLs and pre-JOL retrieval latency, plotted as a function of delay condition (Experiment 1).

To round up the results for retrieval latency, we also calculated the correlations between commission latency and JOLs. The number of items over which this correlation was calculated varied between participants, averaging 29.07 out of 30 items (range 26–30) for the I condition, 23.19 items (range 14–30) for the D1 condition, and 18.41 items (range 9–28) for the D2 condition. The Pearson correlations for the three conditions averaged −.11, $t(26) = 2.49$, $p < .05$; −.42, $t(26) = 10.65$, $p < .0001$; and −.44, $t(26) = 10.10$, $p < .0001$, respectively. A one-way ANOVA yielded $F(2, 52) = 22.06$, $MSE = 0.042$, $p < .0001$. Thus, the results for commission responses also support the hypotheses that (a) JOLs increase with increasing ease of accessing an answer (see Matvey, Dunlosky, & Guttentag, 2001), and (b) reliance on retrieval fluency as a basis for JOLs increases with increasing delay.[3]

[2] It should be noted that the frequency distribution of commission latencies was positively skewed (Skewness = 1.91, Kurtosis = 3.78). About 99% of all responses were made within 6550 ms.

[3] Son and Metcalfe (in press) observed that when participants attempted to retrieve the response before making JOLs, response latency increased monotonically with JOL level, whereas when they only made JOLs, the relationship between JOL latency and JOL level was curvilinear. Thus, our results may be specific to a situation in which JOLs are made after attempted recall.

The relationship between encoding fluency and retrieval fluency

To examine the relationship between encoding fluency and retrieval fluency, we calculated the within-participant Pearson correlations between study time, on the one hand, and accessibility, retrieval latency, and commission latency, on the other hand, for each of the three delay conditions, and their means are presented in Table 2.

The results indicate that encoding fluency and recall fluency were positively correlated: The more time was invested in the study of an item the *less* likely it was to be recalled, and the *longer* was its latency of retrieval. Interestingly, the correlations between study time and retrieval latency tended to be stronger for the D1 delay than for the D2 delay. However, all the correlations in Table 2 are relatively low, yielding further support for the distinction between encoding fluency and retrieval fluency.

The joint effects of encoding fluency and retrieval fluency on JOLs

To examine the joint contribution of encoding fluency and retrieval fluency to JOLs, we carried out a regression analysis for each participant for each delay condition, using study time and retrieval latency as predictors, and JOLs as the dependent variable. The means and standard deviations of the standardized weights of the two predictors, as well as the proportion of explained variance (R^2) by each predictor are presented in Table 3 for each of the three delay conditions.

The results for the R^2 values reflected the general pattern of study time making a stronger contribution to JOLs in the immediate condition, but retrieval latency making a stronger contribution in the two delayed conditions. Thus, one-way ANOVAs confirmed that delay interval exerted significant effects on the R^2 values for both study time, $F(2,52) = 10.05$, $MSE = 0.017$, $p < .0001$ and retrieval latency $F(2,52) = 89.25$, $MSE = 0.028$, $p < .0001$.

Did each of the two predictors make a contribution to JOLs over and above that of the other predictor? To examine this question, we estimated for each participant, in each delay condition, the proportion of variance in JOLs that was explained by study time and retrieval latency combined. This variance averaged .26, .54, and .64, for the I, D1 and D2 conditions, respectively. Thus, overall prediction of JOLs improved with delay interval. By and large, however, combining the two indices did not improve greatly the prediction over and above that achieved by only one of them—study time in the I condition (.26 vs. .22) and retrieval latency in the D1 and D2 conditions (.54 vs .52, and .64 vs. .63, respectively).

In sum, the predictability of JOLs by both encoding fluency and retrieval fluency increased with delay. By and large, however, there was a relatively clear split between the two cues such that JOLs were almost entirely affected by encoding fluency in the immediate condition and by retrieval fluency in the delayed conditions.

The cue validity of encoding and retrieval fluency as predictors of final recall

We shall turn finally to the validity of encoding and retrieval fluency in predicting final recall. Beginning first with encoding fluency, the gamma correlation between study time and final recall was −.17, $t(26) = 4.61$, $p < .0001$. When study times were split at the median for each participant, mean recall for below-median and above median study times averaged 68.78 and 59.75%,

Table 2
Mean within-participant Pearson correlations between study time and accessibility, study time and retrieval latency, and study time and commission latency, as a function of delay condition (Experiment 1)

Delay condition	Study time–accessibility	Study time–retrieval latency	Study time–commission latency
I	−.16 ($N = 2$)	.12 ($p < .05$)	.13 ($p < .01$)
D1	−.17 ($p < .005$)	.22 ($p < .0001$)	.21 ($p < .0001$)
D2	−.08 ($p < .08$)	.10 ($p < .05$)	.08 ($p < .09$)

Table 3
Means and standard deviations of standardized weights and explained variance (R^2) for study time and retrieval latency in predicting JOLs, for each delay condition (Experiment 1)

Delay condition	Standardized weights				R^2			
	Study time		Retrieval latency		Study time		Retrieval latency	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
I	−0.34	0.32	−0.08	0.22	.22	0.20	.06	0.07
D1	−0.03	0.14	−0.67	0.24	.09	0.10	.52	0.26
D2	−0.03	0.14	−0.74	0.25	.07	0.09	.63	0.24

respectively, $t(26) = 3.04$, $p < .01$. Thus, the more time was spent studying an item, the *less* likely was it to be recalled (see also Koriat et al., 2005). This result was quite reliable: The correlation was negative for 22 out of the 27 participants, $p < .001$ by a binomial test. We should note that encoding fluency was also predictive of pre-JOL recall: The correlations for the I, D1, and D2 conditions were −.06 ($N = 5$), −.15 ($N = 25$), and −.12 ($N = 27$), respectively.

