

Using the past to predict the future

MICHAEL R. DOUGHERTY, PETRA SCHECK, and THOMAS O. NELSON
University of Maryland, College Park, Maryland

and

LOUIS NARENS
University of California, Irvine, California

This research addressed three issues. First, we examined whether retrospective confidence judgments (RCJs) and judgments of learning (JOLs) assess memory differently. Second, we examined the relative accuracy of JOLs and RCJs at predicting future recall performance. Third, we examined whether making JOLs improves subsequent recall better than making RCJs or making no metacognitive judgment. Results suggest that RCJs and JOLs are both based on retrievability, but that participants use their memory differently when making JOLs. RCJs were more accurate than JOLs at predicting future recall for some subsets of items, but the reverse was true for other subsets of items. Finally, eventual recall performance was facilitated when participants made JOLs but not when they made RCJs, suggesting that the JOL task helps to improve people's learning of the items.

The study of confidence judgments has played an important role in psychological research. The topic touches several areas within psychology, including judgment and decision making, social psychology, educational psychology, and cognitive psychology, and it has generated related research in cognitive neuropsychology (Shimamura & Squire, 1988) and abnormal psychology (Pappas et al., 1992). Two general classes of confidence judgments that have received increasing attention are those that concern predictions about the future retrieval of events, called *judgments of learning* (JOLs), and those that concern assessments about past retrieval, called *retrospective confidence judgments* (RCJs). When one makes a JOL, one is essentially assessing the likelihood that one will be able to recall, at some point in the future, a particular item (e.g., a word) when cued with its corresponding stimulus. When one makes an RCJ, one is assessing the likelihood that what one has just recalled is indeed correct.

The conceptual distinction between JOLs and RCJs is straightforward. JOLs are assumed to tap a mechanism that enables us to predict future memory performance, but the exact nature of this mechanism still is poorly understood (Nelson & Dunlosky, 1992). In contrast, RCJs are used to assess the correctness of past retrieval and often are assumed to arise from the retrievability of the to-be-judged item from memory (Costermans, Lories, & Ansary,

1992; Dougherty, 2001; Kelley & Lindsay, 1993).¹ However, despite this conceptual difference, little research has directly investigated whether JOLs and RCJs tap functionally different underlying psychological processes. Obviously, an examination of the processes underlying JOLs and RCJs is of considerable theoretical interest, for it may be possible to develop integrative theories to account for both types of judgments.

A comparison of JOLs and RCJs is also of practical interest especially for people in educational psychology. For example, suppose students are interested in assessing how well they have mastered the material for an upcoming examination. As part of their study strategy, they might attempt to recall the answers to a set of study questions. Are they better off predicting the likelihood that they will retrieve the correct answer on the upcoming exam (a JOL), or are they better off assessing their confidence in their answer to the study question (an RCJ)? The students' ultimate goal, of course, is to determine whether they have learned the material well enough to succeed on an upcoming exam. However, which type of judgment, the JOL or the RCJ, is a better predictor of future retrieval?

TWO GENERAL QUESTIONS

This article addresses two questions. First, are JOLs and RCJs based on the same underlying cognitive processes? Although considerable research has investigated JOLs and RCJs, few attempts have been made to study these two types of judgments within a single experiment. Thus, the extent to which JOLs and RCJs are based on the same cognitive processes is largely unknown.

Second, assuming JOLs and RCJs indeed tap different underlying processes, are JOLs better than RCJs in predicting future recall? Although one might presume that

This research was partially supported by Grant SES-0134678 from the Decision Research and Management Sciences program, National Science Foundation, and by Grant 305H030283 from the Cognition and Student Learning (CASL) research program at the Institute of Education Sciences, U.S. Department of Education. Correspondence concerning this article can be addressed to M. R. Dougherty, Department of Psychology, University of Maryland, College Park, MD 20742 (e-mail: mdougherty@psyc.umd.edu).

JOLs are better predictors of future recall, this has not been tested.

Are JOLs and RCJs Based on the Same Information?

Empirical distinction between JOLs and RCJs.

There is much to suggest that JOLs and RCJs are based, at least in part, on similar components of memory. For example, considerable evidence suggests that both RCJs and JOLs involve an assessment of retrievability. RCJs have been proposed to arise from the fluency of retrieval (Kelley & Lindsay, 1993), trace strength (Costermans et al., 1992), and familiarity (Dougherty, 2001) and are often found to be influenced by variables that affect retrievability (Koriat, 1997). Similarly, several studies have suggested that JOLs are based on retrievability. For example, Lovelace (1984) found that JOLs were more predictive of previous recall than of future recall, suggesting that participants' JOLs were influenced by their previous recall attempt (for similar results, see also Connor, Dunlosky, & Hertzog, 1997; King, Zechmeister, & Shaughnessy, 1980). Matvey, Dunlosky, and Guttentag (2001) found that retrieval fluency (i.e., retrieval time) was negatively correlated with JOL magnitude. Sommer, Heinz, Leuthold, Matt, and Schweinberger (1995) found that EEG wave forms elicited for JOLs corresponded to wave forms elicited by a subsequent recognition memory test. Finally, Nelson and Dunlosky (1991) reported that 19 of 20 participants who made delayed JOLs claimed to have attempted to recall the correct answer when forming their judgments. Together, these studies suggest that both RCJs and JOLs are based, at least to some extent, on the retrievability of the item to be judged.

Although the evidence above indicates that both RCJs and JOLs are based on retrievability, there is also evidence that JOLs are based on information other than, or in addition to, retrievability. Busey, Tunnicliff, Loftus, and Loftus (2000) found that the luminance of the faces at study and at test, stimulus duration, and degree of rehearsal all affected JOLs. In contrast, RCJs were affected primarily by degree of rehearsal, with only small effects for test luminance. This suggests that JOLs are based on at least somewhat different cues than are RCJs.

Differences between RCJs and JOLs have also been found in neuropsychological studies. Shimamura and Squire (1988) investigated the relative accuracy of RCJs in Korsakoff amnesiacs, normals, and alcoholic controls and found no decrement in the accuracy of RCJs in Korsakoff's patients. However, Bauer, Kyaw, and Kilbey (1984) found that JOL accuracy was degraded among Korsakoff patients, and Shimamura and Squire (1986) found that the accuracy of feelings of knowing (FOK—i.e., predictions that one will be able to recognize a previously unrecalled item) was more impaired in Korsakoff's patients than it was in the control groups. Apparently, among Korsakoff's patients, the processes responsible for RCJs are intact, but the processes responsible for FOKs and JOLs are damaged. Whereas RCJs, as elaborated in Nelson and Na-

rens (1990), are retrospective judgments (about the past), FOKs and JOLs are both prospective (about the future). Similar dissociations have been found between FOKs and RCJs in Alzheimer's patients (Pappas et al., 1992). These studies suggest that prospective judgments of retrieval (JOLs and FOKs) and retrospective judgments (RCJs) may draw on different processes or cues.

Previous theories of JOLs versus RCJs. The empirical similarities between JOLs and RCJs are paralleled by similarities in theoretical approaches. Two current views of the basis of JOLs are the direct-access (King et al., 1980) and cue-utilization (Koriat, 1997) theories. Direct-access theory posits that the future recallability of an item is assessed by monitoring the strength of the item to be recalled, and JOLs are assumed to be based on a readout of this strength value. According to this view, the correlation between JOLs and recall arises because factors that affect the accessibility (or strength) of the memory trace affect both JOL magnitude and recall (Koriat, 1997).

An alternative view of JOLs is offered by Koriat's (1997) cue-utilization theory, in which it is assumed that JOLs are based to varying degrees on intrinsic cues, extrinsic cues, and mnemonic cues. Intrinsic cues are inherent properties of the to-be-learned items (e.g., word frequency, concreteness, vividness) that might be correlated with later retrieval. Extrinsic cues are the characteristics of the learner or the learning environment, such as the conditions under which learning takes place and the encoding operations used during study. Mnemonic cues are internal assessments of the degree to which an item has been learned. For example, degree of learning might be determined by assessing the accessibility of the to-be-remembered item, processing fluency, or item or cue familiarity (Koriat, 1997). Intrinsic, extrinsic, and mnemonic cues are not independent; intrinsic and extrinsic cues can both affect mnemonic cues. Whereas mnemonic cues are presumed to have only direct effects on JOLs, intrinsic and extrinsic cues are presumed to affect JOLs either directly or indirectly by influencing the mnemonic cues. For example, study time might affect mnemonic processes, such as latency of retrieval, but might also directly affect JOLs if people use study time as a predictor of an item's future recallability (e.g., one might hold the belief that longer study intervals increase the probability of recall).

Just as with research on JOLs, researchers studying RCJs have proposed two main theoretical approaches: the trace-strength view and a cue-based assessment of confidence. According to the trace-strength view, confidence is based on an assessment of the overall strength of the underlying memory representation (Clark, 1997; Dougherty, 2001; Dougherty, Gettys, & Ogden, 1999). Similar to the direct-access view of JOLs, the trace-strength view posits that confidence judgments are based on an assessment of the overall familiarity (or strength) of the item to be judged. Clark (1997), for example, showed that confidence-accuracy inversions could be accounted for by a familiarity-based model of recognition memory

such as Hintzman's (1988) MINERVA 2 model. Likewise, Dougherty (2001) showed that the overconfidence effect (the tendency for percentage correct to lag behind mean confidence) could be accounted for by MINERVA-DM, a modified version of MINERVA 2. Moreover, Dougherty reported that variables that were assumed to affect the trace strength (encoding quality and study repetition) also affected RCJs and the relationship between RCJs and accuracy. In any case, both of these models of RCJs assume that confidence judgments are based on the strength (or familiarity) of the item to be judged, as assessed by a global-matching process.

An alternative view of RCJs that has been applied primarily within the overconfidence literature is embodied in Gigerenzer, Hoffrage, and Kleinböling's (1991) PMM theory. According to PMM theory, confidence judgments in general knowledge questions are based on cues extracted from the judgment environment or from assessments of memory. Confidence is assumed to be based on the ecological validity of the most discriminating cue. Although PMM is fundamentally different from Koriat's (1997) cue-utilization theory in that it focuses on the use of environmental cues, both theories share the idea that confidence is based on the degree to which various cues are predictive of performance. In the case of PMM, cues are assumed to be extracted from the environment. In the case of cue-utilization theory, cues are assumed to be based on the properties of the stimulus materials (intrinsic cues), the properties of the encoding operations (extrinsic cues), and the properties of the retrieval process (mnemonic cues).² Whereas the cue-utilization theory implies that participants integrate information from various cues to form a JOL, PMM theory assumes that participants utilize one particular cue in forming an RCJ (Gigerenzer & Goldstein, 1996; Goldstein & Gigerenzer, 2002).

In sum, theory and research on JOLs and RCJs suggest that these two types of judgments share at least some psychological processes. Also, it is important to note that the two kinds of theory—direct access and utilization of cues—are not mutually exclusive; that is, both could be valid. However, the empirical evidence cited above suggests at least some differences between JOLs and RCJs. The present studies provide a direct test of these differences.

Are JOLs Better Than RCJs at Predicting Future Recall?

Assuming that JOLs and RCJs are based, at least in part, on different information, one might hypothesize that JOLs would be better predictors of future recall than RCJs would, especially because JOLs are specifically targeted at predicting future recall performance. However, it is also possible to imagine that JOLs might actually be worse than RCJs at predicting future recall. Research on test-retest recall performance has revealed that the probability of a successful recall is higher following a previously successful recall than it is following a previously unsuccessful recall (Estes, Hopkins, & Crothers, 1960). Thus, past recall provides a good cue for predicting future recall.

If RCJs are based on retrievability, they should be predictive of future retrieval as well: Items that are retrievable at time t should have a higher probability (in comparison with items not retrieved at time t) of being retrieved at time $t+1$. In contrast, if JOLs are based in part on cues other than retrievability, it is possible for JOLs to be less predictive of future recall than RCJs are. For example, suppose JOLs were based on an assessment of retrievability plus at least one additional cue (e.g., study time). If the additional cues were predictive of future recall performance and participants integrated the cues optimally, then JOLs should be more predictive of future recall than RCJs would be. However, if the additional cues were less related to recall than is retrievability, JOLs might be less accurate than RCJs at predicting future recall.

OVERVIEW OF EXPERIMENTS AND HYPOTHESES

Modification to the Standard JOL Paradigm

To investigate our two questions of interest—that is, are JOLs and RCJs based on different cognitive processes, and, if so, are JOLs functionally better than RCJs at predicting future recall?—we modified the standard delayed-JOL paradigm using paired associates as stimuli. In the standard task, participants typically study a set of paired associates so that they can, at some future point, respond with the target word when prompted with the cue. A short time (e.g., 30 sec) after studying each pair, participants are provided with the cue and are asked to judge the likelihood that they will be able to recall the target at test when prompted with the cue, with the test occurring sometime in the future (e.g., 10 min later). Instead of the standard delayed-JOL paradigm, we used the prejudgment recall and monitoring methodology (PRAM) proposed by Nelson, Narens, and Dunlosky (2004), in which participants are required to make two recall attempts: a prejudgment recall attempt that occurs approximately 30 sec after study and a final recall attempt that occurs approximately 10 min after study. RCJs and/or JOLs were made immediately after prejudgment recall. This allowed us to examine the extent to which both RCJs and JOLs rely on the retrievability of the item to be judged (prejudgment recall) and the extent to which both judgments are predictive of future recall (final recall). Note that the requirement to engage in prejudgment recall should, if anything, bias our results toward the null hypothesis because forcing all participants to engage in recall should make it more likely that retrievability is utilized as a cue. Thus, finding reliable differences in these judgments would be particularly compelling evidence that the two judgments tap different processes.

Hypotheses Under Investigation

Question 1. Our first empirical question concerned whether JOLs and RCJs made following prejudgment recall are based on the same cognitive processes. The trace-access and trace-strength views anticipate that JOLs and RCJs are based on assessments of the retrievability of

the item. Thus, one might expect few, if any, differences between JOLs and RCJs if both types of judgments are based on retrievability. However, if JOLs rely on cues other than, or in addition to, retrievability (as proposed by cue-utilization theory), we would expect several differences between JOLs and RCJs to emerge. These hypothesized differences are elaborated next.

If JOLs and RCJs rely on different processes, this should produce differences in the correlations between JOLs and prejudgment recall rather than between RCJs and prejudgment recall. Moreover, if RCJs rely more on retrievability than JOLs do, then RCJs should be more highly correlated with the latency of prejudgment recall. Also relevant to addressing whether JOLs and RCJs are based on different information is whether both JOLs and RCJs, and their accuracy, are affected similarly by independent variables. In our experiments, we manipulated study time (3 vs. 12 sec) as an independent variable. Under Koriat's (1997) cue-utilization theory, study time is considered an extrinsic cue and should have both indirect effects and direct effects on JOL accuracy. For example, if the amount of time spent studying an item is correlated with recall performance, then study time can affect mnemonic cues (e.g., latency of retrieval). Also, if participants believe that longer study intervals lead to better recall performance, then JOLs should be higher for items studied for 12 sec. Again, if both types of judgments rely on the same processes, study time should not have differential effects on JOL or RCJ magnitude, nor should it have differential effects on the correlation between those judgments and recall accuracy. Whether increased study time improves the accuracy of judgments is an open question. Indeed, study time should improve JOL accuracy only if it is correlated with recall accuracy and if participants utilize the cue appropriately. Overweighting the importance of longer study time might actually lead to a decrease in the accuracy of JOLs.

Question 2. Our second empirical question was whether JOLs are more accurate than RCJs at predicting future (final) recall. This question is addressed by comparing the gamma correlations between final recall accuracy and the metacognitive judgments made after prejudgment recall. On the basis of the fact that JOLs are targeted to predict future recall, we hypothesized that JOLs would be more accurate than RCJs at predicting final recall. However, as previously mentioned, there are several circumstances that might lead JOLs to be less accurate than RCJs in predicting final recall. One potential cue that might differentially affect the accuracy of JOLs and RCJs is study time. For example, if RCJs are based on retrievability and if study time affects retrievability, then both recall and RCJ magnitude should increase with increased study time. The net result might not change the accuracy of RCJs. However, according to cue-utilization theory, study time can have two effects. As noted earlier, extrinsic cues, such as study time, can have both direct and indirect effects. Study time can have a direct effect on JOLs if participants interpret it to be relevant to predicting future recall, and it

can have an indirect effect by influencing mnemonic cues such as latency of recall (which might be an index of retrievability). Thus, if RCJs are based on retrievability but JOLs are based on both retrievability and study time, then the study time variable should have differential effects on JOLs and RCJs.

EXPERIMENT 1

Method

Materials

The items to be remembered were 482 pairs of unrelated, concrete nouns (e.g., *ocean-tree*) from Paivio, Yuille, and Madigan (1968). Of these, 52 were designated as target items (and were the same for all participants). The remaining 430 word pairs served as a pool from which dummy and distractor items could be sampled.

Participants

The participants were 60 undergraduate students who participated to fulfill course requirements and who were enrolled in an introductory psychology course at the University of Maryland.

Design and Procedure

The experimental design was a $3 \times 3 \times 2$ factorial, with number of dummy items (0, 1, or 2), duration of item presentation (3 or 12 msec), and judgment type (RCJs vs. JOLs) as within-participants variables. The experiment consisted of five phases, which were blocked such that participants cycled through the five phases 52 times throughout the course of the experiment.

Study phase. The participants were instructed to study the item pairs so that they could remember the second word when prompted with the first. They were instructed to study all items but were tested on only items designated as targets (whose identity the participants did not know in advance). The first four trials served as practice and were not included in recall. The order of presentation was randomized anew for each participant, and word pairs were displayed on the computer monitor for 3 or 12 sec each. The duration of the presentation for targets was random, except for two restrictions: (1) that half of the targets be presented for 3 sec and the other half be presented for 12 sec, and (2) that no more than two successive trials had targets presented for the same duration.

Preceding the presentation of each target, 0, 1, or 2 dummy items were each displayed for 3 or 12 sec each. Both the number of dummy items (0, 1, or 2) and the presentation rate of the dummy items were random, with the restrictions that, across all trials, (1) half would be presented for 3 sec and half for 12 sec, and (2) no more than two consecutive dummy items would be presented for the same duration.

Immediately after the offset of each target item, distractor items were presented. Distractors were randomly assigned a presentation rate of either 3 or 12 sec, with these restrictions: (1) Half of the distractors were presented for each duration, (2) the total distraction interval was 30 sec, and (3) within each trial, no more than two items of the same presentation duration occurred successively (i.e., there were always two distractors presented at the 3-sec rate and two presented at the 12-sec rate). The dummy and distractor items were included so that participants could not anticipate which item they would be asked to recall at prejudgment recall. (We wanted to prevent the participants from focusing in on the word pair that they would be required to recall.) In addition, the four distractor items also served to provide a filled 30-sec interval between the presentation of the target and prejudgment recall of the target. Therefore, although the number of distractors between target and judgment was constant for all participants, the number of dummy items preceding the target was not necessarily the same for all participants (a maximum of 70 and a minimum of 17 dummy items). Thus, the total number of studied word pairs ranged from a minimum of 277 to a

maximum of 330 word pairs across participants. Note, however, that because the number of dummy items was determined randomly, it was not confounded with our manipulations of interest (study time and judgment type).

Prejudgment recall. Following presentation of the distractor items, the participants completed a self-paced recall test for the target item. Participants saw the first word of the item (the cue) and were required to recall the second word (the target). Recall was achieved by typing the target word into the computer, and recall latency was measured by recording how much time elapsed between the onset of the cue word and when the participants pressed the enter key to submit their response. To prevent omission errors, the participants were required to guess. Immediately after making their recall response, they were prompted to provide their metacognitive judgments.

Metacognitive judgments. The participants made both an RCJ and a JOL for each studied item, with the RCJ always made first. For RCJs, the participants were presented with the cue (the first word of the studied pair) and asked, "How confident are you that the reply you gave for this item is correct (0 = definitely not correct, 20 = 20% sure, 40 . . . , 60 . . . , 80 . . . , 100 = definitely correct)?" For JOLs, the participants were again presented with the cue but were asked, "How confident are you that in about 10 minutes you will be able to recall the second word of the item when prompted with the first (0 = definitely won't recall, 20 = 20% sure, 40 . . . , 60 . . . , 80 . . . , 100 = definitely will recall)?" The participants responded by typing using the number pad on the keyboard.

Final recall test. After the metacognitive judgment on the final item (word pair), the participants completed a self-paced recall test. The participants saw the first word of each item and were required to recall the second word. The 100 targets were divided into four blocks of 25 items per block. To approximately equate the time between study and test for all items, the pairs studied in the first quarter of the study phase were tested in the first quarter of the recall test, pairs studied in the second quarter of the study phase were tested in the second quarter of the recall test, and so on.

Final RCJ. After making the final recall response for a given item, the participants made a second RCJ using the same procedure as described above.

Results and Discussion

All differences that are reported as reliable had $ps < .05$. All pairwise comparisons reported as reliable used Bonferroni adjusted t tests to control familywise alpha at .05.

Are RCJs and JOLs Based on the Same Information?

If RCJs and JOLs are based on the same information and the same weighting of that information, then (as elaborated above) we would expect RCJs and JOLs to be (1) equally well correlated with prejudgment recall, (2) equally well correlated with latency of prejudgment recall, and (3) affected equivalently by the study time manipulation (this should be true both for the judgment magnitude and the correlation between judgment and final recall).

Figure 1 shows the mean gamma correlations between prejudgment recall and metacognitive judgment. We hypothesized that both RCJs and JOLs were based, at least to some extent, on retrievability. Consistent with this hypothesis, both RCJs and JOLs were correlated with prejudgment recall with gammas above .86, all of which were significantly greater than zero (all $ps < .01$). However,

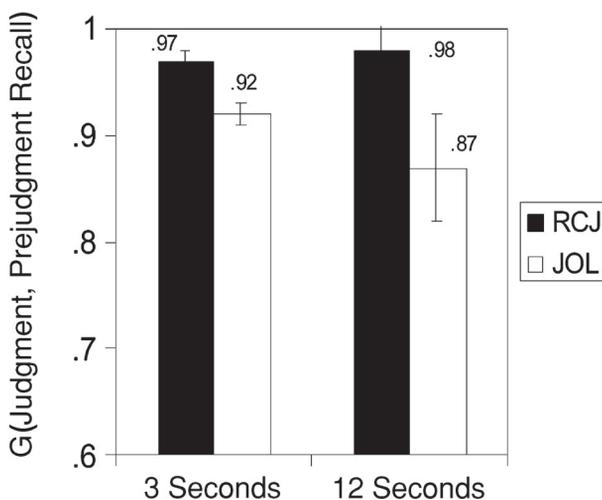


Figure 1. Mean overall gammas between prejudgment recall and magnitude of the judgment for Experiment 1. Error bars represent standard error of the mean.

closer inspection of prejudgment recall and its relationship with judgment suggests that RCJs and JOLs do not rely on prejudgment recall to the same extent. As indicated in Figure 1, the correlation between metacognitive judgment and prejudgment recall was significantly higher for RCJs than for JOLs [$F(1,55) = 7.31, p < .01$]. However, the effect of study time on the correlation between metacognitive judgment and prejudgment recall was non-significant [$F(1,55) = 0.27$] and there was no study time by judgment type interaction [$F(1,55) = 0.41$].³

Particularly diagnostic for examining the relationship between judgment and prejudgment recall are the correlations between judgment magnitude and latency of prejudgment recall. Prior research has revealed that higher RCJs are associated with shorter retrieval latencies (Barnes, Nelson, & Dunlosky, 1999; Costermans et al., 1992; Nelson & Narens, 1990), and the same pattern has been found for JOLs (Nelson, 1996). However, no research that we know of has directly examined whether RCJs and JOLs are correlated to the same degree with retrieval latency. Consequently, we examined the correlation between judgment magnitude and latency, collapsing both across correct and incorrect prejudgment recall and separately for correctly versus incorrectly recalled items. When correctness of prejudgment recall was ignored, both RCJs and JOLs were significantly correlated (gamma correlations) with latency of prejudgment recall [$M_{RCJ} = -.33, t(61) = 11.49, p < .01$; $M_{JOL} = -.21, t(61) = 7.81, p < .01$]. