Journal of Experimental Psychology:
Learning, Memory, and Cognition
2003, Vol. 29, No. 1, 22–34

0278-7393/03/$12.00 DOI: 10.1037/0278-7393.29.1.22

Encoding Fluency Is a Cue Used for Judgments About Learning

Christopher Hertzog
Georgia Institute of Technology

John Dunlosky
University of North Carolina at Greensboro

A. Emanuel Robinson and Daniel P. Kidder
Georgia Institute of Technology

The authors used paired-associate learning to investigate the hypothesis that the speed of generating an interactive image (encoding fluency) influenced 2 metacognitive judgments: judgments of learning (JOLs) and quality of encoding ratings (QUEs). Results from Experiments 1 and 2 indicated that latency of a keypress indicating successful image formation was negatively related to both JOLs and QUEs even though latency was unrelated to recall. Experiment 3 demonstrated that when concrete and abstract items were mixed in a single list, latency was related to concreteness, judgments, and recall. However, item concreteness and fluency influenced judgments independently of one another. These outcomes suggest an important role of encoding fluency in the formation of metacognitive judgments about learning and future recall.

Metacognition focuses on the interplay between processes of monitoring (evaluating the cognitive system) and control (selection and execution of cognitive mechanisms), with the assumption that monitoring can and should be used to adjust control processes to achieve desired goals (e.g., maximizing rates of learning; see Nelson, 1996). For example, when an individual is attempting to encode new material for an upcoming test, how does she or he monitor how well the material has been learned? Effective monitoring of learning is critical for achieving optimal control of subsequent study (Nelson & Narens, 1990; Thiede & Dunlosky, 1999). Accordingly, understanding how an individual assesses his or her own learning may provide insight into various aspects of self-regulated study, such as why self-paced study appears suboptimal under some conditions (Leonesio & Nelson, 1990; Mazzoni & Cornoldi, 1993; Thiede & Dunlosky, 1999). The present research empirically evaluated the hypothesis that the fluency of generating imagery mediators influences metacognitive judgments about learning, even when such fluency is not diagnostic of subsequent recall.

Christopher Hertzog, A. Emanuel Robinson, and Daniel P. Kidder, School of Psychology, Georgia Institute of Technology; John Dunlosky, Department of Psychology, University of North Carolina at Greensboro.

This research was supported by National Institute on Aging Grant R37 AG 13148. We thank Mahira Ali, Ben Blanton, Tara Fogle, and Mike Huston for their assistance in data collection and Amber R. York and C. Michael Herndon for assistance in preparation of the article.

Correspondence concerning this article should be addressed to Christopher Hertzog, School of Psychology, 274 Fifth Street, Georgia Institute of Technology, Atlanta, Georgia 30332-0170. E-mail: christopher.hertzog@psych.gatech.edu

Ease-of-Processing Hypothesis: The Basis of Judgments About Encoding

Hypotheses about how individuals judge memory for recently studied material typically focus on the possible bases of the judgments, or cues, that are available and accessible at the time the judgment is made (Koriat, 1997; Mazzoni & Nelson, 1995). Begg, Duft, Lalonde, Melnick, and Sanvito (1989) developed the ease-of-processing hypothesis to account for people's predictions of later memory performance. According to this hypothesis, a major basis of these judgments is "how easily the [judgment] task was done; this depends on preexperimental differences between items and task-specific differences caused by earlier events in the experiment" (Begg et al., 1989, p. 630). We presume a key idea here is that when the judgment is made immediately after the item is studied, the judgment task involves monitoring that occurs throughout the encoding of the item. For instance, if an individual is studying paired-associate (PA) items using interactive imagery, the judgments may be positively related to the speed with which an image is formed, how many images are generated during encoding, how easily the images are retrieved while making the judgment, and so on.

This hypothesis has been frequently cited as an explanation for variation in metacognitive judgments (see Koriat, 1997). For example, Kelemen and Weaver (1997) stated that Begg et al. (1989) "demonstrated that JOLs are also based on the ease of processing of paired associates at time of judgment" (p. 1395). Nevertheless, the empirical evidence supporting the hypothesis is often rather indirect, at best. Begg et al., Experiment 1, manipulated observable item characteristics (e.g., abstract vs. concrete items) and found that both memorability judgments and ease-of-studying judgments increased with increases in concreteness or imagery level. They also showed that predictive accuracy increased as similarity of the judgment task to the recall test increased (Begg et al., 1989, Experiments 2–4). In no case, however, did Begg et al. explicitly measure ease of processing, per se.

Other viewpoints can account for the empirical findings in Begg et al.'s (1989) experiments without reference to ease of processing as an explanation. Koriat's (1997) cue-utilization framework states that metacognitive judgments about encoding are often based on observable item characteristics (intrinsic cues) or encoding manipulations (extrinsic cues) that influence subsequent performance. Koriat's theory also predicts that judgments of learning can differ between abstract and concrete items because individuals regard concreteness as a stimulus feature that is relevant for subsequent learning. Note that this explanation requires no appeal to monitoring on line encoding processes or to the ease-of-processing information when making metacognitive judgments, as suggested by the ease-of-processing hypothesis.

In our view, the ease-of-processing hypothesis actually subsumes multiple possible mechanisms that could influence metacognitive judgments. In this article we focus our interest primarily on a single aspect of processing countenanced by Begg et al. (1989). Namely, we focus on what we term *encoding fluency*. Encoding fluency, for these purposes, is defined as the efficiency in execution of a chosen method of encoding to-be-learned materials under intentional learning instructions. The encoding fluency hypothesis states that the more quickly an encoding process is successfully concluded, the greater the subjective probability that this encoding will result in retention and subsequent recall. This hypothesis ignores other effects included in the ease-of-processing concept, including the ease with which a metacognitive judgment is made. Instead, the encoding fluency hypothesis concerns itself exclusively with monitoring the ease of execution of controlled processes at encoding as a basis for subjective estimation of item learning.

Evaluating the Encoding Fluency Hypothesis

What is needed to evaluate the encoding fluency hypothesis is a measure of fluency that is formally independent of subjective judgments and directly linked to on line encoding processes. An important parallel can be drawn to existing studies of retrieval fluency and its influence on metacognitive judgments. Benjamin, Bjork, and Schwartz (1998) operationalized retrieval fluency as the speed with which information is accessed from memory, the probability it is accessed, or both. In their Experiment 1, Benjamin et al. showed participants trivia questions and asked them to press a response key as soon as they had retrieved an answer. Response time for retrieval of answers to general knowledge questions was negatively correlated with predicted probability of future recall—faster retrieval was associated with higher predictions of future recall success (mean Goodman–Kruskal gamma correlation = −.35). Likewise, feelings of familiarity in recognition memory can be influenced by the fluency of retrieval of candidates from memory (e.g., Kelley & Jacoby, 1998; Kelley & Lindsay, 1993; see also Whittlesea & Williams, 2000, for a related but more complicated hypothesis incorporating fluency effects on subjective familiarity). Retrieval fluency is also regarded as an important cue in the source monitoring framework of Johnson and colleagues (e.g., Johnson & Raye, 2000).

To study encoding fluency, one must identify a target encoding process and measure its latency to completion. Matvey, Dunlosky, and Guttentag (2001) recently measured latency to generate targets in constrained cue–target pairs and compared learners (those who generated and recalled targets) with yoked observers who could observe learners' generation latencies. Latency to generate a target was an important predictor of learners' and yoked observers' judgments of learning (JOLs) about later cued recall (mean gamma of approximately −.3, averaging over three experiments).

We used a similar procedure in this study. Individuals studied PAs under explicit instructions to generate mediators using interactive imagery (Paivio, 1995; Richardson, 1998). Interactive imagery involves combining both the stimulus and response words in an integrated image that achieves a binding of the words. At test, PA recall is assessed by providing the stimulus word and measuring whether participants produce the associated response. Interactive imagery has been shown to increase the probability of mediator retrieval at test relative to other encoding strategies, including separate imagery, in which the two words are visualized in separate images (Begg et al., 1989; Bower, 1970; Yuille, 1973). In turn, successful retrieval of the image is highly associated with the probability of successful decoding of the response word from the retrieved image (Yuille, 1973).

Fluency of interactive image generation can be measured by having individuals press a response key as soon as an image has been formed, in an analogous manner to the measurement of retrieval fluency in other studies. Yuille and Paivio (1967) originally used this method and generated results consistent with the argument that keypress latencies were a valid measure of image-generation speed (e.g., latencies were longer for images based on abstract vs. concrete items). A critical advantage of this method is that this index of encoding fluency is not dependent on metacognitive judgments collected on the same study trial. The encoding fluency hypothesis predicts that latency to generate an interactive image will be negatively correlated with subjective quality of encoding and subjective probability of recall. Furthermore, the encoding fluency hypothesis stipulates that this effect will be observed irrespective of whether fluency is a valid cue for predicting recall success.

Experiment 1

Experiment 1 was designed to show that encoding fluency is a potentially powerful influence on metacognitive judgments linked to encoding processes. Our basic approach was simple: Individuals studied PAs one at a time. For each item, they pressed a key to indicate successful image formation. They then made a metacognitive judgment for that item.

We used two different kinds of judgment: JOLs and quality-of-encoding ratings (QUEs). JOLs have been widely studied in the metacognitive literature (Nelson, 1996). JOLs are typically scaled as percentage confidence that an item will later be recalled, although studies have also used ordinal Likert-type ratings (e.g., Rabinowitz, Craik, Ackerman, & Hinchley, 1982). JOLs made immediately after an item has been studied have been argued to reflect subjective monitoring of encoding processes (Koriat, 1997; Mazzoni & Nelson, 1995). QUEs have been less frequently studied in the literature (Dunlosky, Kubat, & Hertzog, in press), but they have the potential advantage of not requiring individuals to consider the possibilities of forgetting, test interference, and so forth, so that they may focus only on the nature of the encoding process itself. To the extent that encoding fluency is a potent influence on monitoring of encoding, one would expect that fluency would be correlated with both types of judgment but that correlations would be higher for QUEs than for JOLs. This hypothesis is based on the assumption that JOLs are also influenced by a participant's theory of retention and future recall performance, whereas QUEs need not consider factors related to retrieval at test.

One might also expect that encoding fluency would have maximal impact on metacognitive judgments if those judgments are collected immediately after the keypress to indicate image formation, but that the impact of fluency on the judgments would decrease as more time was allowed to elaborate on or rehearse the interactive image. To the extent that elaborative processing after initial image formation produces additional cues used to make the metacognitive judgments, we expected correlations between latencies and judgments to be attenuated after additional study time. We tested this hypothesis by manipulating the amount of additional study time provided to participants after they reported successful image formation.

Method

Participants. Participants were undergraduates at the Georgia Institute of Technology who received extra credit for their participation. We tested 121 individuals in Experiment 1 in groups of up to 4 persons. There were 61 participants in the JOL condition and 60 in the QUE condition.

Apparatus and materials. The experiment was a PA learning task programmed in HyperCard (Version 2.3, Apple Computer, 1995) and administered on Macintosh computers. All of the word pairs consisted of unrelated, concrete nouns. We used 66 pairs in the task (6 practice). Of the 60 test items, only 57 were retained, because of experimenter error.

Design and procedure. The experiment was a 2 (type of rating) $\times$ 3 (amount of extra study time) $\times$ 2 (randomized or blocked extra study time) mixed design, with type of rating being a between-subjects factor, and amount of extra study time and randomized or blocked study time being within-subjects factors.

The PA task presented each word pair, located at the center of the computer screen initially for up to 6 s. Participants were instructed to press the return key immediately following formation of an image. This was emphasized several times in the instructions and during the practice trials. After pressing the return key, or after a timeout of 6 s, whichever came first, participants were given 0, 2, or 6 s of extra study time. The amount of additional study time was randomly assigned to trials.

Following the allotted study interval, participants were asked whether they were successful in producing an image. If they responded "no," then they were asked reasons why not. If they responded "yes," then they were prompted to provide a metacognitive judgment (either a JOL or a QUE).

Although JOLs are often scaled as percent likelihood of recall, it was critical that JOLs and QUEs were based on comparable scales. Hence, we used an ordinal Likert scale for both judgments. In the JOL condition, participants were questioned as follows: "How confident are you that in about ten minutes from now you will be able to recall the second word of the item when prompted with the first word?" They were instructed to respond by providing a number ranging from 1 (*definitely won't recall*) to 5 (*definitely will recall*).

In the QUE condition, participants were questioned as follows: "What was the quality of the interactive image you made for the last item?" They were instructed to respond by providing a number ranging from 1 (*the lowest quality image*) to 5 (*the highest quality image*). For both the JOL and QUE ratings, the program only accepted responses in the 1–5 range.

The order of item presentation during practice and the experiment was randomized. During recall, the items were presented in a different randomized order. Participants responded during the recall portion at their own pace. They typed the response word of the pairs when presented with the stimulus. A word was scored as correct if the first three letters of the typed word matched the response.

Statistical-procedure. Following recommendations by Nelson (1984), we used Goodman–Kruskal gamma correlations to correlate metacognitive judgments, latency of encoding keypress responses, and item recall for each participant. We then used these estimated gamma correlations as dependent variables in additional analyses.

We also report partial gamma correlations in which the ordinal association between two variables is computed, controlling on a third variable. In particular, we controlled on latency to analyze the contribution of fluency to the relationship between judgments and recall. Unlike ordinal variables, blocking on continuous variables like latency to compute a partial gamma correlation requires aggregation of responses over some range of the continuous variable. Following the examples of Davis (1967) and Quade (1974), we calculated the concordances and discordances between recall and judgments within multiple strata of latency and then summed these concordances and discordances across all strata to calculate the partial gamma. A stratum of latency is a range in the distribution of each individual's response times that will be considered equivalent, so that recall and judgments can be correlated for the set of trials in a given stratum. For example, all recall and judgments made in the top 25% of response latencies can be treated as a stratum. We chose to form approximately equal-size quartiles of response times by doing a median split on the entire data and then doing another median split on each half to determine the four strata.

Results and Discussion

We checked for an effect of mixed versus blocked conditions for extra study time on both the metacognitive judgments and the accuracy of those judgments, and found no statistically reliable main effect or interaction involving this factor. Hence, all results are reported collapsing over blocked versus mixed presentation of additional study time. In all experiments, we used the Type I error rate criterion of $p < .05$ to evaluate statistical significance.

Recall Performance, Imagery Generation, and Judgment Magnitude

In aggregate, recall performance was good for both judgment groups (JOL $M = 0.52$, $SE = 0.03$; QUE $M = 0.60$, $SE = 0.03$). Both groups provided keypresses before the 6-s deadline on a substantial proportion of trials (JOL $M = 0.66$, $SE = 0.03$; QUE $M = 0.73$, $SE = 0.03$). As important, they also reported high levels of successful image formation (JOL $M = 0.79$, $SE = 0.02$; QUE $M = 0.83$, $SE = 0.02$). More than 95% of all trials generating valid responses prior to the 6-s deadline were reported as successful imagery production trials.

Table 1 reports the mean judgments and proportion correct for trials with successful image formation, divided by levels of additional study time (0, 2, and 6 s). Recall performance increased with additional study time, $F(2, 238) = 14.72$, $MSE = 0.02$. This effect did not interact reliably with judgment type (JOL vs. QUE), $F(2, 238) = 2.29$, $MSE = 0.02$, $p > .10$. Neither type of judgment reliably tracked the increasing probability of recall as a function of study time; the study time effect on judgments was not significant ($F < 1$).

Latency to Report Interactive Image

Latency to generate an image did not vary by judgment group (JOL $M = 4.3$ s, $SE = 0.11$; QUE $M = 4.1$ s, $SE = 0.12$; $F < 1$).

Table 1
Means (and Standard Errors) of Judgments and Proportion of Recall in the JOL and QUE Conditions for Experiment 1

	JOL		QUE	
Additional study time	Judgment	Recall	Judgment	Recall
Time-outs included				
0 s	3.03 (0.07)	.52 (.03)	3.11 (0.06)	.61 (.03)
2 s	2.99 (0.07)	.56 (.03)	3.22 (0.08)	.65 (.03)
6 s	3.09 (0.06)	.64 (.03)	3.17 (0.07)	.67 (.03)
Time-outs excluded				
0 s	3.09 (0.06)	.52 (.03)	3.21 (0.07)	.62 (.03)
2 s	3.12 (0.07)	.58 (.03)	3.35 (0.08)	.66 (.03)
6 s	3.15 (0.06)	.62 (.03)	3.31 (0.08)	.69 (.03)

Note. JOL = judgments of learning; QUE = quality-of-encoding ratings.

Unexpectedly, image-generation latency varied by additional study time provided, $F(2, 238) = 10.86$, $MSE = 0.15$. Latencies were reliably shorter in the 0-s condition ($M = 3.8$ s) than in the 2-s and 6-s conditions (*M*s = 4.0 s; *SE*s 0.09 and 0.10, respectively).