Let us turn next to the predictive validity of retrieval fluency, beginning with accessibility. Focusing on the D1 and D2 conditions, the gamma correlations between recall accessibility and final recall averaged .92 for the D1 condition and .96 for the D2 condition, $t(24) = 42.22$, $p < .0001$, and $t(26) = 76.83$, $p < .0001$, respectively. The conditional probabilities of final recall given that an answer was produced or was not produced at D1 were .83 and .14, respectively. The respective probabilities for the D2 condition were .84 and .11, respectively.

The correlations between target accessibility and final recall are open to the interpretation that in the delayed conditions the retrieval of the target prior to providing JOLs improves future recall due to spaced study (Kimball & Metcalfe, 2003; Spellman & Bjork, 1992). Note that accessibility was defined irrespective of the correctness of the response. When accessibility was defined in terms of the retrieval of the *correct* response (see Nelson et al., 2004), the correlation with final recall was even higher: The probabilities of final recall given that an answer was correctly recalled or not correctly recalled at the D1 interval were .91 and .13, respectively. The respective probabilities for the D2 condition were .96 and .14.

However, the second index of retrieval fluency—retrieval latency—is perhaps less vulnerable to the interpretation of the recall-fluency correlation in terms of the "memory hypothesis" (Kimball & Metcalfe, 2003). In fact, Benjamin et al. (1998) found that the longer it took participants to retrieve an answer to a question, the *higher* was their likelihood of recalling that answer at a later time. In this study, in contrast, the gamma correlations between commission latency and final recall were *negative*, averaging −.07 for the I condition, $t(25) = 1.22$, ns, −.50, for the D1 condition, $t(22) = 9.16$, $p < .0001$, and −.53, for the D2 condition, $t(22) = 6.50$, $p < .0001$.

A one-way ANOVA for the effects of delay (using only 20 participants, because 7 participants exhibited perfect recall) yielded $F(2, 38) = 10.59$, $MSE = 0.09$, $p < .001$. Thus, when the analyses were confined to trials in which participants did produce a response, the results indicate that retrieval latency is a diagnostic cue for final recall, and its diagnostic value increases with JOL delay.

We repeated the analysis reported above across all items, with retrieval latency set at 8000 ms for omission responses. The correlations of retrieval latency with recall averaged −.07 for the I condition, $t(25) = 1.24$, ns, −.74, for the D1 condition $t(25) = 15.14$, $p < .0001$, and −.87, for the D2 condition, $t(26) = 34.13$, $p < .0001$. A one-way ANOVA for the effects of delay (using only 25 participants, because 2 participants exhibited perfect recall) yielded $F(2, 48) = 77.21$, $MSE = 0.06$, $p < .0001$.

What was the overall accuracy of JOLs in predicting final recall? The gamma correlation between JOLs and final recall (with $n = 20$) averaged .43 for the I condition, $t(19) = 6.88$, $p < .0001$, .59, for the D1 condition, $t(19) = 7.29$, $p < .0001$ and .70, for the D2 condition, $t(19) = 9.98$, $p < .0001$. The increase in accuracy with delay replicate the delayed-JOL effect (Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991): A one-way ANOVA on these means yielded $F(2, 38) = 5.00$, $MSE = 0.07$, $p < .05$.

Summary

Experiment 1 supported the distinction between encoding fluency and retrieval fluency as potential cues for JOLs. In general, the two cues were positively correlated such that items that were more fluently encoded were more fluently retrieved. However the correlation was quite low. Furthermore, whereas the effects of encoding fluency on JOLs decreased with delay, those of retrieval fluency increased. Importantly, the effects of delay on cue utilization paralleled in part those obtained on cue validity: For immediate JOLs encoding fluency had a stronger effects on JOLs and was also a better predictor of final recall than retrieval fluency, whereas for delayed JOLs, it was retrieval fluency that affected JOLs more strongly and was also the better predictor of final recall. These results may have some bearing on the question of the "choice" of cue for JOLs (see General discussion).

Experiment 2

Experiment 2 was similar in design to Experiment 1, with the important change that presentation time was experimentally determined rather than self-regulated. This change was expected to have little effect on the results pertaining to retrieval fluency: As in Experiment 1, we expect JOLs to correlate positively with retrieval fluency, with the magnitude of this correlation increasing with increasing JOL delay. In parallel, the validity of retrieval fluency in predicting final recall should also increase with delay. In contrast, the effects of presentation time should differ markedly from those of Experiment 1. Recall that in Experiment 1 we expected JOLs to decrease with study time under the assumption that self-paced study time is diagnostic of the ease of encoding of studied items, and that JOLs increase with ease of encoding. When study time is experimenter-determined, however, JOLs should actually increase rather than decrease with study time in the same way that they increase with other factors that affect recall (e.g., Dunlosky & Nelson, 1994; Koriat, 1997; Mazzoni, Cornoldi, & Marchitelli, 1990; Shaw & Craik, 1989). Indeed, this is the pattern that was obtained in Koriat's study (1997, Experiment 4) when presentation time was experimentally manipulated and varied between 2 and 8 s per item within the list. In addition, we do not expect any systematic decrease in the contribution of presentation time to JOLs, unlike what was found in Experiment 1, when study time was self-determined. These predictions, if borne out, should reinforce the assumption underlying Experiment 1, that self-paced study time is diagnostic of encoding fluency. Furthermore, if presentation time in Experiment 2 is found to have the opposite effects on JOLs from what was found in Experiment 1, this would imply that exposure time may have different effects on metacognitive judgments depending on whether it is self- or other-regulated.

Method

Participants

Twenty-four Hebrew-speaking undergraduates (18 women and 6 men) were paid for participating in the experiment.

Materials

The list of paired associates was the same as that of Experiment 1 except that 6 experimental items were added (for counterbalancing considerations). In addition, 24 filler items were also included to permit an equal distribution of I, D1, and D2 items across ordinal positions.[4] Thus, the list consisted of a total of 120 paired-associates.

