Judgments of learning: Evidence for a two-stage process
Lisa K Son; Janet Metcalfe
Memory & Cognition; Sep 2005; 33, 6; ABI/INFORM Global
pg. 1116

Memory & Cognition
2005, 33 (6), 1116-1129

Judgments of learning: Evidence for a two-stage process

LISA K. SON
Barnard College, New York, New York

and

JANET METCALFE
Columbia University, New York, New York

Three experiments tested the hypothesis that people make judgments of learning (JOLs) by attempting to retrieve the target first. If this were the whole story, then the reaction time (RT) functions for making JOLs with no special instructions would parallel those found when people are told to first attempt retrieval and then make a JOL. In the present data, monotonic functions, showing an increase in RT with decreasing JOL, were found when people were instructed to retrieve covertly or overtly and then make a JOL, as would be expected if retrieval fluency entirely determined JOLs. However, the functions for making uninstructed JOLs were different: Low JOLs were made quickly, not slowly, and the curves were inverted U shapes, rather than linear. Furthermore, people's memory performance was somewhat better, especially on low-JOL items, when they were instructed to first retrieve as opposed to when they were told only to make JOLs. To account for these data, we propose a two-stage model of JOLs, with the first stage occurring prior to attempted retrieval.

The question addressed in this article is this: How do people make judgments of learning? Judgments of learning (JOLs) are assessments that people make about how well they have learned particular information—that is, predictions about how likely they will be to remember a target item when later given a cue. These assessments are then, presumably, used to control further study (Benjamin, Bjork, & Schwartz, 1998; Dunlosky & Hertzog, 1998; Mazzoni, Cornoldi, & Marchitelli, 1990; Metcalfe, 2002; Metcalfe & Kornell, 2003; Nelson & Dunlosky, 1991; Nelson, Dunlosky, Graf, & Narens, 1994; Nelson & Narens, 1990, 1994; Son, 2004; Son & Metcalfe, 2000). For example, Nelson and Dunlosky argued that "the accuracy of JOLs is critical, because if the JOLs are inaccurate, the allocation of subsequent study time will correspondingly be less than optimal" (p. 267). Because their role in learning is crucial, understanding the mechanisms underlying these judgments is central to understanding people's control of their own mental processes and for finding ways to ameliorate those processes. We will argue here that JOLs made at a delay are based on two distinct processes rather than on only the single retrieval process that previous researchers have emphasized. If so, then understanding both of those processes, and the pitfalls therein, may be pivotal to improving people's performance. In this article, we provide data that challenge the assumption that delayed JOLs are made only by an assessment of retrieval goodness or fluency. We argue and provide evidence that JOLs, like other metacognitive judgments, instead may involve two distinct stages: (1) a quick preretrieval stage based, perhaps, on cue familiarity, which determines whether or not the second stage—retrieval—will occur, and (2) a later stage in which the judgments are based on an assessment of the goodness and/or fluency of retrieval or on other aspects of the target that can become apparent once retrieval has been attempted.

In the typical delayed-JOL paradigm that we investigated, participants studied cue–target pairs. Following study, they were presented with only the cue and were asked to judge how confident they were that they would be able to remember the target when they were given the cue on a future memory test. Then, at some later time, they were given a test. The prevailing view of how such JOLs are made is that when given the cue, people try to retrieve the target and base their judgments on their perception of some of the qualities of that attempted retrieval of the target, such as how quickly, easily, or fluently it comes to mind, or on how replete the retrieved information is (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Benjamin et al., 1998; Dunlosky & Nelson, 1994; Dunlosky, Rawson, & McDonald, 2002; Kelemen & Weaver, 1997; Kelley & Lindsay, 1993; Maki, 1998; Nelson & Dunlosky, 1991; Nelson & Narens, 1990; Nelson, Narens, & Dunlosky, 2004; Spellman & Bjork, 1992; Weaver & Kelemen, 1997).

This research was supported by NIMH Grant MH60637 to J.M. The experiments were conducted as part of a doctoral dissertation completed by L.K.S. at Columbia University. The authors thank Brady Butterfield, Dan Kimball, Nate Kornell, Jason Kruk, Jennifer Mangels, Jasia Pietrzak, Robert Remez, and Herb Terrace for their help and many insightful comments. Correspondence regarding this article should be addressed to the first author at Department of Psychology, Barnard College, 3009 Broadway, New York, NY 10027 (e-mail: lson@barnard.edu).

Copyright 2005 Psychonomic Society, Inc.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The delayed-JOL effect (Nelson & Dunlosky, 1991)—which suggests that JOLs are more accurate at predicting later performance when they are made after a delay has intervened between study and the judgments than when the judgments are made immediately—has triggered more precise formulations of how people make JOLs. Three explanations have emerged. Nelson and Dunlosky proposed the monitoring-dual-memories hypothesis, which assumes that the JOLs are made by retrieving information from both short-term memory (STM) and long-term memory (LTM). In the immediate-JOL condition, STM information is accessible and dominant. The highly transient STM information is no longer available at final test. The presence of STM information during the judgment, therefore, adds nondiagnostic information to the judgment process, reducing the accuracy of the JOLs. By contrast, the delayed JOL—which is based primarily on retrieval of information from LTM (with little input from STM)—is more accurate in predicting final test performance, which is also based on LTM alone.

The second explanation (Begg et al., 1989; Glenberg, 1987)—a transfer-appropriate processing view—proposes that the delayed-JOL effect occurs because of differences between the two JOL conditions in the degree of contextual match from the time of the judgment to the time of the test. Making a JOL in a situation that is as similar as possible to that of the test should maximize its accuracy. Insofar as the retrieval, which forms the basis of the JOL, is more similar to that of a delayed test in the delayed-JOL than it is in the immediate-JOL condition, the former judgments are predicted to be more accurate.

The third explanation (Kimball & Metcalfe, 2003; Spellman & Bjork, 1992) locates the increase in gamma accuracy between immediate and delayed JOLs in the memory system rather than in the metamemory system. This so-called Heisenberg explanation says that people attempt retrieval to make their JOLs but, in the delayed-JOL condition, they are successful with only some of those attempts. They benefit from retrieval practice if they are successful, but this occurs on only some of the items. The items that receive this extra practice are not distributed randomly across the JOL range. Rather, they are those given high JOLs. Those items that people fail to retrieve are given low JOLs and get no additional study. Thus, the high-JOL items benefit from an extra (spaced) study trial, whereas the low-JOL items receive no additional practice. This differential study has an effect on memory that bolsters the predictive value of the ratings. In the immediate-JOL condition, virtually all items are assumed to receive an extra study trial during the judgment, but it is massed rather than spaced, and because it is uniform across the entire JOL range, it does not improve the accuracy of the JOLs.

Although these three explanations of the delayed-JOL effects are still being hotly debated, in one way they are all the same. They all assume that, when asked to make a JOL, people attempt to retrieve the target and base their JOLs on some aspect of the outcome.

Factors such as the a priori difficulty of the items, their associative relatedness, their imagery value, or the number of times or the spacing with which an item has been presented may, of course, influence JOLs. These are what Koriat (1997), in his cue-utilization theory of JOLs, called "extrinsic and intrinsic factors." When the judgments are delayed, though, as they are in the present study, these factors probably influence the judgments indirectly in making the retrieval itself more or less fluent or replete. The other class of information specified by Koriat (1997) as contributing to JOLs is mnemonic. Mnemonic information includes the ease with which the information comes to mind and the ease of processing, both of which we here call *retrieval fluency*. It also includes people's memory for the ease of acquisition and their memory for the outcome of previous recall attempts. This information, presumably, is retrieved along with the target as target-specific source attributes. All of these cues are depend on appraisal of the process or of the attributes associated with attempted retrieval of the target.

Koriat (1997) also included the familiarity of the cue as mnemonic information contributing to JOLs, but he cited only feeling-of-knowing (FOK) findings. FOKs are metacognitive judgments made concerning the probability of later being able to recognize an item in a multiple-choice test that the person is unable to recall at the time the judgment is made. Although cue familiarity has been shown to be an important factor in FOK judgments, as will be outlined shortly, there are no extant data in the JOL literature to support its inclusion as a factor in determining JOLs. As we will detail shortly, however, we will make a case that Koriat (1997) was correct to include cue familiarity on his list, and we will provide supporting data in this article. Furthermore, we suggest that this factor may be distinctly different from other mnemonic factors that depend on the attempt to retrieve the target. Indeed, it may stem from a separate stage of processing occurring prior to that of attempted target retrieval.