Relations Among Latency of Imagery Reports, Judgments, and Recall Performance

The critical data for evaluating the encoding fluency hypothesis are provided in Table 2: Goodman–Kruskal gamma correlations among the three key variables (image latency, judgment [JOL or QUE], and item recall). Image latency was negatively correlated with both JOL and QUE ratings ("Latency, judgment" column of Table 2). The magnitude of gamma did not vary by additional study time, $F(2, 222) = 1.33$, $MSE = 0.06$, $p > .25$. Thus, encoding fluency appears to have been a potent cue for judgments that did not lose its potency with additional time for elaborative encoding or mediator rehearsal. The main effect of judgment type was reliable, $F(1, 111) = 7.25$, $MSE = 0.13$. QUEs were more negatively correlated with image-formation latency than were JOLs (marginal $M = -0.42$ vs. -0.32, respectively). Given that additional study time did not influence the correlations, Figure 1 presents the marginal means for each gamma correlation, collapsing over study time.

The "Latency, recall" column of Table 2 reports gamma correlations between image latency and recall. These correlations were also negative, but near zero and small relative to the relationship of latency with judgments. The "Judgment, recall" column reports gamma correlations of judgments with recall. Judgment accuracy was reliably greater than chance ($\gamma = 0$) in the 0 and 2-s conditions but not in the 6-s condition. However, the effect of additional study time on relative accuracy of the judgments was not reliable, $F(2, 204) = 1.95$, $MSE = 0.22$, $p > .10$. On average, JOLs produced higher relative accuracy than QUEs, $F(1, 102) = 3.98$, $MSE = 0.23$. Marginal means were 0.21 for JOLs and 0.10 for QUEs, however, indicating that the judgments discriminated relatively poorly between successful and unsuccessful item recall.

A potential limitation in the data just reported is inclusion of a few trials that were time-outs (maximum 6-s latency) followed by

Table 2
Gamma Correlations (and Standard Errors) Between Latency, Judgments, and Recall for Experiment 1

	Latency, judgment		Latency, recall		Judgment, recall	
Additional study time	JOL	QUE	JOL	QUE	JOL	QUE
All trials						
0 s	−.29 (.03)	−.41 (.04)	−.08 (.05)	.01 (.06)	.21 (.06)	.21 (.06)
2 s	−.37 (.04)	−.44 (.03)	−.18 (.07)	−.01 (.05)	.26 (.07)	.04 (.06)
6 s	−.28 (.05)	−.41 (.04)	.01 (.06)	−.23 (.05)	.12 (.07)	.09 (.07)
Time-outs excluded						
0 s	−.22 (.04)	−.33 (.04)	−.08 (.06)	.02 (.06)	.24 (.07)	.17 (.07)
2 s	−.24 (.05)	−.34 (.04)	−.16 (.07)	−.04 (.06)	.27 (.08)	.06 (.07)
6 s	−.22 (.06)	−.32 (.04)	−.03 (.06)	−.14 (.06)	.13 (.08)	.06 (.08)

Note. JOL = judgments of learning; QUE = quality-of-encoding ratings.

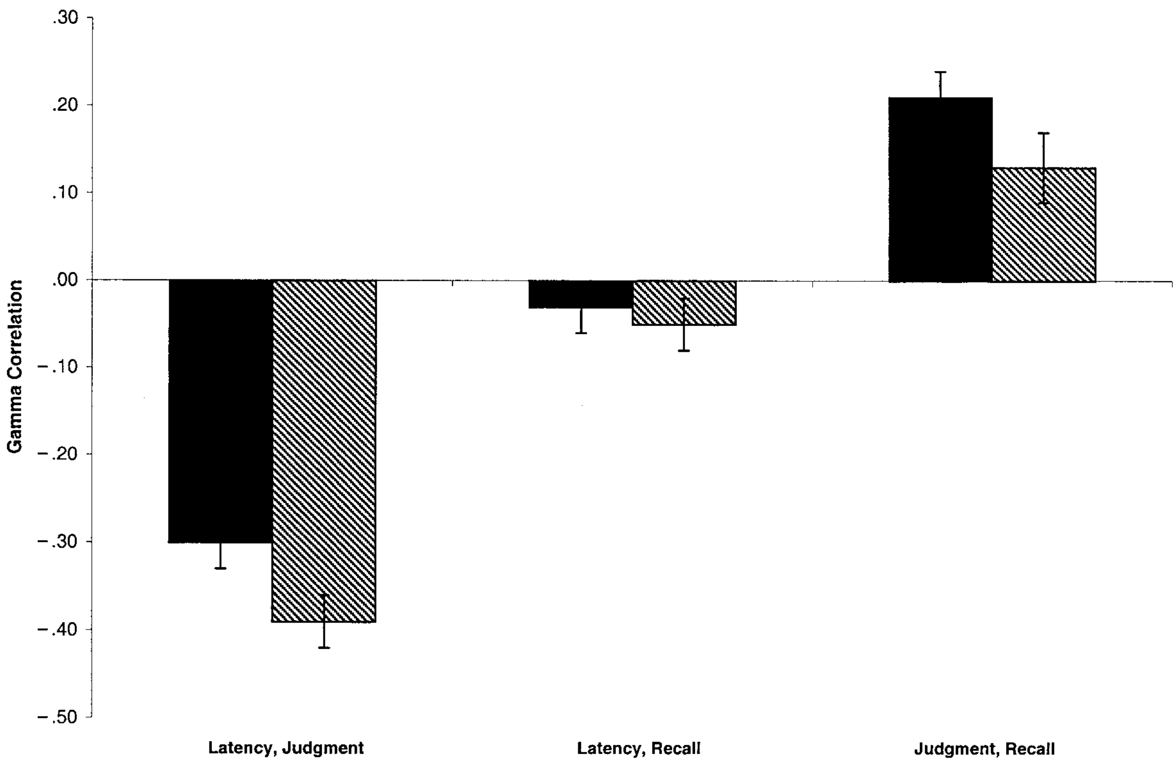

Figure 1. Gamma correlations between latency of image formation, recall, and metacognitive judgments (judgments of learning [JOL] or quality-of-encoding ratings [QUE]) pooling over amount of additional study time (Experiment 1). Solid bars represent JOLs; hatched bars represent QUEs.

a report of successful image formation. This outcome could indicate forgetting to make a keypress; image formation after the 6-s deadline, especially in the 2-s and 6-s additional time conditions; or a confabulated report of image-generation success. Data excluding time-out trials are also reported in Table 2. Exclusion of default 6-s latencies reduced the gamma of latency with judgments (*M* $\gamma = -.23$, JOL; $-.30$, QUE), but the basic pattern of results remained the same.

The data reported in Table 2 suggest indirectly that fluency affects metacognitive judgments, but they cannot account for the predictive accuracy for PA recall. This inference is based on the fact that the correlations of fluency with recall were small in magnitude. To formally test this hypothesis, we computed partial gamma correlations of judgments with recall, controlling on latency to generate an image. We ignored the amount of extra study time when we computed these correlations because this variable had little impact on relative accuracy and because doing so increased the number of data points available at each quartile of the latency distribution (see *Method* sections). For JOLs, the mean partial gamma correlations with recall were .19 ($SE = .04$) for all trials and .20 ($SE = .05$) excluding time-out trials. For QUEs, the corresponding mean partial gamma correlations were .13 ($SE = .04$) for all trials and .12 ($SE = .05$) excluding time-out trials. These values were close to the marginal means previously reported for both judgments, indicating little effect of controlling on latency. In all cases, these partial gamma correlations were significantly greater than zero.

Thus, the results from Experiment 1 are consistent with the hypothesis that encoding fluency has an impact on metacognitive judgments. Moreover, the impact appears to be greater for QUEs, which we expected to be more influenced by evidence about the encoding experience itself, than for JOLs, which require that encoding experiences be translated into expectations about retention and subsequent recall. However, the fact that judgment accuracy was above chance, even when controlling on latency, indicates that a cue or cues other than the fluency in generating mediators are also accessed when making metacognitive judgments about encoding and learning. We found it surprising that fluency remained a potent cue for metacognitive judgments even when additional study time was provided. Despite the facts that extra study time (a) provides an opportunity to elaborate on or otherwise rehearse the mediator and (b) benefited recall, it affected neither the magnitude nor the accuracy of metacognitive judgments.