Design and procedure

The design of Experiment 2 was similar to that of Experiment 1 except that presentation time was experimenter controlled: Half of the pairs were presented for 3 s and half for 12 s each, with the assignment of items to each presentation time counterbalanced across participants. Note that 3 and 12 s represent about −1 STD and +1 STD, respectively, from the mean self-paced study time in Experiment 1. In addition, the assignment of I, D1, and D2 items to different ordinal positions was modified so that the three types of items were now equally represented in each ordinal position.

For the D1 pairs, the prompt for pre-JOL recall occurred exactly after 5 trials whereas for the D2 pairs it occurred exactly after 20 trials. To achieve this distribution, all filler items occurred after all experimental items had been studied. Half of the filler items were assigned to the I condition and half were assigned to the D1 condition. On average, the temporal interval between the end of study of an item and the recall prompt was approximately 1, 45, and 185 s, for the I, D1, and D2 conditions, respectively.

Results

The effects of delay on pre-JOL recall, retrieval latency, JOLs, and final recall

Table 4 presents the means of pre-JOL recall, retrieval latency (with 8000 for omission responses), JOLs, and final recall as a function of both JOL delay and presentation time. We shall first examine the effects of delay. Percent pre-JOL recall decreased with delay, $F(2,46) = 108.33$, $MSE = 100.18$, $p < .0001$, similar to what was found in Experiment 1. Unlike in Experiment 1, however, this decrease was paralleled by a drop in JOLs from 59.12 to 46.88%, $F(2,46) = 17.63$, $MSE = 56.41$, $p < .0001$. Final recall, on the other hand, yielded a curvilinear trend, being highest for the D1 condition, $F(2,46) = 9.09$, $MSE = 71.18$, $p < .001$. Thus, the similar curvilinear trend that was observed in Experiment 1 does not seem to derive from the unequal distribution of I, D1, and D2 items across different ordinal study positions in that experiment. Several explanations for this curvilinear trend may be offered but additional data are needed before these explanations can be seriously considered.

The results also replicated the curious observation from Experiment 1: For the D2 condition, there was little drop from pre-JOL recall to final recall [In fact, as in Experiment 1, there was a significant increase, $t(23) = 2.74$. $p < .05$]. Also replicated is the underconfidence bias: Across all delay conditions, JOLs and final recall averaged 54.29 and 61.30%, respectively, $t(23) = 1.96$, $p < .07$.

The effects of presentation time on JOLs

We now examine the effects of presentation time on JOLs. As expected, unlike what was found in Experiment 1, JOLs now *increased* with study time (Table 4) from 50.91% for the 3-s presentation to 57.54% for the 12-s presentation, $t(23) = 4.27$, $p < .0001$. Furthermore, whereas in Experiment 1 the effects of self-paced study time on JOLs were strong for the immediate condition and quite weak for the delayed conditions, here, if anything, the pattern was reversed. Indeed, a Presentation Time × Delay ANOVA on JOLs yielded significant effects for presentation time, $F(1,23) = 22.20$, $MSE = 74.20$, $p < .0001$, for delay, $F(2,46) = 17.98$, $MSE = 112.86$, $p < .0001$, and for the interaction, $F(2,46) = 8.21$, $MSE = 29.07$, $p < .001$. The interaction possibly reflects the observation that the effects of presentation time were weakest for the I condition.

To describe the results in a format like that used in Experiment 1, we calculated the within-participant Pearson correlations between presentation time and JOLs for

[4] Only 20 filler items were actually needed; the additional four filler items were added by mistake.

Table 4
Mean pre-JOL, pre-JOL latency, JOLs, and final recall as a function of delay condition and presentation time (Experiment 2)

Delay condition	Presentation time(s)	Pre-JOL recall	Pre-JOL retrieval latency (ms)	JOL	Final recall
I	3	98.35	960.89	58.31	51.96
	12	99.22	1023.12	59.96	61.87
D1	3	68.09	3378.34	51.46	59.54
	12	81.89	2667.30	61.30	74.13
D2	3	48.91	4341.69	42.35	51.96
	12	63.18	3686.29	51.15	67.27

each delay condition. These correlations were positive, averaging .04, .17, and .13, for the I, D1, and D2 conditions, respectively (based on 23 participants because one participant gave only 100% JOLs in the I condition). Only the last two correlations were significantly different from zero, $t(23) = 4.78$, $p < .0001$, for the D1 condition, and $t(23) = 4.57$, $p < .0001$, for the D2 condition. Recall that the respective correlations in Experiment 1 averaged −.35, −.19, and −.10.

The effects of presentation time on memory performance

In parallel to its effects on JOLs, presentation time also affected several aspects of memory performance. First, across the D1 and D2 condition, pre-JOL recall increased from 58.50% for the 3-s presentation to 72.54% for the 12-s presentation (pre-JOL recall was nearly perfect for the I condition). A Presentation Time × Delay (D1 vs. D2) ANOVA for the delayed conditions yielded $F(1,23) = 44.17$, $MSE = 107.08$, $p < .0001$, for presentation time, $F(1,23) = 40.65$, $MSE = 212.03$, $p < .0001$, for delay, and $F < 1$ for the interaction.

Second, retrieval speed was also similarly affected: Across all items, pre-JOL retrieval latency was shorter (3176 ms) for the 12-s than for the 3-s (3860 ms) presentation rate. As can be seen in Table 4, this effect too was observed only for the delayed conditions. A Presentation Time × Delay (D1 vs. D2) ANOVA yielded $F(1,23) = 22.76$, $MSE = 492{,}169$, $p < .0001$, for presentation time, $F(1,23) = 32.34$, $MSE = 728{,}973$, $p < .0001$, for delay, and $F < 1$ for the interaction.