Because, in the experiments that follow, we used RTs as our main dependent measure, and RTs have been interpreted as being especially relevant to one particular aspect of the retrieval—retrieval fluency—we now review those data. RT data have been used to support the suggestion that JOLs are based on the fluency of either original encoding (in the immediate JOL case) or of retrieval itself (Benjamin et al., 1998; see also Kelemen, 2000; Matvey, Dunlosky, & Guttentag, 2001; Rawson & Dunlosky, 2002, for related processing-fluency effects on JOLs). In support of the retrieval-fluency hypothesis, in the first experiment reported by Benjamin et al., participants were asked general information questions. They were told that the time that it took to answer each question was of primary interest to the experimenters. Thus, participants were told to press the enter key as soon as possible after coming up with the answer. Once the enter key was pressed, the question disappeared, and participants typed in their answers. Then, they were asked to indicate the likelihood (on a scale of 0 to 100) that they would remember the answer on a free recall test to be given 20 min later. Benjamin et al. called this judgment *predicted free recall*, but it is very much like a traditional JOL, so we will use

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the latter terminology. After a 10-min distractor task, the participants were given a free recall test. RT scores on retrieval just prior to making JOLs were calculated, and each participant's individual data set was split into quartiles based on his/her own median response latency. For each quartile, for each participant, average JOLs were calculated. Items for which an answer was provided were included in the analyses. Their results showed that JOLs increased with decreasing retrieval time. The mean JOL for their fastest RT quartile was 90; for the second fastest, 68; for the third, 65; and for the slowest, 52. Similar results were obtained in Experiment 2 using randomly presented words, and in Experiment 3 using cue–target pairs and a cued test. Benjamin et al. interpreted these results as indicating that people monitor their retrieval time and base their JOLs upon it: If the retrieval is quick and fluent, people give a high JOL; if the retrieval is slow and labored, people give a low JOL.

One problem with this interpretation, however, is that the participants were explicitly asked to retrieve prior to making the judgment. From these data then, it is unknown whether people would respond in the same way when they made a JOL without instructions to first retrieve. This issue has been addressed in only one study. Kelemen (2000) presented participants with a list of category names along with six exemplars—for example, a type of fuel: petroleum, alcohol, butane, water, uranium, and charcoal. Then, either immediately or after a delay, they were asked to make four types of responses. In Condition A, participants were asked to make a percentage confidence rating of how well they would recall the exemplars of the category when shown only the category name on a future test—a common JOL. In Condition B, participants were asked to say how many exemplars they would be able to recall on a future test—a modified type of JOL. In Condition C, participants were asked to say how many exemplars of a given category they were currently able to recall—requiring explicit retrieval. And, in Condition D, participants were asked to both say how many exemplars they could currently recall and tell how many they predicted that they would be able to recall on a future test. The delayed results showed that the judgments were most accurate in Conditions C and D, where participants had been required to attempt retrieval. Thus, the results of this study suggest that people's spontaneous judgments may be different in important ways from judgments that are elicited when the experimenter explicitly requires that people first retrieve. Thus, inferences about the mechanisms underlying spontaneous JOLs gleaned from studies such as those of Benjamin et al. (1998) or of Nelson et al. (2004), in which the experimenter has required certain processes—such as target retrieval—may be faulty. Experimenters may err by placing too much emphasis on the retrieval fluency heuristic or by omitting scrutiny of other stages or processes that may also be important.

A second problem with Benjamin et al.'s (1998) interpretation of their own findings as providing evidence for the retrieval fluency hypothesis is that their method for clustering the data, which showed a decreasing JOL accompanied by increasing RTs, does not reveal the relation between speed and JOL over the entire range. Even the lowest RT quartile reported by Benjamin et al. had a JOL mean of 52 (on a scale from 0 to 100), which is right in the middle of allowable JOLs. This method of clustering provides no information on what happens to RTs with low JOLs. It is possible that the decreasing JOLs observed by Benjamin et al. as a function of RT and interpreted as supporting the retrieval fluency hypothesis were ascribable to inappropriate averaging due to their method of clustering. Alternatively, they might have occurred because mid-range JOLs are especially slow, and would not truly show a monotonic increase in RT with decreasing JOL.

If retrieval fluency, operationalized as speed, is used as the cue for determining what JOL a person should assign to a given stimulus, then a monotonic effect should obtain whether one examines the data by investigating JOL as a function of RT (which is what Benjamin et al., 1998, did) or RT as a function of JOL. No matter which way one analyzes, low JOLs should be accompanied by slow RTs. Retrieval fluency would be unusable as a heuristic cue to indicate what JOL should be assigned unless it were monotonic over the range of JOLs. If sometimes a fast RT meant the person should give a high JOL and other times it meant they should give a low JOL, then RT alone would not be diagnostic.

There is reason to suppose that the RTs of the low JOLs, which may be obscured by Benjamin et al.'s (1998) analysis, might be especially important for understanding the processes contributing to the judgments and might not be particularly slow. In a different experimental paradigm, Kolers and Palef (1976) showed that people are sometimes able to make very fast "don't know" judgments. Very low JOLs might be of this ilk. Furthermore, Nelson and Narens (1990) investigated people's RTs with a different metacognitive judgment, the FOK. FOKs are typically made on items that the person is unable to recall. Interestingly, although participants cannot recall the target item at the time of the judgment, they can predict rather accurately whether they will later be able to recognize the answer. Several researchers have investigated the relations among various metacognitive judgments, such as FOKs, JOLs, and confidence judgments (about whether or not a person was correct in his or her recall, for example; see Bacon et al., 1998; Costermans, Lories, & Ansay, 1992; Nelson, 1984; Roy-Byrne et al., 1987; Schnyer et al., 2004). The conclusion has been (Leonesio & Nelson, 1990) that these judgments are typically not correlated with one another and, furthermore, that a person who is good at (or impaired on) one type of judgment is not necessarily good at (or impaired on) another. There does not appear to be a global metacognitive capability that governs all of these judgments. Therefore, we cannot conclude that because a finding has obtained with one type of judgment it will necessarily also obtain with another. Nevertheless, the fact that Nelson and Narens (1990) showed that the FOK reaction times were inverse-U shaped—both extremely high and extremely low FOKs were made quickly—is suggestive. The same might be true of JOLs.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Although in the JOL paradigm the theoretical emphasis has been on retrieval, in the FOK paradigm, as noted above, researchers have isolated two processes contributing to the judgments: one retrieval-related, but the other related to familiarity with the cue. Regarding the former, Koriat (1993) showed that retrieval fluency and target accessibility (information related to retrieval of the target) can affect people's FOKs. He had participants memorize four-letter nonsense strings to remember for a later memory test. At test, participants were asked to recall the strings aloud and as quickly as possible. The latency to initiate recall, or to pronounce the first letter, was measured by having participants speak into a voice-activated microphone. After recall, participants were asked to make FOKs on every item, regardless of the accuracy of the letters reported. As expected, Koriat's primary results showed that FOK increased as the number of letters reported increased. Furthermore, in his analyses, mean recall latencies for each participant were split at the median. Then the mean FOKs were calculated for slow and fast latencies. His results showed that the mean FOKs at the slow level were significantly lower than those at the fast level, suggesting a retrieval fluency mechanism. Koriat (1993) concluded that recall latencies are diagnostic of the accuracy of the information retrieved and that FOKs are based on them.

There is now overwhelming empirical support, in the FOK paradigm, for the contribution of another factor—the familiarity of the cue—in the determination of these judgments (Glenberg, Sanocki, Epstein, & Morris, 1987; Klin, Guzman, & Levine, 1999; Metcalfe, 1993; Metcalfe, Schwartz, & Joaquim, 1993; Miner & Reder, 1994; Nhouyvanisvong & Reder, 1998; Reder, 1987, 1988, 1996; Reder & Ritter, 1992; Reder & Schunn, 1996; Schunn, Reder, Nhouyvanisvong, Richards, & Stroffolino, 1997; Schwartz & Metcalfe, 1992). For example, Schwartz and Metcalfe presented participants with paired associates and asked for FOKs in the presence of only the cue. Prior to presenting the list, however, they also manipulated the accessibility of some of the cues by preexposing them in a pleasantness-rating task, and they did the same for a different set of target items. Their results showed that priming the cues influenced FOKs, but that priming targets did not. They concluded that the strength of the memory trace—that is, target retrievability—did not influence the FOKs, but that the familiarity of the cue did. Metcalfe et al. found similar results in an interference theory paradigm in which the familiarity of the cues and the retrievability of the targets were systematically varied.