Experiment 2

Experiment 1 demonstrated a robust correlation between encoding fluency, as measured by image-formation latency, and metacognitive judgments. One major goal for Experiment 2 was to obtain metacognitive judgments on every study trial, not just trials for which respondents reported successful image generation. This procedure is more representative of standard experiments using JOLs, in which individuals rate every item, irrespective of encoding outcomes. It also had an additional advantage for assessing the ease-of-processing hypothesis. In particular, this procedure allowed us to separate fluency for successful mediator generation (i.e., encoding fluency) from unsuccessful mediator generation (i.e., inability to produce an image in the allotted time). Thus, we could evaluate whether encoding fluency and mediator generation were both cues that were correlated with metacognitive judgments.

In standard PA learning experiments used to collect JOLs, respondents may be attempting spontaneously to implement mediational strategies (Dunlosky & Hertzog, 2001). If so, then they could base JOLs on whether a mediator was produced, as well as on the fluency with which mediators were successfully produced. Indeed, it could well be the case that the most accessible and influential cue would be success of mediator formation rather than the fluency of mediator formation. Both types of influence would be subsumed under the ease-of-processing concept. One possible explanation of the low correlations of metacognitive judgments with recall in Experiment 1 was that individuals made judgments only for successful image-formation trials, preventing variation in successful versus unsuccessful mediator generation from influencing the judgments and their accuracy. We hypothesized that the accuracy of JOLs and QUEs would be highest in the aggregate, when individuals could use the cue of success or failure in generating a mediator as a basis for judgments. However, we also expected the encoding fluency hypothesis to hold when interactive images had been successfully completed in the allotted time.

A goal of the procedure was to obtain as accurate a report of imagery use as possible. One concern about the strategy report format for Experiment 1 was that it asked whether an image had been successfully generated (participants responded "yes" or "no"). This type of format may be prone to demand characteristics. That is, individuals may be prone to falsely report successful imagery use. To minimize this possibility, we used an item-level mediator-strategy report technique developed and validated by Dunlosky and Hertzog (1998, 2001). Individuals were informed of the existence of alternative strategies, including interactive imagery, sentence generation, and rote repetition. They were instructed to use imagery if possible but to report whatever strategy was actually used for an item.

Method

Participants. We tested 133 individuals in groups of up to 6 persons. Of that total 133, data for 15 participants were excluded because they did not report using interactive imagery, as instructed, on at least 33% of the trials. After exclusion, there were 57 participants in the JOL condition and 61 participants in the QUE condition.

Apparatus and materials. The experiment used the same apparatus as in Experiment 1. We used a new PA list, consisting of 66 pairs of unrelated, concrete terms. We used 60 pairs in the experimental study-test trial, with 6 pairs used for practice making metacognitive judgments.

Design and procedure. The experiment was a two-way (type of rating) between-subjects design. Each session lasted approximately 1 hr. The PA task presented each word pair, located at the center of the computer screen, for 10 s. Participants were explicitly instructed to use interactive imagery to encode each pair and were instructed to press the *Return* key immediately following formation of an image. This was emphasized several times in the instructions and during the practice trials. Immediately following the presentation of each word pair, individuals were presented with a metacognitive judgment (either JOL or QUE).

For the JOL condition, participants were presented with the following question: "How confident are you that in about ten minutes from now you will be able to recall the second word of the item when prompted with the first word?" Participants were instructed to respond by providing a number ranging from 0 (*definitely won't recall*) to 100 (*definitely will recall*). In the QUE condition, participants were presented with the following question: "What was the quality of your encoding for the last item?" They were instructed to respond by providing a number ranging from 0 (*the lowest quality image*) to 100 (*the highest quality image*). For both conditions, the program accepted responses only in that range.

After a metacognitive judgment was entered, participants were asked to give a strategy report modeled after the procedure used by Dunlosky and Hertzog (1998). During the initial instructions in the experiment, participants were given explanations and examples of three possible strategies that could be used (e.g., interactive imagery, sentence generation, and rote repetition). They were told explicitly to use interactive imagery for the task, but they were also instructed to report whatever they actually used for a given trial, even if it was not interactive imagery. For the strategy report, the pair was presented at the top of the screen with the following statement: "How did you originally study the pair above?: 1) interactive imagery, 2) rote repetition, 3) sentence generation, 4) other, 5) no strategy, and 6) ran out of time." After entering one of these choices, the participant was presented with the next word pair.

Results and Discussion

Recall performance, mediator reports, and judgment magnitude. PA recall did not differ between the judgment groups (JOL $M = .68$, $SE = .03$; QUE $M = .66$, $SE = .03$). Judgments were close to the midpoint of the scale (JOL $M = 48\%$, $SE = 3\%$; QUE $M = 53\%$, $SE = 3\%$). On average, 75% of all JOL group trials and 78% of all QUE group trials were reported to result in successful imagery production. These outcomes are comparable to those of Dunlosky and Hertzog (1998), in which about 63% of the trials with 8-s presentation rates in an instructed-imagery condition were reported to yield successful imagery generation. Other reported mediator outcomes were relatively infrequent. The mean percentages of sentence generation, rote repetition, other strategy, no strategy, and self-reported time-out trials, averaged over JOL and QUE conditions, were 6%, 8%, 3%, 3%, and 5%, respectively. When we excluded time-out trials, a mean of 85% of the trials with valid latencies for the JOL group were reported to yield successful image formation. The corresponding mean for the QUE group was 89% of the valid trials.

Relations among latency of imagery reports, judgments, and recall performance. Across all trials, ignoring mediator outcomes, keypress latency for mediator reports was strongly correlated with metacognitive judgments. Figure 2 depicts the three relevant gamma correlations. Latencies were negatively correlated with judgments, and this correlation was again greater for QUEs than for JOLs, $t(114) = 1.96$. Latencies and judgments were both significantly related to recall. However, the magnitude of the correlations of judgments with recall was modest ($\gamma < .25$). The sample correlations of judgment with recall was larger for QUEs, compared with JOLs, but this difference was not reliable, $t(114) = 1.29$, $p > .10$.

Partial gamma correlations of judgments with recall, controlling on fluency, were also computed. Because latency to generate mediators was significantly related to recall outcomes, latency was a valid cue for recall. Controlling on latency reduced the gamma correlations to .18 ($SE = .03$) for QUEs and to .10 ($SE = .04$) for

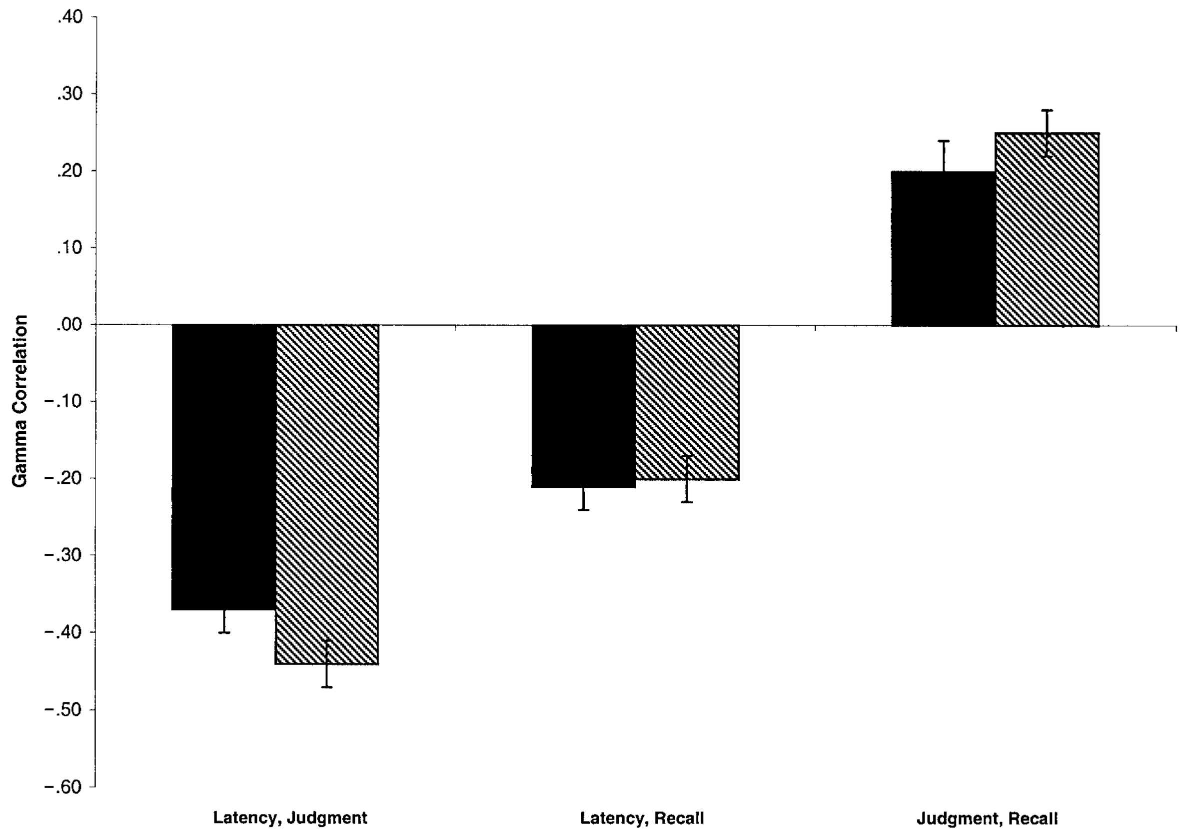

Figure 2. Gamma correlations between latency of image formation, recall, and metacognitive judgments (judgments of learning [JOL] or quality-of-encoding ratings [QUE]) for all trials (Experiment 2). Solid bars represent JOLs; hatched bars represent QUEs.