Finally, final recall also increased with presentation time (Table 4): A similar ANOVA, as above, including all three delay conditions, yielded $F(1,23) = 56.08$, $MSE = 113.01$, $p < .0001$, for presentation time, $F(2,46) = 9.01$, $MSE = 140.22$, $p < .001$, for delay, and $F(2,46) = 1.40$, ns, for the interaction, It is interesting to note that the pattern reported by Koriat (1997, Fig. 9) of a weaker effect of presentation time on JOLs than on recall appears to hold most strongly for the immediate condition. The effect of presentation time on JOLs amounted to 1.65, 9.84, and 8.80, for the I, D1, and D2 conditions, respectively, whereas the respective values for recall were 9.91, 14.59, and 15.31.

The effects of presentation time on the ease and success of recall may be contrasted with the effects observed in Experiment 1, where study time correlated *negatively* with pre-JOL recall, speed of pre-JOL retrieval, and final recall. Note that the relationships observed in Experiment 1 were interpreted in the correlational sense, as indicating that items that are difficult to encode are also difficult to retrieve. In Experiment 2, in contrast, the similar relationship with presentation time should be interpreted as a causal relation: Presentation time affects future recallability.

In sum, unlike the pattern observed in Experiment 1, in which longer study time was generally associated with lower pre-JOL recall, longer retrieval latency, and lower rate of final recall, the results of Experiment 2 yielded a diametrically opposed pattern in which all measures of retrievability were correlated in the opposite direction with presentation time.

The effects of retrieval fluency on JOLs

We now examine the effects of retrieval fluency on JOLs. These effects were expected to be similar to those obtained in Experiments 1. We focus first on pre-JOL accessibility. This index averaged 99.2, 82.8, and 71.6%, for the I, D1, and D2 conditions, respectively. The output-bound accuracy of recall amounted to 99.6, 90.9, and 80.3%, respectively, for these conditions. As in Experiment 1, we calculated a within-participant gamma correlation between JOLs and accessibility. Pooling data across the two presentation times, this correlation averaged .84, .92, and .95, for the I, D1, and D2 conditions, respectively. The mean correlation for the I condition was based only on 4 participants, but the means for the D1 and D2 correlations were based on 23 participants each. Consistent with the results of Experiment 1 (and Nelson et al., 2004), the latter two correlations were highly significant, $t(22) = 24.50$, $p < .0001$, and $t(22) = 48.48$, $p < .0001$, respectively. Across the D1 and D2 conditions JOLs averaged 64.75% when a response was produced compared with only 7.28% when no response was produced.

Turning next to retrieval latency, retrieval latency increased with delay as in Experiment 1, averaging 992, 3014, and 4008 ms, for the I, D1, and D2 conditions, respectively, $F(2,46) = 134.25$, $MSE = 422{,}280$, $p < .0001$. The within-participant Pearson correlations between retrieval latency and JOLs averaged −.10, for the I condition, $t(22) = 1.73$, $p < .10$, −.75, for the D1 condition, $t(23) = 21.61$, $p < .0001$, and −.77, for the D2 condition,

$t(23) = 23.16$, $p < .0001$. A one-way ANOVA (based on 23 participants) for the effects of delay yielded $F(2,44) = 82.95$, $MSE = 0.040$, $p < .0001$. Whereas the comparison between the I and D1 conditions yielded $t(22) = 9.27$, $p < .0001$, that between the D1 and D2 conditions yielded $t(22) = 0.93$, ns.

We also calculated the correlations between commission latency and JOLs. Commission latency increased with delay, averaging 933, 1989, and 2415 ms, for the I, D1, and D2 conditions. The Pearson correlations for these conditions averaged −.05, $t(21) = 1.02$, ns, −.51, $t(22) = 11.15$, $p < .0001$, and −.55, $t(22) = 12.20$, $p < .0001$, respectively. (These correlations were based, respectively, on 29.3, 24.8, and 20.2 items on average). A one-way ANOVA (based on 22 participants) yielded $F(2,42) = 43.50$, $MSE = 0.039$, $p < .0001$.

In sum, the results for retrieval fluency agree with those found in Experiment 1, suggesting that JOLs increase with increasing ease of accessing the target at the time of making JOLs, and that the reliance on retrieval fluency as a basis for JOLs increases with the interval between the study phase and the elicitation of JOLs.

The effects of presentation time on JOLs: A search for the underlying mechanism

The contrast between the effects of study time (Experiment 1) and those of presentation time (Experiment 2) on JOLs invites an interpretation in terms of the dual-basis view of metacognitive judgments (Koriat et al., 2004; Koriat & Levy-Sadot, 1999). According to that interpretation, whereas the inverse relationship between study time and JOLs (Experiment 1) is due to a heuristic, mnemonic-based inference, the direct relationship (Experiment 2) reflects a theory-based judgment: Learners draw explicitly on the belief that items studied longer are more likely to be remembered than items studied for a shorter duration (along with the "total time hypothesis," Cooper & Pantle, 1967).

In contrast to this view, however, we explored the possibility that the effects of presentation time on JOLs in Experiment 2 also derive from a heuristic-based process, although this process now yields the opposite effect from that found in Experiment 1. Indeed, as can be seen in Fig. 2 (top panel), the effects of presentation time as well as delay (D1 vs. D2) on JOLs mimic closely the effects of these variables on retrieval fluency: JOLs as well as retrieval fluency (shorter latency) increased with increasing presentation time and decreased with delay, and the effects of presentation time and delay appear to be additive in both cases. A similar pattern for the effects of delay on JOLs and on retrieval fluency is also evident for the results of Experiment 1 (bottom panel).

Is it possible, then that the effects of presentation time in Experiment 2 are entirely mediated by the effects of presentation time on fluent retrieval? To examine this possibility, we considered only the D1 and D2 conditions for which retrieval fluency appears to be a diagnostic cue for JOLs. Across both of these conditions, we split retrieval latencies at the median for each participant and for each presentation time, and calculated mean JOLs for below-median (short) and above-median (long) retrieval latencies. The JOL means are plotted in Fig. 3 for the 3 and 12 s presentation conditions. As can be seen, JOLs decreased with increasing retrieval latency for each of the presentation times, and the functions relating JOLs to latency are almost perfectly superimposed on each other.