Another line of evidence that has been used to support the notion that people use cue familiarity for making a judgment rather than using only retrieval information comes from the RTs that they exhibit. People are sometimes quicker to make a metacognitive judgment than they are to retrieve (Miner & Reder, 1994; Reder, 1987; Reder & Ritter, 1992). It follows that they cannot, therefore, be making those judgments on the basis of what they have retrieved because they simply have not had time to retrieve yet. For example, in several experiments that used a game-show paradigm, Reder and colleagues (e.g., Reder, 1987) asked participants a series of trivia questions. The participants were told to imagine that they were competing against another contestant and to make FOKs as quickly as possible. The results showed that participants could make these FOKs quicker than they could retrieve the answers. Furthermore, the judgments were usually accurate in predicting subsequent accuracy of the answer. Reder (1987) also found that priming words in the question led to increased subjective estimates of knowing the answer, despite the fact that this exposure did not improve actual rates of producing the correct answer. Reder and her colleagues concluded that FOKs have a quick early mechanism that relies on cues presented prior to retrieval (Reder, 1996; Reder & Schunn, 1996). These results provide support for the idea that people use cue familiarity when making FOKs and that they may do so rather quickly. Indeed, even Koriat and Levy-Sadot (2001) have recently acknowledged the importance of the cue-processing stage.

Here, we investigate whether JOLs are based on retrieval alone or whether they too, like FOKs, might have an earlier quick stage of processing based, perhaps, on cue familiarity. Basically, either participants made JOLs alone or they were told to retrieve and then make JOLs (as was the case in the Benjamin et al., 1998, study, and as Nelson et al., 2004, recommended as being a methodological advance useful for studying spontaneous JOLs). If people normally retrieve in order to make their JOLs, then the two conditions should produce nearly identical results. However, if the processes involved in making spontaneous JOLs differ from those involved when the person first retrieves and then makes the judgment, we would expect RT differences to occur, especially at the low end of the JOL scale. The experiments that follow, which differ in the details of the procedure, addressed this issue.

EXPERIMENT 1A

The purpose of Experiment 1A was to investigate the hypothesis that JOLs are based only on retrieval. Participants were presented with cue–target pairs to study. Some participants were told to make only JOLs (JOL only); others were told to retrieve covertly and then make JOLs (covert retrieval + JOL); and still others were told to retrieve overtly and make JOLs (overt retrieval + JOL). The latter condition was included as a comparison condition to make sure that when told to covertly retrieve, people were actually attempting retrieval as they would when forced to give an answer aloud. The retrieval hypothesis predicted that, except for the time needed to express the response, in the overt condition, the RT data for all three conditions would be the same, and that JOLs would increase as RTs decreased. More specifically, if JOLs were based on attempted target retrieval alone, and if the evaluation process following attempted retrieval takes a constant amount of time, then the RT functions for JOLs (retrieval + evaluation) should be slower but have the same form as those observed for retrieval alone, and they should be

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the same as those given when the participants are required to retrieve before giving a JOL. Furthermore, if people rely on the speed of retrieval as a heuristic to determine their JOLs, the functions relating RT to JOLs in all three conditions should be monotonic: Slower times should be associated with lower JOLs.

Method

Participants

Fifty-eight introductory psychology students from Columbia University participated for course credit. There were 12 participants in the JOL-only condition, 22 in the covert retrieval + JOL condition, and 24 in the overt retrieval + JOL condition. All of the participants were treated in accordance with the ethical standards of the American Psychological Association.

Materials

The stimuli were 20 first name–last name pairs, randomly selected and paired for each participant from a list of 130 frequently used female first names and 150 frequently used last names in the United States, as identified by the 1990 U.S. Census Bureau (www.census.gov/genealogy/names/). The names are presented in the Appendix.

Procedure

The session consisted of three lists, each of which included a study phase, a judgment/retrieval phase, and a test phase. For analyses, we collapsed over lists. For each list, the participants were shown 20 name pairs one at a time on the computer screen for 6 sec each, with each pair presented three times in a different random order. The participants were instructed to study each pair for a later test. Six of these items were not included in the judgment/retrieval phase but were tested later. Across all three lists, a total of 60 name pairs were studied.

Following study, the participants were randomly assigned to one of three groups: (1) JOL only, (2) retrieve covertly and make JOLs, or (3) retrieve overtly and make JOLs. When covert retrieval was required, participants were told,

> You will now be presented with only the first names that you saw. Your task is to retrieve the target last name or as much of the target last name as you can, in your head. Once you think that you have recalled as much as you can, press return. Do not press return before you have recalled what you think is the correct last name or as close to it as possible.

When overt retrieval was required, the participants were told,

> You will now be presented with only the first names that you saw. Your task is to try to retrieve the target last name for each first name. Once you have recalled the name, say it aloud because we will be recording your responses. And, it is also important that at the same time you say it aloud, you press the return button. If you cannot recall it, then say "don't know" out loud at the same time you press the return button.

For the JOL-only condition, the participants were told,

> Your task will be to type in a number, on a scale from 0 to 10, of how well you think that you would be able to remember the target last name on a future memory test. If you are absolutely sure that you will remember, then type in a 10. If you feel that you will have no chance of remembering, then type a 0. If you are somewhere in the middle, type in a number in between accordingly. This is called a judgment of learning, or JOL.

The participants in the retrieval conditions were given both the retrieval instructions and the JOL instructions. The JOL buttons were positioned on the top row of the keyboard, from the tilde key to the "0" key. Stickers were placed over the keys, labeled "0" to "10." The participants were made aware of the sticker keys prior to beginning the experiment.

Notice that in the retrieval + JOL conditions, the participants made two responses. The first was a buttonpress to indicate that they had retrieved as much as they could, allowing the researchers to obtain RTs for the retrieval process. The other response was the JOL itself, which was also timed. Together, these two RTs allowed us to test whether the judgment added a constant to the retrieval process, as we had predicted on the basis of previous research. After a 5-min distractor task, the participants were tested on the name pairs in a random order.

Results

A probability level of $p < .05$ was used as the criterion for statistical significance. Estimates of effect size (η^2) were also calculated. We analyzed the data using both absolute RTs and normalized z scores for each participant. All of the statistical results were similar, so we present only the absolute data here.

RT Functions for Retrieving and Then Making the Judgment Versus Just Retrieving

We first plotted, in panel A of Figure 1, as individual points, the RTs to make the first buttonpress—that is to complete retrieval—and the RTs to make the second buttonpress—that is, to make the JOL—in the covert and overt retrieval conditions. As can be seen from the figure, the judgment stage added a constant to the amount of time needed for retrieval.

In order to analyze statistically, we first fitted bivariate regression scores in both the overt and covert retrieval + JOL conditions to the differences between the retrieval RTs and the retrieval + JOL RTs. The quadratic component was −.004 in the covert condition and .003 in the overt condition, indicating that nothing curvilinear was going on. We then conducted ANOVAs, one for slopes and one for intercepts, on the linear components of the overt retrieval + JOL and covert retrieval + JOL conditions, each at two levels: the retrieval RTs and retrieval + JOL RTs. There was no effect of slope ($F < 1$), showing that the two functions were parallel. There was a main effect of intercept [$F(1,43) = 260.01$, $MS_e = 0.42$], indicating that making the JOL added a fixed amount of time to retrieval. These data indicate that when people are asked to retrieve and then make a JOL, they engage in a process of attempted retrieval in which RTs are monotonically related to JOL in a judgment stage, which adds a constant amount of time to the process.