JOLs. In both cases, the null hypothesis ($\gamma = 0$) could still be rejected, indicating above-chance accuracy. Nevertheless, the magnitude of the correlations was reduced by controlling on latency.

We also restricted the analysis by including only reported imagery trials with valid generation latencies. This is the analogous report condition to Experiment 1 and allows latency to be treated as a measure of imagery encoding fluency (i.e., speed of successful image generation). Doing so did not change the general pattern of effects, but it reduced the correlations between recall and both mediator latencies and judgments (see Figure 3). Apparently, unsuccessful mediator generation served as a cue for judgments in the aggregate data. However, latency of successful image generation was still correlated with metacognitive judgments, and the correlation between QUEs and latency was still reliably higher than the correlation between JOLs and latency, $t(114) = 2.85$. When the analysis was restricted to successful imagery trials, the difference between JOLs and QUEs in the correlation of judgment with recall was not reliable, $t(111) = 1.26$, $p > .05$.

The results of Experiment 2 were relatively straightforward. Evidently, success or failure of generating a mediator is an important cue for metacognitive judgments. JOLs and QUEs showed above-chance correlations with PA recall when all trials were included, but accuracy of these judgments was close to nil when the analysis was restricted to trials for which imagery mediators had been successfully generated. On the other hand, latency of successful image formation was correlated with judgments but not with recall. This relationship was reliably larger for QUEs, which focus on encoding quality alone, than for JOLs, which require a prediction of future recall performance, presumably based in part on subjective encoding experiences (Mazzoni & Nelson, 1995). Thus, imagery encoding fluency is a cue that observers apparently use to make metacognitive judgments about PA learning, even when this cue lacks validity for predicting subsequent item recall. It seems that both the success of image generation and the fluency of generation influence metacognitive judgments about encoding and learning, consistent with the ease-of-processing view.

Experiment 3

An important issue in contemporary research on metacognitive judgments is whether individuals attend to multiple cues when making their judgments (Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002; Koriat, 1997). The first two experiments established that encoding fluency may be an important cue for judgments, even though such fluency is not necessarily related to recall. It is well-known that intrinsic stimulus cues such as associative relatedness or item concreteness influence JOLs (e.g., Koriat, 1997; Lovelace, 1984). Indeed, as noted earlier, Begg et al. (1989, Experiment 1) used item concreteness as a means of operationally

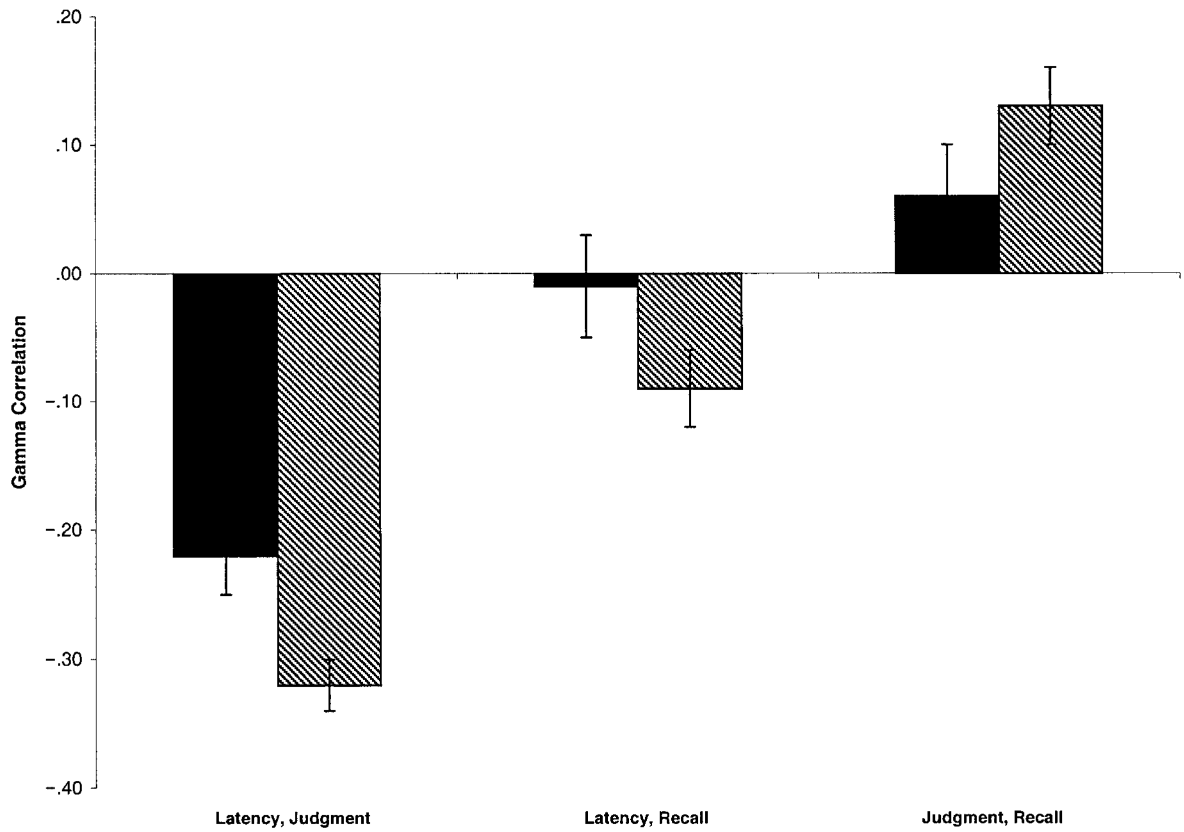

Figure 3. Gamma correlations between latency of image formation, recall, and metacognitive judgments (judgments of learning [JOL] or quality-of-encoding ratings [QUE]) for successful imagery trials only (Experiment 2). Solid bars represent JOLs; hatched bars represent QUEs.

defining the ease-of-processing concept. In Experiment 3, we also manipulated item concreteness under identical experimental conditions to Experiment 2. Individuals again received imagery instructions and reported mediator outcomes in addition to providing metacognitive judgments for all items.

This experimental design has several potential benefits. First, we hypothesized that successful mediator generation is more likely for concrete than for abstract items and that the fluency of successful generation would be greater for concrete items. One consequence of this hypothesized pattern is that the manipulation of concreteness should create positive cue validity for encoding fluency. The existing literature on mediators in PA learning suggests that, to generate an image for abstract items, the word elements must be recoded into a concrete, imageable form (Paivio, 1995; Yuille, 1973). Hence, latency for successful image generation should be longer for abstract items. The success rate in generating imagery mediators for abstract items is also lower, and abstract noun pairs are also more difficult to recall. In combination, these effects should increase the correlation of latency with judgments and with recall when the data are analyzed in the aggregate, ignoring concreteness. These effects would stand in contrast to those of the first two experiments, in which latency for successful image generation was essentially unrelated to recall.

A second benefit of the design is that it permits an evaluation of the hypothesis that encoding fluency would be used as a cue for judgments in addition to (and independently of) whether abstract or concrete items are studied. Here, as for Experiment 2, we analyze data from trials for which successful image generation was reported. We hypothesized that fluency would be correlated with judgments even when the intrinsic cue of concreteness is available to participants as a basis for metacognitive judgments. Indeed, we expected that both cues would influence judgments independently of one another. Thus, controlling on latency, we hypothesized a significant correlation of concreteness with the metacognitive judgments. Conversely, we hypothesized that, for concreteness, encoding fluency for successful image generation would be related to judgments but would not be a valid cue (i.e., not diagnostic of the probability of recall within concrete or abstract item classes). This outcome would replicate and extend results from Experiments 1 and 2, which used concrete PA items alone.

We also designed Experiment 3 to address the potential issue of reactive effects of pressing a response key to designate a mediator outcome. It is possible that the experimental procedure makes the cue of latency of encoding salient to participants and that they base their judgments on this cue for that reason alone. That is, fluency may not be a metacognitive cue used by participants unless it is made salient by the experimental procedures used in this study. Although this hypothesis is difficult to evaluate directly, we included a control condition in which individuals made metacognitive judgments and mediator reports but did not press a key to indicate successful mediator generation. We hypothesized that the keypress is actually inert and does not influence metacognitive judgments. Thus, we hypothesized no difference between groups who did and did not report image generation by keypresses on the following variables: PA recall, magnitudes of metacognitive judgments, effects of concreteness to metacognitive judgments, and correlations of judgments with recall.

Method

Participants. Participants were undergraduates at the Georgia Institute of Technology. All participants received one hour of extra credit for participation. We tested 162 individuals in groups of up to 6. Nine individuals were excluded because they pressed the bar even though they were in a control group and there was no mention of pressing the bar as soon as an image was formed. Similarly, participants who did not use interactive imagery on at least 33% of the trials were excluded. After exclusion, there were 37 participants in the JOL experimental condition, 36 participants in the JOL control condition, 34 participants in the QUE experimental condition, and 38 in the QUE control condition.

Apparatus and materials. The same apparatus used in Experiments 1 and 2 was used. Half of the word pairs consisted of unrelated, concrete terms and half consisted of unrelated abstract terms. We used 66 pairs in the task (3 abstract and 3 concrete pairs were used as practice).

Design and procedure. The experiment was a 2 (type of rating) $\times$ 2 (item type) $\times$ 2 (latency prompt) design. Type of rating (QUE vs. JOL) and latency prompt (*yes* or *no*) were between-subjects factors, and item type (concrete vs. abstract) was a within-subject factor. Participants were tested in groups of up to 6 persons in a session lasting approximately 1 hr.

The task and procedure for the two encoding fluency groups were identical to that used in Experiment 2. The only change in task instructions and procedures was for the control groups. Unlike in Experiment 2, the JOL and QUE control groups were not given any prompt to press a key on completion of encoding, and speed of image formation was not mentioned at all in the task instructions.

Table 3
Proportion of Successful Images, Judgments, Recall, and Response Latency by Judgment Type (QUE vs. JOL) and Experimental Latency Prompt (Control vs. Latency) for Experiment 3

	Control		Latency	
Judgment type	Abstract	Concrete	Abstract	Concrete
		All trials		
QUE				
Imagery	.39 (.04)	.88 (.02)	.51 (.05)	.90 (.02)
Judgment	47.72 (2.92)	66.35 (2.91)	36.97 (4.20)	52.95 (3.94)
Recall	.25 (.03)	.79 (.03)	.31 (.04)	.74 (.04)
Latency			6.44 (0.38)	4.66 (0.33)
		Successful imagery trials		
Judgment	60.28 (3.27)	69.01 (2.88)	47.13 (4.09)	56.76 (3.97)
Recall	.34 (.05)	.79 (.03)	.34 (.05)	.77 (.04)
Latency			5.13 (0.29)	4.24 (0.28)
		All trials		
JOL				
Imagery	.39 (.05)	.88 (.03)	.43 (.04)	.89 (.02)
Judgment	44.20 (2.99)	57.20 (2.37)	39.38 (2.87)	53.16 (3.04)
Recall	.37 (.03)	.85 (.03)	.35 (.03)	.82 (.02)
Latency			6.98 (0.30)	4.80 (0.27)
		Successful imagery trials		
Judgment	49.10 (3.09)	59.23 (2.45)	50.14 (3.40)	55.00 (3.13)
Recall	.38 (.05)	.87 (.02)	.51 (.04)	.85 (.02)
Latency			5.31 (0.27)	4.46 (0.22)

Note. JOL = judgments of learning; QUE = quality-of-encoding ratings. Standard errors are in parentheses.

Results and Discussion

Effects of mediator generation keypresses. Table 3 reports the means and standard errors of the major variables in each experimental condition. There was only limited evidence in the data of a reactive effect of responding to indicate latency of mediator formation. The top panel of Table 3 reports the data for all trials, ignoring mediator outcomes. All variables were analyzed in separate Latency Prompt (control vs. latency) $\times$ Judgment Type (JOL vs. QUE) $\times$ Concreteness (concrete vs. abstract) analyses of variance (ANOVAs). Control and latency participants differed in mean metacognitive judgments (marginal M = 46%, latency group; M = 54%, control group), $F(1, 141) = 7.65$, MSE = 645.02. However, all interactions involving latency prompt were not significant ($F < 2$, $\eta^2 \leq .01$). Concreteness showed a robust main effect on judgments, $F(1, 141) = 204.89$, $MSE = 83.22$, with higher ratings for concrete items (marginal M = 57%) than for abstract items (marginal M = 42%). The main effect of judgment type was not reliable ($F < 1$). No other interactions were reliable.

Likewise, there was no evidence of a difference between control and latency conditions on PA recall ($F < 1$) nor did interactions involving latency prompt achieve statistical significance. Recall was reliably higher for concrete items (marginal $M = 0.80$) than for abstract items (marginal $M = 0.32$), $F(1, 140) = 955.57$, $MSE = 0.02$. Nor did providing latency materially influence the proportion of reported imagery use, $F(1, 141) = 2.89$, $MSE = 0.05$. Individuals in the latency condition had slightly higher reported imagery use than individuals in the control condition (marginal $M = 0.68$ vs. 0.64, respectively). The sample mean differences were larger for QUE judgments (see Table 3; QUE marginal $M = 0.78$ vs. 0.65). None of the associated interactions was reliable. As expected, concreteness had a robust relationship to the proportion of successful imagery trials, with higher success rates for concrete items (marginal $M = 0.89$) relative to abstract items (marginal $M = 0.43$), $F(1, 141) = 397.54$, MSE = 0.04.

Concreteness effects on relative accuracy of metacognitive judgments. In general, therefore, the data supported the argument that requesting keypresses to indicate mediator formation did not materially affect the patterns of data. This inference was reinforced when the relative accuracy of the metacognitive judgments was evaluated. Table 4 reports the gamma correlations between judgments and recall for the control and latency conditions. Correlations for all trials, excluding mediator outcomes, are reported in the "All trials" portion of the table. When correlations were computed for all trials (ignoring concreteness), the relative accuracy of judgments in the aggregate ("Aggregate" columns for QUE and JOL in Table 4) was reliably greater than zero in all four cases. Moreover, a 2 (judgment type) $\times$ 2 (latency prompt) ANOVA showed that the aggregate gamma correlations did not differ be-

Table 4
Gamma Correlations (and Standard Errors) of Metacognitive Judgments (JOLs and QUEs) for Control and Latency Conditions Aggregated Over All Items and Separately for Abstract and Concrete Items in Experiment 3

Condition	Aggregate	Abstract	Concrete
	QUE		
All trials			
Latency	.44 (.06)	.20 (.08)	.15 (.10)
Control	.47 (.05)	.34 (.06)	.10 (.09)
Image generation			
Latency	.27 (.08)	.13 (.12)	.10 (.10)
Control	.27 (.08)	.12 (.13)	.08 (.10)
	JOL		
All trials			
Latency	.44 (.05)	.28 (.06)	.24 (.09)
Control	.45 (.04)	.28 (.08)	.06 (.09)
Image generation			
Latency	.29 (.07)	.28 (.10)	.08 (.11)
Control	.29 (.06)	.34 (.12)	+.01 (.10)

Note. JOL = judgments of learning; QUE = quality-of-encoding ratings.

tween latency and control participants ($F < 1$). The aggregate gamma correlations also did not differ between QUEs and JOLs ($F < 1$).

Table 4 also reports gamma correlations separately for abstract and concrete items. These correlations were lower than when the data were aggregated across both types of items. For the aggregate data, the marginal mean gamma was .45; when data were disaggregated into abstract and concrete items, the marginal mean gamma was .21. Relative accuracy was much higher in the aggregate because individuals systematically gave higher ratings to concrete items and also recalled more concrete items. This finding confirms that concreteness is an important cue for both JOLs and QUEs and hence may contribute to overall levels of relative accuracy (Begg et al., 1989). Whether the relationship to concreteness can be attributed to the effects it has on mediator outcomes is considered below.

A $2 \times 2 \times 2$ (Concreteness $\times$ Judgment Type $\times$ Latency Prompt) ANOVA on the gamma correlations reported in Table 4, with concreteness as a within-subject factor, revealed no reliable main effect of latency prompt, nor any interactions associated with latency prompt ($p > .10$). There was, however, a reliable effect of concreteness, $F(1, 110) = 5.58$, $MSE = 0.18$. Relative accuracy was higher for abstract items (marginal $M\ \gamma = .27$) than for concrete items (marginal $M\ \gamma = .14$).