A similar analysis was carried out for the effects of delay (pooling data across all presentation times) and the results appear in Fig. 4 for both Experiment 2 (top panel) and Experiment 1 (bottom panel). One conspicuous feature in these figures is the clear separation between the results for the I condition and those for the D1 and D2 conditions, suggesting that retrieval fluency plays little role in mediating JOLs in the case of immediate JOLs. In contrast, the results for the D1–D2 comparison are clearly consistent with the hypothesis that the reduction in JOLs that occurs with delay can be explained almost entirely in terms of the corresponding reduction in the fluency of retrieving the target at the time of making JOLs.

In sum, the results presented above suggest that both the negative effect of study time (Experiment 1) and the positive effect of presentation time on JOLs reflect the operation of heuristic-based inference. However, while the effects of self-paced study time are mediated by an implicit interpretation of study time as an index of encoding fluency, the effects of experimenter-paced study time are mediated by retrieval fluency.

The accuracy of JOLs in predicting final recall

Across both presentation times, the within-person gamma correlation between JOLs and final recall averaged .47 for the I condition, $t(22) = 5.59$, $p < .0001$, .77, for the D1 condition, $t(23) = 23.08$, $p < .0001$, and .86, $t(23) = 28.58$, $p < .0001$, for the D2 condition. A one-way ANOVA (based on 23 participants because one participant gave only 100% JOL in the I condition) yielded a significant effect for delay, $F(2,44) = 16.91$, $MSE = 0.057$, $p < .0001$.

With regard to the cue validity of accessibility and latency, the gamma correlations between accessibility and final recall averaged .88 for the D1 condition and .97 for the D2 condition, $t(22) = 10.21$, $p < .0001$, and $t(22) = 80.92$, $p < .0001$, respectively. As for latency, the gamma correlations with final recall averaged .06 for the I condition, $t(23) = 1.06$, ns, −.69, for the D1 condition, $t(23) = 12.67$, $p < .0001$, and −.83, for the D2 condition, $t(23) = 28.84$, $p < .0001$. A one-way ANOVA for the effects of delay yielded $F(2,46) = 147.68$, $MSE = 0.037$, $p < .0001$. Thus, as in Experiment 1, accessibility as well

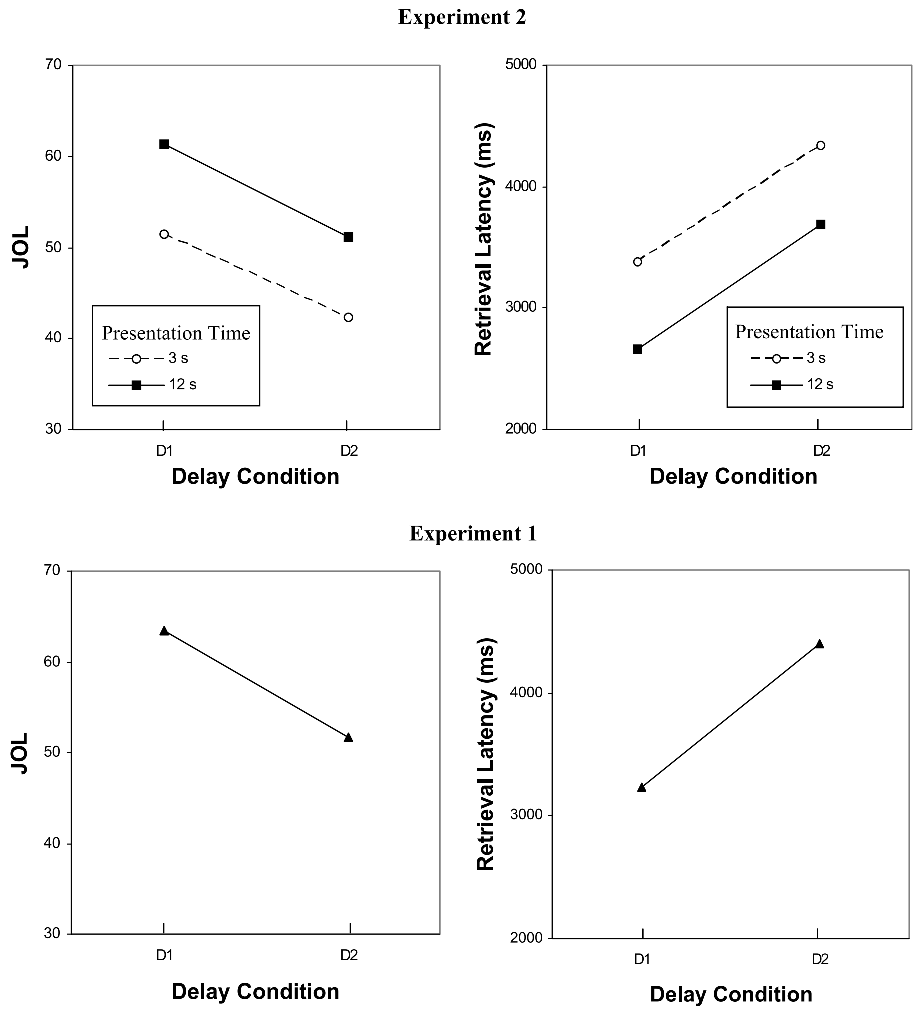

Fig. 2. The top panel presents mean JOLs (left panel) and pre-JOL retrieval latency (right panel) as a function of delay condition (D1 and D2) plotted separately for each presentation time (Experiment 2). The bottom panel presents mean JOL (left panel) and pre-JOL retrieval latency (right panel) as a function of delay condition (D1 and D2) (Experiment 1).

as retrieval latency were valid predictors of final recall, and their validity increased with JOL delay.