RT Functions for Retrieving and Then Making a JOL Versus Simply Making a JOL

Our main question concerned what people did when they were not asked to retrieve. If they based their JOLs on retrieval, then the RT function for the JOL-only condition should have been parallel to, but slower than, the RT functions for the retrieval components of the retrieval + JOL conditions. These RT functions, based on JOLs alone, are shown in panel B of Figure 1 along with the RT points for the overt and covert retrieval conditions. The data fitted in this graph for the retrieval conditions are for the first buttonpress—that is, for the time needed to retrieve. As can be seen in the figure, the results for the JOL-only condition were not as expected: Its RT function was not parallel to those of the overt and covert retrieval conditions.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

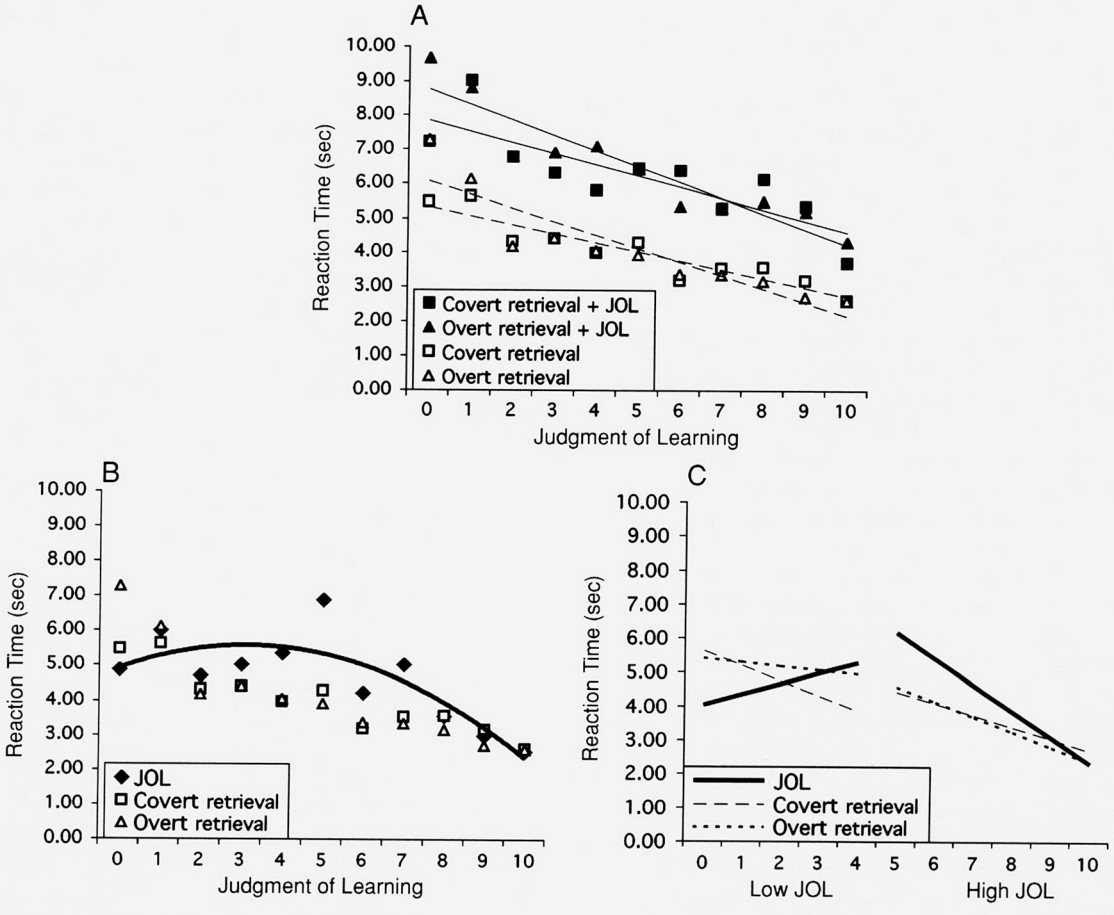

Figure 1. RTs as a function of JOL for Experiment 1A. (A) Mean RTs for the retrieval components and JOL components of the covert retrieval + JOL and overt retrieval + JOL conditions, showing that the judgment time is additive. (B) Mean RTs in the JOL-only condition and for the retrieval components in the covert retrieval + JOL and overt retrieval + JOL conditions. The curve shown is the best-fitting quadratic function for the JOL-only condition, which had a significant quadratic component; the functions for the other conditions, which had no quadratic components, were linear. (C) Mean RT best-fitting lines for the three conditions for high-JOL (between 5 and 10) and low-JOL (between 0 and 4) items.

After fitting bivariate regression curves for each participant's RTs as a function of level of JOL for each of the three conditions, we conducted one-sample *t* tests on both the linear and quadratic regression coefficients for each condition individually. Results showed a *quadratic* coefficient that was significantly positive [$M = .77$; $t(11) = 4.88$, $SE = 0.157$] in the JOL-only condition.[1] Neither of the other conditions in which retrieval had been required revealed a trace of a quadratic component. Because of the nonlinearity in the JOL-only condition, it was inappropriate to use a simple ANOVA to compare this condition to the retrieval conditions. A different analysis, therefore, is presented below.

To investigate further and analyze these results, the data were divided for each subject into JOLs from 5 to 10 and from 0 to 4.[2] We fitted two lines for each participant who had sufficient data for this analysis:[3] one for the high-JOL items and one for the low-JOL items. As can be seen from the mean best-fitting lines presented in panel C of Figure 1, although the functions for both of the retrieval conditions and the JOL-only condition were similar for the high-JOL items, this was not true for the low-JOL items. Two ANOVAs—one for the slopes of the functions and the other for the intercepts (with the intercepts for the high-JOL items category being calculated at the JOL = 5 level)—were conducted in a 2 (JOL level: high-JOL vs. low-JOL items) × 3 (condition: covert retrieval + JOL, overt retrieval + JOL, and JOL-only) design.

Slopes. There was a significant interaction between condition and JOL level for the slopes [$F(2,39) = 8.64$, $MS_e = 2.55$, $\eta^2 = .31$], such that for the high-JOL items, there were no differences among the conditions (JOL-only condition, $M = -0.77$; covert retrieval condition, $M = -0.35$; overt retrieval condition, $M = -0.44$), and all were significantly negative. In contrast, for the low-JOL items, whereas the two retrieval conditions had negative slopes (covert retrieval condition, $M = -0.56$, overt retrieval condition, $M = -0.39$), the JOL-only condition showed a positive slope ($M = 0.31$). In *t* tests, it was found that although there was no difference between

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the two retrieval conditions on the low-JOL items' slopes, the JOL-only condition was different from both the covert retrieval [$t(17) = 4.92$, $SE = 0.17$] and the overt retrieval [$t(32) = 3.35$, $SE = 0.21$] conditions for these items. These results contradict the hypothesis that JOLs, when people are not explicitly told to first retrieve and then make their judgments, are based entirely on retrieval fluency. The very fast low JOLs in the JOL-only condition suggest that a quick stage of processing was occurring prior to attempted target retrieval.

Intercepts. In interpreting this analysis, one should recall that in the two retrieval conditions, we took, as our dependent measure, the time it took to make the first buttonpress by which the participant indicated that he or she had attempted retrieval. (This RT was used, in part, because in the overt retrieval condition, the time to make a JOL depended on people's speed at saying aloud the retrieved answer, which was not important for our purposes. Also, as we showed in the first section of the results, making the judgment added a constant amount of time to this first buttonpress.) In the JOL-only condition, the dependent measure was the total time to make the JOL. Because we here compared the time to retrieve and the time to make a spontaneous JOL, we can use the logic introduced by Reder in her 1987 article, in which she argued that if the metacognitive judgment RT is as fast as or faster than the time to retrieve, then the person cannot be retrieving first to make the judgment. In terms of the following analysis, if people were retrieving first in the JOL-only condition, then the RT intercept in that condition should have been higher (i.e., slower) than those for retrieval, as given by the RT intercepts in the two retrieval + JOL conditions. An ANOVA showed that there was a significant interaction between JOL level and condition for the intercepts [$F(2,39) = 3.51$, $MS_e = 56.31$, $\eta^2 = .15$]. For the high-JOL items, there was a trend indicating that the intercept for the JOL-only condition ($M = 10.66$) was higher than those for both the covert retrieval condition ($M = 6.53$) [$t(17) = 1.77$, $SE = 2.33$, $p = .09$] and the overt retrieval condition ($M = 6.95$) [$t(32) = 1.73$, $SE = 2.14$, $p = .09$]. This is consistent with the idea that for the high-JOL items, the participants, even in the spontaneous JOL-only condition, were first attempting retrieval. However, for the low-JOL items, the intercept in the JOL-only condition ($M = 4.04$) was not different (and, indeed, was numerically lower!) than those for both the covert retrieval condition ($M = 5.63$) [$t(17) = 1.64$, $SE = 0.97$, $p = .11$] and the overt retrieval condition ($M = 5.41$) [$t(32) = 1.48$, $SE = 0.93$, $p = .14$], which were not different from one another. This interaction indicated that the JOLs, at least for the lowest JOL items in the JOL-only condition, could not have been made by first attempting retrieval and then making a judgment based on the outcome.

The analyses based on the slopes and the intercepts converged: Low-JOL items in the JOL-only condition were different from the equivalent items in either of the conditions in which people were explicitly told to retrieve.

Final Performance

Means are presented in Table 1. There was a main effect of JOL on final performance [$F(1,39) = 183.98$, $MS_e = 5.85$, $\eta^2 = .83$], such that people performed better on the high-JOL items than on the low-JOL items, as expected. There was also a significant effect of condition [$F(2,39) = 5.39$, $MS_e = 0.29$, $\eta^2 = .22$], such that the overt retrieval + JOL condition yielded better performance than did the JOL-only condition. The interaction was not significant.

Discussion

The most important finding in Experiment 1A was that the RTs for the low-JOL items in the JOL-only condition were unlike those found for either of the conditions in which the participants were required to retrieve before making a JOL. These results disconfirmed our hypothesis that JOLs are made solely on the basis of retrieval goodness or fluency. To be specific, for the low-JOL items, RTs became faster as the JOLs decreased when people were making JOLs alone, whereas RTs became slower as JOLs decreased when people had to retrieve, in both the overt and covert retrieval conditions.

The one obvious methodological difference between the retrieval conditions and the JOL-only condition was that in the retrieval conditions, people made two buttonpresses—one for retrieving the target and another for making the JOL. In the JOL-only condition, only one buttonpress—the one for making the JOL—was made. This methodology was important in allowing us to directly compare the time people needed to retrieve to the time it took them to make a JOL, providing the theoretically important information that people can make low JOLs more quickly than they can retrieve. Nevertheless, it was important to investigate the RTs directly (without differential intervening buttonpresses) for making simple JOLs

Table 1
Mean Final Performance for Low-JOL and High-JOL Items for Experiments 1A, 1B, and 2

	Experiment 1A		Experiment 1B		Experiment 2	
Condition	Low JOL	High JOL	Low JOL	High JOL	Low JOL	High JOL
JOL only	.08	.74	.18	.67	.18	.68
Covert retrieval + JOL	.23	.81	.39	.60	.37	.73
Overt retrieval + JOL	.35	.86				

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

as compared with making JOLs following retrieval. In the next experiment, we sought to replicate our findings in Experiment 1A while avoiding the double buttonpress.

EXPERIMENT 1B

Experiment 1B was similar to Experiment 1A, except that only two conditions were included: (1) JOL only and (2) (covert) retrieval + JOL. Given that the RT functions were highly similar for the overt and covert retrieval conditions in Experiment 1A, it appeared that we could eliminate the overt condition. Furthermore, because our previous analyses had indicated that the judgment itself adds a constant amount of time following retrieval, it was not necessary to measure retrieval RTs separately. Thus, we could eliminate the double buttonpress in this experiment. Therefore, in the retrieval + JOL condition, although we asked people to retrieve before making the JOL, their only overt response was their JOL rating. According to the retrieval hypothesis, the RTs should have been the same in both conditions and monotonically increasing with JOL. However, the results of Experiment 1A suggested that we would find an interaction such that RTs would increase monotonically with decreasing JOL in the retrieval condition, whereas in the JOL-only condition the RT JOL function would be nonmonotonic, with fast very high and fast very low JOLs in evidence.

Method

Thirty-two introductory psychology students from Columbia University participated for course credit. Participants were randomly assigned to either the JOL-only or the retrieval + JOL condition. There were 16 participants in each condition. Instructions and methods were the same as in Experiment 1A except that in the retrieval + JOL condition, participants did not overtly respond after retrieval.

Results

RT Functions

As in Experiment 1A, we fitted bivariate regression curves for each participant's RTs as a function of JOL. One-sample t tests showed that the JOL-only condition had a *quadratic* coefficient that was significant (M = 1.53) [$t(15) = 6.48$, $SE = 0.023$; see note 1]. The retrieval + JOL condition did not have a quadratic component. These RTs are shown in panel A of Figure 2.

The two best-fitting lines for the high-JOL items and low-JOL items in each of the two conditions are shown in panel B of Figure 2. There was a significant interaction between condition and JOL level (high/low) for the slopes [$F(1,28) = 4.52$, $MS_e = 10.43$, $\eta^2 = .14$]. For the low-JOL items, the slope of the retrieval + JOL condition (M = −1.08) was different from the slope of the JOL-only condition ($M = 0.96$) [$t(28) = 2.89$, $SE = 0.71$]. For the high-JOL items, the slopes were negative for both the JOL-only condition ($M = -0.80$) and the retrieval + JOL condition ($M = -1.17$) and did not differ from one another.

The interaction was also significant for the intercepts [$F(1,28) = 8.65$, $MS_e = 193.12$, $\eta^2 = .24$]. The low-JOL item intercept in the JOL-only condition ($M = 3.62$) was significantly lower than was that of the retrieval + JOL condition ($M = 12.69$) [$t(28) = 4.19$, $SE = 2.16$]. There were no reliable intercept differences for the high-JOL items.

Final Performance

Means are presented in Table 1. There was a main effect of JOL level on final performance [$F(1,28) = 35.15$, $MS_e = 1.85$, $\eta^2 = .87$], such that people performed better on the high-JOL items than on the low-JOL items. There was also a significant interaction between condition and JOL level [$F(1,28) = 5.41$, $MS_e = 0.28$, $\eta^2 = .24$]: No difference in recall was found for the high-JOL items, but a significant difference [$t(28) = 2.00$, $SE = 0.10$] did exist between the JOL-only and the retrieval + JOL conditions for the low-JOL items. On these items, people in the retrieval + JOL condition performed better than did those in the JOL-only condition.

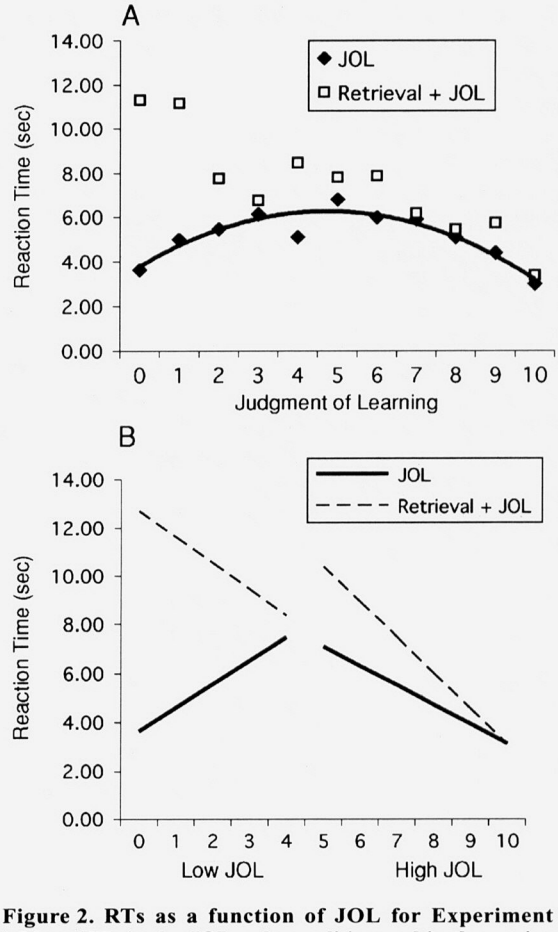

Figure 2. RTs as a function of JOL for Experiment 1B. (A) Mean RTs in the JOL-only condition and in the retrieval + JOL condition. The curve shown is the best-fitting quadratic function for the JOL-only condition, which had a significant quadratic component; the function for the retrieval + JOL condition, which had no quadratic component, was linear. (B) Mean RT best-fitting lines for the two conditions for high-JOL (between 5 and 10) and low-JOL (between 0 and 4) items.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Discussion

In Experiment 1B, we replicated the RT results that were found in Experiment 1A. The pattern of RT data that favors the hypothesis that retrieval fluency alone underlies JOLs was obtained when the participants were asked to retrieve and then make a JOL, but the pattern was different when the participants were asked simply to make JOLs. In this case, very low JOLs were made especially quickly, rather than especially slowly.

The second interesting finding was that in the JOL-only condition, recall performance was impaired on the items that were assigned low JOLs. It is possible that when the participants were instructed to attempt to retrieve first, and then make their JOLs, the retrieval attempt itself had a beneficial effect on performance. When the participants simply made their JOLs without any retrieval attempt, some items that might have benefited from the retrieval effort failed to receive that boost in performance. Thus, the final recall data in Experiment 1B suggest that there may be a second important difference in terms of what people do when they make JOLs without instructions as compared with when they are told to do so by attempting to retrieve. These findings correspond well to those presented by Kelemen (2000).

In these experiments, we used a 0–10 scale to assess judgments of learning. It is possible, however, that participants had some difficulty interpreting what the numbers on this scale referred to[4] and that, had we used a standard Likert scale based on the specific probability of remembering an item on a later test, such as that which has been used in other studies (Benjamin et al., 1998; Kelemen, 2000; Kelemen & Weaver, 1997; Matvey et al., 2001; Nelson & Dunlosky, 1991), our results might have been different. In Experiment 2, we conducted the same procedure as in Experiment 1B, except that rather than use a JOL scale from 0 to 10, we used a scale with percentage levels for JOLs and included very clear (standardly used) instructions about the meaning of the various levels of JOLs.

EXPERIMENT 2

The purpose of Experiment 2 was to replicate the previous two experiments and to confirm that our use of a particular scale (0–10) that might have been open to varied interpretation had not contributed to the discrepancy between our observed RT findings in the JOL-only condition and what we had expected—namely, monotonic increases in RT with decreases in JOL.

Method

Forty-three introductory psychology students from Barnard College participated for course credit. The participants were randomly assigned to either the JOL-only or the retrieval + JOL condition. There were 22 participants in the JOL-only condition and 21 participants in the retrieval + JOL condition. The methods were similar to those used in Experiment 1B except that participants were told to make JOLs on the exact scale that had been used in the previous JOL literature, with the following instructions: "How confident are you that in about 10 min from now, you will be able to recall the last name if given the first name of the pair?" Participants were asked to rate themselves as 0% confident (*definitely will not recall*), 20%, 40%, 60%, 80%, or 100% confident (*definitely will recall*). These were the same instructions that had been used in previous JOL research (Benjamin et al., 1998; Kelemen, 2000; Kelemen & Weaver, 1997; Matvey et al., 2001; Nelson & Dunlosky, 1991). The responses were made by pressing one of six keys on the top row of the computer keyboard labeled with 0%, 20%, 40%, 60%, 80%, and 100% stickers. The participants were told to use their middle three fingers on both hands and to keep their hand positions as stable as possible when making JOLs.

Results

RT Results

Figure 3 displays the mean RTs for each of the six JOL categories, 0%–100%, for both the JOL-only and retrieval + JOL conditions. As can be seen in the figure, the participants in the retrieval + JOL condition had monotonically increasing JOL–RT functions: shorter RTs for high JOLs and longer RTs for low JOLs—consistent with our earlier results. In contrast, the JOL-only condition had very quick, low JOLs, replicating nonlinear RT

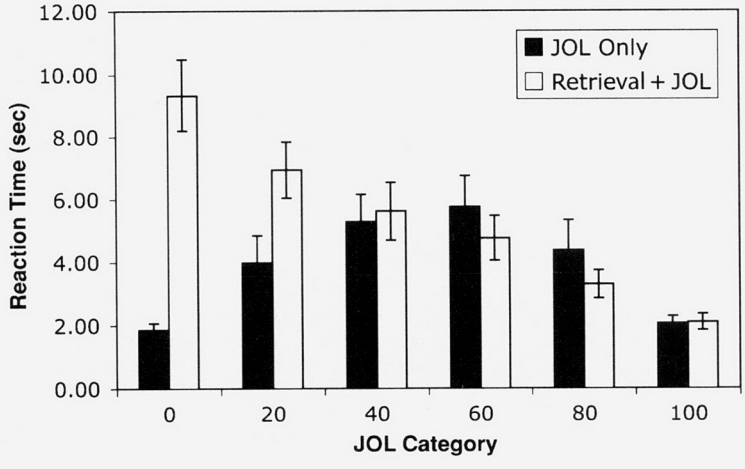

Figure 3. RTs as a function of JOL category (0%–100%) for both the JOL-only and the retrieval + JOL conditions for Experiment 2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

functions from the previous two experiments (see note 1). An ANOVA (see note 3) revealed a significant main effect of JOL category (0%–100%) [$F(5,135) = 18.62$, $MS_e = 50.81$, $\eta^2 = .41$] and a significant main effect of condition [$F(1,27) = 5.27$, $MS_e = 74.64$, $\eta^2 = .16$]. Most important, there was a significant interaction between JOL category and condition [$F(5,135) = 26.64$, $MS_e = 72.69$, $\eta^2 = .50$]. Post hoc t tests revealed faster RTs for the JOL-only condition at the JOL 0% level [$t(36) = 6.42$, $SE = 1.17$; JOL-only condition, $M = 1.85$ sec; retrieval + JOL condition, $M = 9.33$ sec] and at the JOL 20% level [$t(35) = 3.11$, $SE = 0.96$; JOL-only, $M = 3.97$ sec; retrieval + JOL, $M = 6.94$ sec]. The only other significant difference between the two conditions' RTs was at the JOL 80% level [$t(35) = 2.28$, $SE = 0.48$], where participants in the JOL-only condition spent slightly more time retrieving ($M = 4.38$ sec) than did those in the retrieval + JOL condition ($M = 3.28$ sec).

Final Performance

To investigate final performance, the data were grouped into high-JOL (60%, 80%, and 100%) and low-JOL (0%, 20%, and 40%) levels. The mean proportion correct for the high- and low-JOL levels for both conditions is presented in Table 1. The ANOVA resulted in a main effect of JOL level (high/low) as expected [$F(1,40) = 103.52$, $MS_e = 3.84$, $\eta^2 = .72$], such that people performed better on the high-JOL items than on the low-JOL items. There was also a condition trend in the direction that the retrieval + JOL condition did better than the JOL-only condition [$F(1,40) = 216.49$, $MS_e = 20.04$, $\eta^2 = .84$, $p = .07$]. Finally, there was a trend for the condition × JOL level interaction [$F(1,40) = 2.69$, $MS_e = 0.10$, $\eta^2 = .06$, $p = .10$], suggesting that any differences in performance between the two conditions occurred mainly at the low-JOL level.

GENERAL DISCUSSION

These results challenge the notion that JOLs are based only on the retrieval of target information. In three experiments, people studied cue–target pairs and were then asked either to make JOLs or to retrieve and then make JOLs. The retrieval-only hypothesis predicted that the RT functions would be parallel. Contrary to this prediction, in all three experiments, whereas the RT functions paralleled one another for the high-JOL items, they were in opposite directions for the low-JOL items, with the lowest spontaneous JOLs being made very quickly, not very slowly. These findings suggest that people do not spontaneously retrieve when making all JOLs and that low JOLs may be based on a stage of processing other than target retrieval. In particular, we suggest that JOLs, like FOKs, may be based on evaluation of the cues that are available prior to retrieval. In a manner similar to a quick preliminary assessment that can terminate retrieval for FOKs (Reder, 1987, 1996), it is plausible to suppose that there is a quick preliminary assessment stage that allows people to make a fast, low JOL, indicating that they "don't know."

Although a number of researchers have shown that fast "no" responses in a recognition test might be based on a stage prior to recollective retrieval involving an assessment of the global familiarity of the target (Atkinson, Hermann, & Wescourt, 1974; Glucksberg & McCloskey, 1981; Juola, Fischler, Wood, & Atkinson, 1971; Kolers & Palef, 1976; Nelson & Narens, 1990), we suggest that the analogy to our situation may not be entirely appropriate. For example, Kolers and Palef (1976) presented participants with a list of nonwords mixed with both high- and low-frequency nouns and asked them to judge whether each item could be used in a sentence. They plotted RTs on the y-axis and word type on the x-axis for nonwords and from low frequency to high frequency for real words. RT results showed an inverted-U–shaped function much like our own: The participants were not only quickest in responding "yes" to high-frequency words, but also unusually quick to respond "no" to nonwords. The investigators argued for a two-stage recognition process in which items of high and low familiarity may be identified quickly in an initial prerecollection or preretrieval assessment. A number of other researchers have proposed that recognition responses may be made via two processes: a fast, global appraisal of familiarity and a slow and deliberate recollective retrieval process (Gruppuso, Lindsay, & Kelley, 1997; Hintzman, Caulton, & Levitin, 1998; Jacoby, 1991; Jacoby, Jones, & Dolan, 1998; Juola & Atkinson, 1971; Mandler, 1980; Mandler & Boeck, 1974; Metcalfe, 1993, 1998). Although our data, like those cited above, indicate that people can quickly assess that they do not know, the familiarity mechanism suggested above seems inappropriate for our data because in those paradigms, the target was given and the task was to make a recognition decision, whereas in our case, no target was given. Thus, although our fast "don't know" judgments bear a surface similarity to fast "no" judgments in recognition, we think that their basis is, or at least might be, different—namely, that our fast "don't knows" might not entail evaluation of the target.

However, Nelson and Narens (1990) proposed a model of FOKs similar to the recognition memory models given above. Importantly, the FOK situation they modeled is similar to the JOL paradigm insofar as the target was not presented there either. They proposed that there are two counters, or accumulators—one for "know" information and one for "don't know" information—as is consistent with some recognition random walk models (e.g., Meyer, Irwin, Osman, & Kounios, 1988; Ratcliff, 1978, 1988) and race models (e.g., Jones & Anderson, 1982; Pike, 1973; Ratcliff & McKoon, 1997). The "affirmative-FOK" counter keeps track of accumulating affirmative information or features, whereas the "negative-FOK" counter keeps track of the accumulating negative or null information. Each counter has its own criterion level. Thus, if information is built up quickly in the negative counter, then a fast "don't know" FOK can occur. If information accumulates quickly in the affirmative counter, then fast "know" judgments can occur. Nelson and Narens supported their theory empirically with a nonmonotonic function similar

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

to ours, and the longest FOK response times occurred for the middle levels (Nelson & Narens, 1990).

It is easy to extend this reasoning to the JOL paradigm, and we do not dispute that such a dual-counter mechanism based on the retrieval of target information could produce the U-shaped function. The only logical point mitigating against this model as an explanation of our data is that if this were the mechanism underlying JOLs, and it entailed attempted target retrieval, there is no reason whatsoever for the participants not to have used it both when they were told to retrieve to make their JOLs and when they were told to simply make their judgments. Both conditions should have, by this mechanism, revealed U-shaped functions, but they should have, nevertheless, been the *same* as one another. Thus, we reject the target-based, dual-counter mechanism as an explanation of our data in favor of the two-stage mechanism in which the cue is recognized before (or, indeed, even in parallel with) attempted target retrieval.

In summary, the data presented here contradict a single-process retrieval mechanism of JOLs in which knowing that one knows or does not know occurs as a result of evaluating one's attempted target retrieval. Such an evaluation of retrieval cannot be the sole mechanism underlying JOLs. If it were, then when people were asked to retrieve, their data should have been identical to those found when they were simply asked to make JOLs, but they were not identical. Furthermore, low JOLs should have been associated not with very short but rather with very long RTs, such as those observed when people were instructed to retrieve before making their JOLs. When simply asked to make JOLs, fast, not slow, low JOLs were observed, indicating that people apparently knew that they did not know without having to explicitly attempt to retrieve. Indeed, the RTs for the very low judgments were as fast as or faster than the time to retrieve. These data suggest that JOLs, like other types of metacognitive judgments, are driven by two processes—an assessment of cue familiarity that may terminate further processing if unsuccessful and, conditionally, an evaluation of the results of attempted target retrieval.

REFERENCES

ATKINSON, R. C., HERMANN, D. J., & WESCOURT, K. T. (1974). Search processes in recognition memory. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium*. Potomac, MD: Erlbaum.

BACON, E., DANION, M.-J., KAUFFMANN-MULLER, F., SCHELSTRAETE, M.-A., BRUANT, A., SELLAL, F., & GRANGE, D. (1998). Confidence level and feeling of knowing for episodic and semantic memory: An investigation of lorazepam effects on metamemory. *Psychopharmacology*, **138**, 318-325.

BEGG, I., DUFT, S., LALONDE, P., MELNICK, R., & SANVITO, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory & 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.

COSTERMANS, J., LORIES, G., & ANSAY, C. (1992). Confidence level and feeling of knowing in question answering: The weight of inferential processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18**, 142-150.

DUNLOSKY, J., & HERTZOG, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 249-276). Mahwah, NJ: Erlbaum.

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 & Language*, **33**, 545-565.

DUNLOSKY, J., RAWSON, K. A., & MCDONALD, S. L. (2002). Influence of practice tests on the accuracy of predicting memory performance for paired associates, sentences, and text material. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 68-92). Cambridge: Cambridge University Press.

GLENBERG, A. M. (1987). Temporal context and recency. In D. S. Gorfein & R. R. Hoffman (Eds.), *Memory and learning: The Ebbinghaus Centennial Conference* (pp. 173-190). Hillsdale, NJ: Erlbaum.

GLENBERG, A. M., SANOCKI, T., EPSTEIN, W., & MORRIS, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, **116**, 119-136.

GLUCKSBERG, S., & MCCLOSKEY, M. (1981). Decisions about ignorance: Knowing that you don't know. *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 311-325.

GRUPPUSO, V., LINDSAY, D. S., & KELLEY, C. M. (1997). The process-dissociation procedure and similarity: Defining and estimating recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 259-278.

HINTZMAN, D. L., CAULTON, D. A., & LEVITIN, D. J. (1998). Retrieval dynamics in recognition and list discrimination: Further evidence of separate processes of familiarity and recall. *Memory & Cognition*, **26**, 449-462.

JACOBY, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, **30**, 513-541.

JACOBY, L. L., JONES, T. C., & DOLAN, P. O. (1998). Two effects of repetition: Support for a dual-process model of know judgments and exclusion errors. *Psychonomic Bulletin & Review*, **5**, 705-709.

JONES, W. P., & ANDERSON, J. R. (1982). Semantic categorization and high-speed scanning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **8**, 237-242.

JUOLA, J. F., & ATKINSON, R. C. (1971). Memory scanning for words versus categories. *Journal of Verbal Learning & Verbal Behavior*, **10**, 522-527.

JUOLA, J. F., FISCHLER, I., WOOD, C. T., & ATKINSON, R. C. (1971). Recognition time for information stored in long-term memory. *Perception & Psychophysics*, **10**, 8-14.

KELEMEN, W. L. (2000). Metamemory cues and monitoring accuracy: Judging what you know and what you will know. *Journal of Educational Psychology*, **92**, 800-810.

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, & Cognition*, **23**, 1394-1409.

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 & Language*, **32**, 1-24.

KIMBALL, D. R., & METCALFE, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Memory & Cognition*, **31**, 918-929.

KLIN, C. M., GUZMAN, A. E., & LEVINE, W. H. (1999). Prevalence and persistence of predictive inferences. *Journal of Memory & Language*, **40**, 593-604.

KOLERS, P. A., & PALEF, S. R. (1976). Knowing not. *Memory & Cognition*, **4**, 553-558.

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

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

KORIAT, A., & LEVY-SADOT, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 34-53.

LEONESIO, R. J., & NELSON, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 464-470.

MAKI, R. H. (1998). Predicting performance on text: Delayed versus immediate predictions and tests. *Memory & Cognition*, **26**, 959-964.

MANDLER, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, **87**, 252-271.

MANDLER, G., & BOECK, W. J. (1974). Retrieval processes in recognition. *Memory & Cognition*, **2**, 613-615.

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. (1993). Novelty monitoring, metacognition, and control in a composite holographic associative recall model: Implications for Korsakoff amnesia. *Psychological Review*, **100**, 3-22.

METCALFE, J. (1998). Cognitive optimism: Self-deception or memory-based processing heuristics? *Personality & Social Psychology Review*, **2**, 100-110.

METCALFE, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General*, **131**, 349-363.

METCALFE, J., & KORNELL, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General*, **132**, 530-542.

METCALFE, J., SCHWARTZ, B. L., & JOAQUIM, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 851-861.

MEYER, D. E., IRWIN, D. E., OSMAN, A. M., & KOUNIOS, J. (1988). The dynamics of cognition and action: Memory processes inferred from speed–accuracy decomposition. *Psychological Review*, **95**, 183-237.

MINER, A. C., & REDER, L. M. (1994). A new look at feeling of knowing: Its metacognitive role in regulating question answering. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 47-70). Cambridge, MA: MIT Press.

NELSON, T. O. (1984). A comparison of current measures of the accuracy of feeling of knowing predictions. *Psychological Bulletin*, **95**, 109-133.

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

NELSON, T. O., DUNLOSKY, J., GRAF, A., & NARENS, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science*, **5**, 207-213.

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). San Diego: Academic Press.

NELSON, T. O., & NARENS, L. (1994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1-25). Cambridge, MA: MIT Press.

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.

NHOUYVANISVONG, A., & REDER, L. M. (1998). Rapid feeling-of-knowing: A strategy selection mechanism. In V. Y. Yzerbyt, G. Lories, & B. Dardenne (Eds.), *Metacognition: Cognitive and social dimensions* (pp. 35-52). London: Sage.

PIKE, A. R. (1973). Response latency models for signal detection. *Psychological Review*, **80**, 53-68.

RATCLIFF, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**, 59-108.

RATCLIFF, R. (1988). Continuous versus discrete information processing: Modeling accumulation of partial information. *Psychological Review*, **95**, 238-255.