These data confirm the importance of the intrinsic cue of item concreteness for producing accurate judgments, in the aggregate. Hertzog et al. (2002) reported similar attenuation of gamma correlations for JOLs provided for related PA items (e.g., sugar–coffee). Like concreteness, associational relatedness is a potent cue that influences both judgments and recall (Koriat, 1997) and contributes to relative accuracy of the judgments, in the aggregate. However, individuals do not appear to discriminate well between related items they will and will not recall.

When the data are restricted to successful image-generation trials ("Image generation" portion of Table 4), there was still little indication of a difference between latency and control conditions in relative accuracy ($F < 1$). The relative accuracy of the metacognitive judgments decreased when unsuccessful imagery generation trials were eliminated, although the aggregate gamma correlations were greater than zero in all four cases. This outcome is consistent with Experiment 2 and shows that an important cue contributing to judgment accuracy (one which covaries with concreteness) is whether a mediator was successfully created. The data again produced higher gamma correlations for abstract items (marginal $M = .24$) versus concrete items (marginal $M = .04$), $F(1, 81) = 5.69$, $MSE = 0.30$.

Thus, these results clearly indicate that the relative accuracy of metacognitive judgments was unaffected by requesting a keypress to indicate mediator generation. Most important, they also show expected effects of item concreteness and mediator generation on judgment magnitudes and relative accuracy.

Concreteness effects on encoding fluency. The remaining analyses were conducted only with participants in the latency condition. Our first hypothesis regarding latency was that concreteness would have a major impact on encoding fluency, consistent with the ease-of-processing concept. Indeed, there were substantial effects of concreteness on latency to report mediator completion, replicating earlier work by Yuille and Paivio (1967). When all trials were analyzed, there was a robust concreteness main effect, $F(1, 44) = 89.21$, $MSE = 1.56$. Keypress latencies were reliably faster for concrete items (marginal $M = 4.7$ s) than for abstract items (marginal $M = 6.7$ s). Similar results were obtained when data were restricted to successful interactive imagery trials only. Given that successful image generation was less likely for abstract items, the concreteness main effect was smaller for the reduced data set, $F(1, 67) = 26.40$, $MSE = 0.97$. Successful image generation was reliably faster for concrete items ($M = 4.4$ s) than for abstract items ($M = 5.2$ s). Thus, concreteness was positively related to encoding fluency but was not isomorphic with it.

Encoding fluency and metacognitive judgments. Given the robust effects of concreteness on recall, judgments, and encoding fluency, the second major hypothesis of interest was a relationship of all three variables with each other when the relevant gamma correlations were computed, as in Experiments 1 and 2. Relative accuracy of the judgments is already reported in Table 4 for latency condition participants. The upper panel of Table 5 shows the two remaining correlations involving latency for all trials, irrespective of mediator outcomes. Note that for both QUEs and JOLs, the aggregate data (pooling concrete and abstract items) yielded robust correlations ($\gamma < -.4$) of keypress latency with metacognitive judgments. Note also that, unlike in the earlier experiments, there was a robust, negative relationship between keypress latency and recall for both groups. This relationship arises because, normatively, abstract items require more time to generate a mediator outcome and also yield lower recall. Given the cue validity for fluency, it is perhaps not surprising that the relative accuracy of the judgments for predicting recall is about 4 for both types of judgments. This pattern was attenuated, but not greatly altered, when the data were restricted to successful imagery trials (reported in the lower panel of Table 5). Thus, both the success at generating a mediator and the fluency for generating imagery

Table 5
Gamma Correlations (and Standard Errors) Between Encoding Fluency, Metacognitive Judgments (JOLs and QUEs), and Recall for Experiment 3

Judgment type	Fluency, judgments	Fluency, recall
	All trials	
QUE		
Aggregate	−.46 (.05)	−.38 (.05)
Concrete	−.39 (.04)	−.07 (.08)
Abstract	−.41 (.06)	−.19 (.07)
JOL		
Aggregate	−.44 (.03)	−.46 (.04)
Concrete	−.27 (.04)	−.27 (.08)
Abstract	−.46 (.04)	−.31 (.06)
	Successful imagery trials	
QUE		
Aggregate	−.31 (.04)	−.25 (.05)
Concrete	−.31 (.05)	−.04 (.08)
Abstract	−.15 (.07)	−.11 (.09)
JOL		
Aggregate	−.25 (.04)	−.26 (.06)
Concrete	−.22 (.04)	−.10 (.08)
Abstract	−.27 (.07)	−.06 (.11)

Note. JOL = judgments of learning; QUE = quality-of-encoding ratings.

mediators apparently influenced both types of metacognitive judgments.

Table 5 also reports the data separately for concrete and abstract items. When all trials were included, the gamma correlations of keypress latencies with judgments were about as large for concrete and abstract items, separately, as when the two types of items were pooled. This was not the case, however, for the correlations involving recall. Indeed, for QUEs several of the correlations were not reliably greater than zero. For the JOLs, the correlations were attenuated but significantly greater than zero in all four cases. These data show that item concreteness does not fully entrain and cannot completely capture the encoding fluency effects on metacognitive judgments. Hence, the data suggest that individuals were able to attend to at least two different cues when generating metacognitive judgments—item concreteness and encoding fluency, specifically.

We provided additional support for this interpretation by computing partial gamma correlations of judgments with recall, controlling on keypress latency. Using all trials, the partial correlations of JOLs and QUEs with recall were .21 (SE = .06) and .27 (SE = .07), respectively. Both correlations were reduced from about .4 by controlling on keypress latency. However, both were reliably greater than zero, suggesting that concreteness was still a useful cue, even when the latency of mediator reports was statistically controlled. By contrast, the partial gamma correlations of JOLs for concrete and abstract items were .10 (SE = .11) and .09 (SE = .08) respectively, neither of which was statistically different from chance ($p > .25$). For QUEs, the partial correlations for concrete and abstract items were .23 (SE = .11) and .09 (SE = .09). Only the former correlation was reliably greater than zero. Taken together, these two sets of correlations indicate that concreteness is a valid cue, above and beyond any effects of latency, but that within the classes of concrete and abstract items, controlling on keypress latency reduces relative accuracy of the judgments.

Thus, the pattern of results indicates that concreteness was an important cue for generating above-chance accuracy in a way that is not completely dependent on ease-of-processing at encoding. This result is consistent with the view that the cues of concreteness, mediator generation success, and encoding fluency have correlated but distinct effects on metacognitive judgments and on the relative accuracy of those judgments.

Experiments 1 and 2 suggested that QUEs may be more influenced than JOLs by encoding fluency. Experiment 3 produced no main effect of type (QUEs vs. JOLs) for correlations of judgments with keypress latencies ($F < 1$). However, there was a robust Type × Concreteness interaction, $F(1, 67) = 4.01$, $MSE = 0.05$. QUEs for concrete items appeared to be more highly correlated with keypress latencies than were JOLs, but this was not the case for abstract items (see Table 5). When differences in correlations for concrete items alone were tested (as in earlier experiments), there was a reliable difference in the correlation of latency on judgments, with a higher gamma correlation for QUEs than for JOLs, $t(68) = 1.96$, one-tailed $p < .05$. The basic pattern of gamma correlations was similar when the analysis was restricted to successful imagery trials (bottom panel of Table 5), with a reliable Type × Concreteness interaction, $F(1, 62) = 4.05$, $MSE = 0.09$. However, the difference in correlations for concrete items alone was not reliable, $t(68) = 1.39$, one-tailed $p > .05$.

In sum, Experiment 3 confirms that item concreteness influences metacognitive judgments, but also demonstrates that this effect was not isomorphic with the effects of successful mediator generation or encoding fluency on the judgments. The outcomes support the view that concreteness, as an intrinsic cue, has mnemonic processing consequences (Koriat, 1997) but that the cues of mediator generation success and the fluency of successful mediator generation operate independently of the concreteness cue, per se.

General Discussion

The three experiments reported here indicate that encoding fluency, defined as the latency to generate imagery mediators, is a potent cue influencing metacognitive judgments about encoding and learning. This work therefore extends the existing literature showing retrieval fluency to be a salient metacognitive cue (e.g., Benjamin et al., 1998; Matvey et al., 2001) by demonstrating that the fluency of encoding processes that govern initial acquisition also influence judgments occurring at the time of encoding. Undoubtedly, there are boundary conditions on whether encoding fluency would influence judgments about learning. For example, delaying JOLs for a brief period of time increases their accuracy (the delayed-JOL effect; Nelson & Dunlosky, 1991; see also Kelemen & Weaver, 1997) and apparently causes delayed JOLs to behave in ways that are qualitatively different from the kind of immediate JOLs studied here (see also Dunlosky & Nelson, 1994). One can therefore hypothesize that encoding fluency would have little impact on delayed JOLs.

Encoding fluency has its influence despite the fact that it is often not a valid, diagnostic cue. In Experiments 1 and 2, robust correlations of latency and both types of metacognitive judgments were observed, consistent with the encoding fluency hypothesis, even though the correlations of latency with recall outcomes were essentially zero. In this sense, reliance on encoding fluency, like retrieval fluency, can be misleading because individuals may be relying on a source of information about item encoding that is not associated with between-items differences in probability of recall. Individuals apparently judge rapidly generated mediators to be more effective for learning, even when they are not. However, when fluency is diagnostic of future recall, as in the case of generating mediators for mixed lists of concrete and abstract PAs, it may act to enhance the relative accuracy of metacognitive judgments.

One of the intriguing aspects of the present research is that it suggests that the above-chance relative accuracy of immediate JOLs and QUEs can be attributed in large part to the cue of unsuccessful mediator generation. In all three experiments, relative accuracy of judgments for the subset of trials resulting in successful image generation was low and often not greater than zero (i.e., at chance). Invariably, relative accuracy was lower for items with successful image generation than for the entire set of PA items. This outcome is admittedly indirect, but it suggests that participants discriminated effectively between successful and unsuccessful mediator generation and adjusted their metacognitive judgments accordingly. However, it also suggests that participants did not successfully discriminate, at encoding, mediators that will later be retrieved from those that will not.

One surprising aspect of the present results was the finding in Experiment 1 that providing additional study time did not attenuate the relationship between encoding fluency and metacognitive judgments. Perhaps the additional opportunity for elaboration or rehearsal of the mediator did not produce new cues that influenced the judgments, despite the fact that providing additional study time did increase the probability of item recall. There are multiple demonstrations in the literature that metacognitive judgments do not necessarily track variables that influence recall (e.g., Mazzoni & Nelson, 1995) and that there are influences on judgments that are apparently unrelated to recall (e.g., Benjamin et al., 1998).

Broadly speaking, then, the results of this series of experiments are consistent with Koriat's (1997) cue-utilization framework for JOLs (but see Dunlosky & Matvey, 2001). In particular, it appears that individuals' metacognitive judgments may be influenced by multiple cues in ways that are consistent with observable stimulus properties (Koriat's intrinsic cues), such as concreteness or relatedness of the word pairs, but that mnemonic cues such as mediator generation success and encoding fluency also influence JOLs. Available evidence also indicates that metacognitive judgments do not depend only on individual item properties in isolation but also partly depend on the contrast between different items in a heterogeneous list. Unrelated PA items are given lower JOLs when they are either mixed with or preceded by easier related items (Dunlosky & Matvey, 2001; Hertzog et al., 2002). JOLs track von Restorff effects on recall of a salient isolated item late in the list, but not early in the list when contrasts are not possible (Dunlosky, Hunt, Clark, 2000). Finally, Begg et al. (1989, Experiment 2) showed that JOLs are higher for effective interactive imagery mediators than for less effective separate-imagery mediators when imagery type was a within-subject (not a between-subjects) factor. Further progress in understanding of the multiple cues observers use to make JOLs will apparently require additional attention to the ways in which cues become more or less salient across sets of items as a function of the experimental context. How, and how consistently, do individuals weight different sources of information when making metacognitive judgments? By analogy, one might expect to identify conditions under which the importance of encoding fluency might depend on the nature of task experience, item domain sampling, and the constellation of independent variables that cause variation in the range of cues accessed at the time the judgments are made.

One useful feature of the present set of experiments was the demonstration that a relatively new metacognitive judgment, a QUE, behaves much like a JOL, despite the fact that it does not require individuals to rate confidence in future recall. If anything, QUEs were more sensitive to encoding fluency than JOLs, and it may be the case that they would also be more sensitive to different cues available at encoding as well. If the goal is to understand how encoding behavior is monitored and translated into metacognitive judgments, QUEs may prove to be more sensitive than JOLs to experimental manipulations of encoding parameters.

We note in passing that there were limitations of the present study that should be addressed in future research. We relied on correlations between keypress latencies, judgments, and recall to make inferences about encoding fluency. Inferences about the salience of encoding fluency as a cue would be enhanced by studying experimental manipulations of fluency that can then be evaluated in terms of differential impact on judgments and recall (see Benjamin et al., 1998). Likewise, we do not yet know whether individuals are aware of the variations in encoding fluency and whether they can be effectively instructed to use or to ignore fluency as a cue when making metacognitive judgments. The effects observed in this study warrant more elaborate attention to the encoding fluency phenomenon and, more generally, the extent to which mediator outcomes play an important role in shaping metacognitive judgments during encoding.

In summary, even though fluency of processing has often been cited as an important basis of metacognitive judgments, research to date had not established empirically that the fluency of encoding was a basis of JOLs. In a defining article, Begg et al. (1989) proposed this ease-of-processing hypothesis and demonstrated that item concreteness affected JOLs. Although this effect was viewed as consistent with the hypothesis, a more recent theory of JOLs offers an alternative explanation based on the idea that the item characteristics directly influence JOLs (Koriat, 1997). The experiments reported here provide consistent and more definitive support for the hypothesis that both success of mediator generation and the fluency of successful generation influence metacognitive judgments and that these effects are related to, but not isomorphic with, an intrinsic cue—concreteness—that has consequences for ease of processing.

References

Apple Computer, Inc. (1995). Hypercard: Script language guide (Version 2.3). Cupertino, CA: Author.

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., & 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.

Bower, G. H. (1970). Imagery as a relational organizer in associative learning. *Journal of Verbal Learning and Verbal Behavior, 9,* 529–533.

Davis, J. A. (1967). A partial coefficient for Goodman and Kruskal's gamma. *American Statistical Association Journal, 62,* 189–193.

Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy use? *Psychology and Aging, 13,* 597–607.

Dunlosky, J., & Hertzog, C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory & Cognition, 29,* 247–253.

Dunlosky, J., Hunt, R. R., & Clark, E. (2000). Is perceptual salience needed in explanations of the isolation effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 649–657.

Dunlosky, J., Kubat, A. K., & Hertzog, C. (in press). Do age differences exist in monitoring of encoding? Effects of aging on the magnitude and accuracy of quality-of-encoding judgments. *American Journal of Psychology.*

Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic–extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1180–1191.

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.

Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging, 17,* 209–225.

Johnson, M. K., & Raye, C. L. (2000). Cognitive and brain mechanisms of false memories and beliefs. In D. L. Schacter & E. Scarry (Eds.), *Memory, brain, and belief* (pp. 35–86). Cambridge, MA: Harvard University Press.

Kelemen, W. L., & Weaver, C. A. (1997). Enhanced metamemory at delays: Why do judgments of learning improve over time? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1394–1409.

Kelley, C. M., & Jacoby, L. L. (1998). Subjective reports and process disassociation: Fluency, knowing, and feeling. *Acta Psychologica, 98,* 127–140.

Kelley, C. M., & Lindsay, S. D. (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.

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

Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 3,* 464–470.

Lovelace, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 756–766.

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. (1993). Strategies in study time allocation: Why is study time sometimes not effective? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1263–1274.

Mazzoni, G., & Nelson, T. O. (1995). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1263–1274.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 84,* 93–116.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist, 51,* 102–116.

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., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–141). New York: Academic Press.

Paivio, A. (1995). Imagery and memory. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 977–986). Cambridge, MA: MIT Press.

Quade, D. (1974). Nonparametric partial correlation. In H. M. Blalock (Ed.), *Measurement in the social sciences* (pp. 369–398). London: Macmillan.

Rabinowitz, J. C., Craik, F. I. M., Ackerman, B. P., & Hinchley, J. L. (1982). Aging and metamemory: The roles of relatedness and imagery. *Journal of Gerontology, 37,* 688–695.

Richardson, J. T. E. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review, 5,* 597–614.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1024–1037.

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

Yuille, J. C. (1973). A detailed examination of mediation in PA learning. *Memory & Cognition, 1,* 333–342.

Yuille, J. C., & Paivio, A. (1967). Latency of imaginal and verbal mediators as a function of stimulus and response concreteness-imagery. *Journal of Experimental Psychology, 75,* 540–544.

Received November 12, 2001
Revision received July 1, 2002
Accepted August 3, 2002 ■