Discussion

Experiment 2 differed from Experiment 1 in that the allocation of study time to different items was experimenter determined rather than self-determined. This change had a dramatic effect on JOLs: Whereas in Experiment 1 JOLs decreased with increased study time, in Experiment 2 they increased with study time. The same pattern of diametrically opposed effects was observed for the three measures of memory performance: pre-JOL recall, retrieval latency, and final recall. These results testify for the flexible and adaptive nature of cue utilization. First, study duration appears to be interpreted differently when it is self-regulated than when it is experimenter regulated, and accordingly affect JOLs in opposite directions. Second, the different effects of study duration on JOLs mimic those observed for actual memory performance.

General discussion

Our investigation into the heuristic bases of metacognitive judgments during learning indicates that these judgments are finely tuned. First, Experiment 1 indicated that the relative sensitivity of JOLs to encoding fluency and retrieval fluency changes systematically with the phase at which JOLs are elicited. Second, Experiment 2 indicated that increased exposure duration can have diametrically opposed effects on JOLs, possibly depending on whether the learner has control over study duration

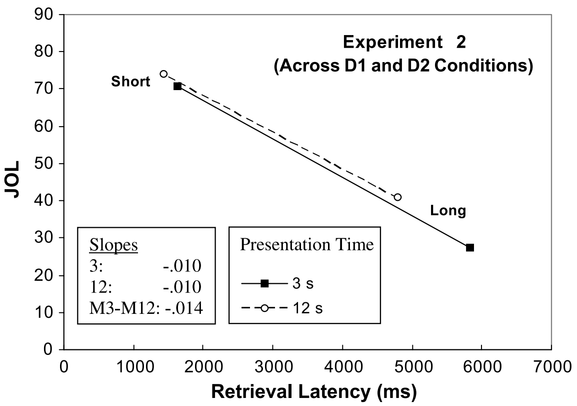

Fig. 3. Mean JOLs for below-median (short) and above-median (long) retrieval latencies (across D1 and D2 delay conditions), plotted separately for each presentation time (Experiment 2). The inset box indicates the slopes of the function relating JOLs to retrieval latency for the 3-s and 12-s presentation times, as well as the slope of the function relating mean JOLs to mean retrieval latency for the 3-s and 12-s conditions (M3–M12).

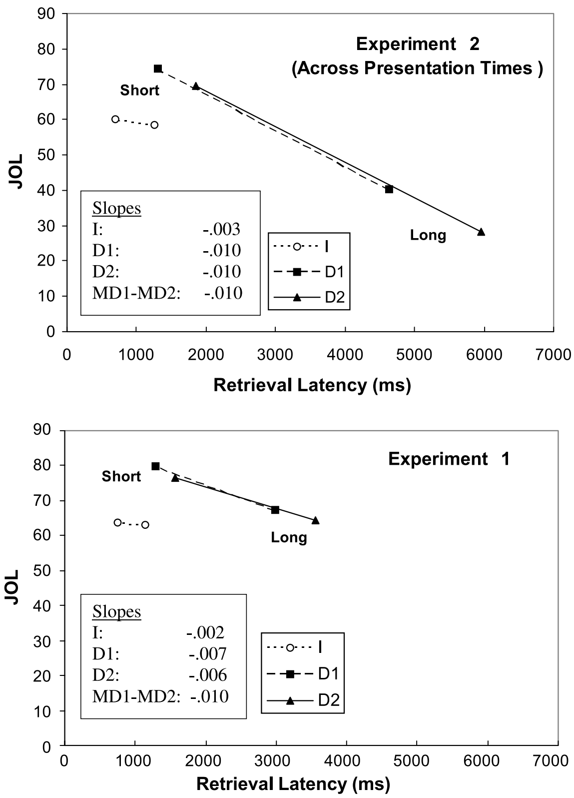

Fig. 4. Mean JOL for below-median (short) and above-median (long) retrieval latency, plotted separately for each delay condition. Top panel presents the results for Experiment 2 (across presentation times), whereas the bottom panel presents the results for Experiment 1. The inset box indicates the slopes of the function relating JOLs to retrieval latency for each of the delay conditions, as well as the slope of the function relating mean retrieval latency to mean JOLs for the D1 and D2 conditions (MD1–MD2).

and how the increased exposure is interpreted. In what follows we discuss these two findings in turn.

Cue utilization: The effects of encoding fluency and retrieval fluency on JOLs

Experiment 1 examined the usefulness of distinguishing between encoding fluency and retrieval fluency as potential cues for metacognitive judgments. Whereas previous discussions have focused either on the ease with which studied items are processed during encoding (e.g., Begg et al., 1989; Koriat et al., 2004), or on the ease with which to-be-remembered items are retrieved in the course of learning (e.g., Benjamin & Bjork, 1996; Benjamin et al., 1998; Matvey et al., 2001; Nelson et al., 2004; Son & Metcalfe, in press), none of these studies included both types of cues within the same experiment.

The use of self-paced study time as a measure of encoding fluency was based on the findings of Koriat et al. (2005) suggesting that in making recall predictions learners rely on the memorizing effort heuristic according to which items that are quickly mastered stand a better chance to be recalled than items that take longer to master. Indeed, JOLs in Experiment 1 decreased with increasing study time. Furthermore, recall performance too yielded an inverse relationship with study time, supporting the validity of the memorizing effort heuristic. The correlation between study time and JOLs may be mediated in part by the direct effects of judged a priori difficulty of the items. However, the decrease in the contribution of study time to JOLs with increased delay suggests that the effects due to judged a priori difficulty are not very strong. In addition, the observation that experimentally determined presentation time is actually positively correlated with JOLs (Experiment 2) suggests that it is not exposure duration per se that affected JOLs in Experiment 1, but possibly what study time (or effort) is taken to signify—difficulty of encoding.

As far as the effects of retrieval fluency are concerned, our results are consistent with Nelson et al.'s (2004) finding that the mere accessibility of a target before making JOLs makes a very substantial contribution to JOLs in the delayed conditions. As noted earlier, because Nelson et al. focused on the accuracy of JOLs, they stressed correct target recall rather than the mere accessibility of the target. However, although the majority of produced responses—about 90%—were correct, it is interesting to compare the contribution of correct commissions and commission errors to JOLs. Across the two delayed conditions, JOLs in Experiment 1 were significantly higher for correct commissions (74.38%) than for commission errors (38.67%), $t(23) = 7.89$, $p < .0001$. However, JOLs for commission errors were also higher than JOLs for omission responses (14.34%), $t(23) = 6.39$, $p < .0001$. A similar pattern was found for Experiment 2 where JOLs were higher for correct commissions (69.96%) than for commission errors (35.02%), $t(22) = 8.48$, $p < .0001$, and higher for commission errors than for omission responses (7.59%), $t(22) = 5.90$, $p < .0001$. Thus, although the mere accessibility of a response enhanced JOLs, learners could nevertheless discriminate between correct and incorrect commissions. One cue that they might have used to make such discrimination is retrieval latency. In Experiment 1, retrieval latency across the two delayed conditions averaged 2235 ms for correct commissions and 4111 ms for commission errors, $t(23) = 5.44$, $p < .0001$. The respective means in Experiment 2 were 1964 and 3368 ms, $t(22) = 6.34$, $p < .0001$. Thus, retrieval latency provides a refined basis for discriminating between retrieved items. Indeed, the results indicate that retrieval latency for commission responses also yields the same pattern as that observed for accessibility.

We should stress that our conclusions regarding the bases of JOLs were based on correlational data. Subsequent research must attempt to evaluate these conclusions with the help of experimental manipulations such as those used in cue-familiarity studies (e.g., Metcalfe et al., 1993).

Cue validity: The diagnostic value of encoding fluency and retrieval fluency

The assumption that JOLs rest on mnemonic cues implies that the accuracy of JOLs in predicting final recall depends on the predictive validity of the specific mnemonic cues on which they rest. In this study, an impressive correspondence between cue utilization and cue validity was observed, testifying for the efficient use of different mnemonic cues according to their predictive validity. First, like JOLs, encoding fluency was negatively correlated with final recall in Experiment 1: The more time was spent studying an item, the *less* likely it was to be recalled (see also Koriat et al., 2005). Second, in both Experiment 1 and Experiment 2, the likelihood of final recall increased with both indices of retrieval fluency—accessibility and speed of retrieval—and the correlations between final recall and retrieval fluency increased steadily with delay. In fact, whereas encoding fluency in Experiment 1 was a better predictor of recall than retrieval fluency at the immediate condition, retrieval fluency was a much better predictor than encoding fluency at the delayed conditions. This pattern mimics the one that was observed for cue utilization. Finally, as discussed later, the change from a self-paced to an experimenter-paced regime resulted in a change in cue utilization that mimicked the change in cue validity: JOLs as well as recall now increased rather than decreased with exposure duration.

"Choice" of mnemonic cues for JOLs

The correspondence between cue utilization and cue validity suggests that the inference rules underlying the use of mnemonic cues incorporate knowledge about the predictive validity of different cues under different conditions. How, then, does the learner "choose" which cue to use and how to apply it at any given condition?

Consider the flexible "choice" of mnemonic cues that is suggested by the results of Fig. 1. One possibility is that learners' choice is based on a strategic decision that is informed by explicit knowledge about the relative validity of different cues under different conditions. This possibility is consistent with the correspondence just noted between cue utilization and cue validity. We doubt, however, that participants have insight regarding the relative validity of different mnemonic cues under different conditions, and that their reliance on these cues reflects a premeditated, reasoned strategy. Benjamin et al. (1998) also argued against the possibility that the process by which participants make recall predictions incorporates an understanding of which subjective cues are diagnostic and when that diagnosticity may be compromised. An alternative account that we propose is that the cues that are chosen on each occasion are those that are the most subjectively accessible or salient: The salient cue immediately after study is the relative effort that was needed to commit the item to memory. After some delay, in contrast, the most salient cue is retrieval difficulty. The assumption is that learners are tuned to cues that disclose differences between items at the time of making JOLs (see Begg et al., 1989; Shaw & Craik, 1989).

Let us examine this account more closely with respect to immediate JOLs. Why does retrieval fluency have little effect on immediate JOLs? In discussing the delayed-JOL effect, Nelson and his associates (e.g., Nelson et al., 2004) explained that because practically all studied targets are still in short-term memory when JOLs are solicited immediately after study, their retrievability should be of little diagnostic value. This argument is certainly valid when the focus is on the accuracy of JOLs (as in Nelson et al., 2004). When it comes to the basis of JOLs, we would argue that participants do not rely on ease of retrieval in making immediate JOLs simply because they do not have the opportunity to experience it when the target is still in short-term memory. This is like a situation investigated by Kelley and Jacoby (1996): They found that after solving an anagram, participants were quite successful in predicting the difficulty that the anagram would pose to other participants, presumably relying on their own experience in solving the anagram. However, predictions were less accurate when made in the presence of the solution to the anagram, presumably because the solution deprived participants of the experience of solving the anagram. We propose that in like manner, the presence of the target in short-term memory in the immediate condition deprives learners of the experience of trying to retrieve it. Of course, the question remains why a "choice" that is based on the relative salience of different cues appears also to reflect the relative validity of these cues.

Achievement: The accuracy of JOLs in predicting final recall

We shall finally comment on the last component in Brunswick's (1956) conceptualization, which is referred to as achievement—the success of mnemonic-based JOLs in predicting recall. In Experiment 1, JOLs were accurate in predicting final recall even in the immediate condition, but they were clearly more accurate in the delayed condition, consistent with the delayed-JOL effect (Nelson et al., 2004). The increase in accuracy with delay is consistent with the observation that the predictive validity of retrieval fluency exceeded that of encoding fluency. Indeed, as noted earlier, the overall prediction of JOLs by both study time and retrieval fluency combined (R^2) increased with delay in Experiment 1, averaging 0.26, 0.54, and 0.64, for the I, D1, and D2 conditions, respectively. In parallel, the accuracy of JOLs in predicting final recall also increased: The respective gamma correlations averaged .43, .59, and .70, respectively.