RATCLIFF, R., & MCKOON, G. (1997). A counter model for implicit priming in perceptual word identification. *Psychological Review*, **104**, 319-343.

RAWSON, K. A., & DUNLOSKY, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 69-80.

REDER, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, **19**, 90-138.

REDER, L. M. (1988). Strategic control of retrieval strategies. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 277-259). San Diego: Academic Press.

REDER, L. M. (ED.) (1996). *Implicit memory and metacognition*. Mahwah, NJ: Erlbaum.

REDER, L. M., & RITTER, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18**, 435-451.

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. 45-78). Mahwah, NJ: Erlbaum.

ROY-BYRNE, P. P., UHDE, T. W., HOLCOMB, H., THOMPSON, K., KING, A. K., & WEINGARTNER, H. (1987). Effects of diazepam on cognitive processes in normal subjects. *Psychopharmacology*, **91**, 30-33.

SCHNYER, D. M., VERFAELLIE, M., ALEXANDER, M. P., LAFLECHE, G., NICHOLLS, L., & KASZNIAK, A. W. (2004). A role for right medial prefrontal cortex in accurate feeling-of-knowing judgments: Evidence from patients with lesions to frontal cortex. *Neuropsychologia*, **42**, 957-966.

SCHUNN, C. D., REDER, L. M., NHOUYVANISVONG, A., RICHARDS, D. R., & STROFFOLINO, P. J. (1997). To calculate or not to calculate: A source activation confusion model of problem familiarity's role in strategy selection. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 3-29.

SCHWARTZ, B. L., & METCALFE, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 1074-1083.

SON, L. K. (2004). Spacing one's study: Evidence for a metacognitive control strategy. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 601-604.

SON, L. K., & METCALFE, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 204-221.

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.

WEAVER, C. A., III, & KELEMEN, W. L. (1997). Judgments of learning at delays: Shifts in response patterns or increased metamemory accuracy? *Psychological Science*, **8**, 318-321.

NOTES

1. When we plotted JOL as a function of RT quartile, as Benjamin et al. (1998) had done, the results of the JOL-only condition showed a linear effect in all three experiments. In Experiment 1A, the mean JOLs for each RT quartile, from fastest to slowest, were 7.98, 7.18, 5.72, and 4.29. In Experiment 1B, the mean JOLs for each quartile were 6.73, 6.35, 6.00, and 4.72. In Experiment 2, the mean JOLs for each quartile were 71.48, 66.24, 64.65, and 60.92.

2. The split was also conducted for items with JOL = 6–10 (high) and JOL = 0–5 (low). This change did not alter the pattern of any of the results.

3. There were a few cases in which individuals did not select every JOL rating on the scale, making the degrees of freedom inconsistent across some of the statistical analyses.

4. A reviewer had asked to see the JOL rating frequency distributions for the retrieval and JOL-only conditions, which we analyzed for all three experiments (0–10 in Experiments 1A and 1B; 0%–100% in Ex-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

periment 2), because he or she thought that differences in the frequency distributions might be attributable to systematic differences in the interpretation of the JOL rating scale between conditions. For all three experiments, there was a main effect of JOL [Experiment 1A, $F(10,260) = 15.20$, $MS_e = 426.31$, $\eta^2 = .37$; Experiment 1B, $F(10,300) = 44.79$, $MS_e = 1{,}032.80$, $\eta^2 = .60$; Experiment 2, $F(5,210) = 24.82$, $MS_e = 2{,}838.56$, $\eta^2 = .37$], such that responses to the highest JOL rating were most frequent. There were no main effects or interactions with condition in Experiment 1A. In Experiment 1B, there was an interaction between JOL and condition [$F(10,300) = 4.95$, $MS_e = 114.12$, $\eta^2 = .14$], such that people in the retrieval + JOL condition ($M = 26.19$) responded more frequently to the JOL-10 items than did those in the JOL-only condition ($M = 15.50$) [$t(30) = 2.84$, $SE = 3.76$], whereas at the JOL-2 [$t(30) = 3.97$, $SE = 1.24$] and JOL-3 [$t(30) = 2.41$, $SE = 0.99$] levels there was a slight tendency for a higher frequency in the JOL-only condition (4.50 vs. 0.81 and 3.38 vs. 1.00, respectively). None of the other JOL levels showed any difference. There were no differences in the frequency distributions in Experiment 2.

APPENDIX

First Names

ABIGAIL	ELIZABETH	LEAH	ROBIN
ALICE	EMILY	LILLIAN	ROSE
AMANDA	EVELYN	LINDA	ROSIE
AMY	FAITH	LISA	RUBY
ANDREA	FELICIA	LOIS	RUTH
ANGELA	FRANCES	LORI	SANDRA
ANNA	GABRIELLA	LOUISE	SARA
ANNIE	GERALDINE	MARGARET	SHARON
ASHLEY	GLORIA	MARIA	SHIRLEY
BARBARA	GRACE	MARILYN	SOPHIA
BECKY	HEATHER	MARTHA	STELLA
BETTY	HEIDI	MARY	STEPHANIE
BEVERLY	HELEN	MELISSA	SUSAN
BONNIE	HOPE	MICHELLE	SYLVIA
BRENDA	IRENE	MILDRED	TAMMY
CARLA	ISABEL	MINDY	TERRY
CAROLYN	JACQUELINE	NANCY	THERESA
CATHERINE	JANET	NATASHA	TIFFANY
CECILIA	JANICE	NICOLE	TINA
CHERYL	JEAN	NORMA	TRACEY
CHRISTINE	JENNIFER	OLIVIA	VANESSA
COURTNEY	JESSICA	PAIGE	VICTORIA
CYNTHIA	JOAN	PAMELA	VIRGINIA
DARLENE	JOYCE	PATRICIA	VIVIAN
DEBORAH	JUDY	PAULA	WANDA
DENISE	JULIA	PEARL	WENDY
DIANE	KAREN	PEGGY	WHITNEY
DOLLY	KATHLEEN	PHYLLIS	WILDA
DONNA	KATHRYN	RACHEL	WILMA
DORIS	KATHY	RAMONA	WONDA
DOROTHY	KELLY	REBECCA	YVONNE
DOTTIE	KIMBERLY	RENEE	
ELAINE	LAURA	RHONDA	

Last Names

ADAMS	CARTER	FISHER	HAYES
ALEXANDER	CLARK	FLORES	HENDERSON
ALLEN	COLE	FORD	HENRY
ANDERSON	COLLINS	FOSTER	HERNANDEZ
BAILEY	COOK	FREEMAN	HICKS
BAKER	COOPER	GARCIA	HILL
BARNES	COX	GIBSON	HOLMES
BELL	CRAWFORD	GOMEZ	HOWARD
BENNETT	CRUZ	GONZALEZ	HUGHES
BLACK	CUNNINGHAM	GORDON	HUNTER
BOYD	DANIELS	GRAHAM	JACKSON
BROOKS	DAVIS	GRAY	JAMES
BROWN	DIAZ	GREEN	JENKINS
BRYANT	DIXON	GRIFFIN	JOHNSON
BURNS	EDWARDS	HALL	JONES
BUTLER	ELLIS	HAMILTON	JORDAN
CAMPBELL	EVANS	HARRISON	KELLY

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

APPENDIX (Continued)

KENNEDY	MURRAY	RICHARDSON	THOMPSON
KIM	MYERS	RIVERA	TORRES
KING	NELSON	ROBERTS	TUCKER
LAGOS	ORTIZ	ROBINSON	TURNER
LEE	OWENS	RODRIGUEZ	WALKER
LEWIS	PARKER	ROGERS	WALLACE
LONG	PATTERSON	ROSS	WALSH
LOPEZ	PEREZ	RUSSELL	WARD
MARSHALL	PERRY	SANCHEZ	WARREN
MARTIN	PETERSON	SANDERS	WATSON
MARTINEZ	PHILLIPS	SCOTT	WEBB
MASON	PORTER	SHAW	WELLS
MCDONALD	POWELL	SIMMONS	WEST
MILLER	PRICE	SIMPSON	WHITE
MITCHELL	RAMIREZ	SMITH	WILLIAMS
MOORE	RAMOS	STEVENS	WILSON
MORALES	REED	STEWART	WRIGHT
MORGAN	REYES	SULLIVAN	YATES
MORRIS	REYNOLDS	TAYLOR	YOUNG
MURPHY	RICE	THOMAS	

(Manuscript received December 19, 2003;
revision accepted for publication October 14, 2004.)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.