These results may have some relevance to the question whether memory or metamemory changes with delay (e.g., Kimball & Metcalfe, 2003). They suggest that with increased delay a change occurs in the *basis* of JOLs in the direction of greater reliance on cues that are more diagnostic of recall. Thus, this change alone can explain part of the improved accuracy of JOLs with increased delay.

Self-regulated vs. experimenter-regulated study time

We turn now to the results of Experiments 2. What is the mechanism underlying the change that occurs from when study time is self-regulated to when it is experimenter-regulated? It may be argued that this change reflects a shift from experience-based JOLs to theory-based JOLs (Koriat, 1997; Koriat et al., 2004). In contrast, however, we proposed that both the decrease in JOLs with increasing study time (Experiment 1) and the increase in JOLs with increasing presentation time (Experiment 2) are mediated by a heuristic, experience-based process, except that self-paced study time is used to signal encoding fluency, whereas experimenter-paced study time affects JOLs through its effects on retrieval fluency. This implies that different intuitive theories may underlie cue utilization, depending on whether attention is directed towards encoding fluency or retrieval fluency. Indeed, several recent findings (see Schwarz, 2004, for a review) demonstrate that the effects of mnemonic cues on judgments vary depending on the naïve theory of mental processes that people bring to bear in each case.

Assuming that indeed the effects of both presentation time and JOL delay are mediated by retrieval fluency, a question that arises is why these effects are not discounted in making JOLs. Previous studies by Whittlesea and Leboe (2003) suggest that participants do not respond to the absolute magnitude of fluency but to the experienced fluency relative to what would be expected. The results of Experiments 1 and 2 are consistent with this suggestion with regard to the contrast between immediate and delayed JOLs. Thus, immediate JOLs were considerably lower than would be expected from retrieval latency (Fig. 4), suggesting that participants may have taken into account that to-be-remembered items are more fluently accessed immediately after study than later on. In contrast, although retrieval latency increased from the D1 to the D2 condition (Fig. 2), there was no indication that that increase was discounted in making JOLs. If such were the case, then the functions relating JOLs to retrieval latency (Fig. 4) should have been shifted upward for the D2 condition relative to the D1 condition. In a similar manner, in making JOLs, learners in Experiment 2 did not seem to take into account the decrease in retrieval latency with increasing time (Fig. 2). Rather, JOLs decreased with retrieval latency regardless of presentation time (Fig. 3). Perhaps the fact that presentation time and length of delay changed unexpectedly from trial to trial made it more difficult to partial out their effects on retrieval fluency in making JOLs.

In sum, the results presented in this study suggest that the monitoring of one's competence during learning involves a delicate use of encoding and retrieval fluency in a manner that reflects their relative validity in predicting memory performance.

References

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology, 81*, 126–131.

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., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. Reder (Ed.), *Implicit memory and metacognition* (pp. 309–338). Hillsdale, NJ: Erlbaum.

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.

Brunswick, E. (1956). *Perception and representative design in psychological experiments*. Berkley, CA: University of California Press.

Cohen, R. L., Sandier, S. P., & Keglevich, L. (1991). The failure of memory monitoring in a free recall task. *Canadian Journal of Psychology, 45*, 523–538.

Cooper, E. H., & Pantle, A. J. (1967). The total time hypothesis in verbal learning. *Psychological Bulletin, 68*, 221–234.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20*, 374–380.

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur?. *Journal of Memory and Language, 33*, 545–565.

Dunlosky, J., & Nelson, T. O. (1997). Similarity between the cue for judgments of learning (JOL) and the cue for test is not the primary determinant of JOL accuracy?. *Journal of Memory and Language, 36*, 34–49.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56*, 208–216.

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., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32*, 1–24.

Kelley, C. M., & Rhodes, M. G. (2002). Making sense and nonsense of experience: Attributions in memory and judgment. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 41, pp. 293–320). San Diego, CA: Academic Press.

Kimball, D. R., & Metcalfe, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Memory & Cognition, 31*, 918–929.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*, 609–639.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General, 124*, 311–333.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349–370.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133*, 643–656.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490–517.

Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 483–502). New York, NY: Guilford Press.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2005). *The intricate relationships between metacognitive monitoring and metacognitive control*. Manuscript in preparation.

Matvey, G., Dunlosky, J., & Guttentag, R. (2001). Fluency of retrieval at study affects judgments of learning (JOLs): An analytic or nonanalytical 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.

Metcalfe, J. (2000). Metamemory: Theory and data. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 197–211). London: Oxford University Press.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 851–864.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect". *Psychological Science, 2*, 267–270.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect". *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 676–686.

Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research on metamemory: Pre-judgment Recall And Monitoring (PRAM). *Psychological Methods, 9*, 53–69.

Reder, L. M., & Schunn, C. D. (1996). Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 45–77). Mahwah, NJ: Erlbaum.

Robinson, M. D., Johnson, J. T., & Herndon, F. (1997). Reaction time and assessments of cognitive effort as predictors of eyewitness memory accuracy and confidence. *Journal of Applied Psychology, 82*, 416–425.

Rubinsten, O., Henik, A. Anaki, D., Faran,Y., & Drori, S. (2005). Hebrew word association norms. In A. Henik, O. Rubinsten, & D. Anaki (Eds.), *Hebrew word norms*. Manuscript in preparation.

Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology, 14*, 332–348.

Shaw, R. J., & Craik, F. I. M. (1989). Age differences in predictions and performance on a cued recall task. *Psychology and Aging, 4*, 131–135.

Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science, 3*, 315–316.

Son, L. K., & Metcalfe, J. (in press). Judgments of learning: Evidence for a two-stage process. *Memory & Cognition*.

Vernon, D., & Usher, M. (2003). Dynamics of metacognitive judgments: Pre- and postretrieval mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 339–346.

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

Zakay, D., & Tuvia, R. (1998). Choice latency times as determinants of post-decisional confidence. *Acta Psychologica, 98*, 103–115.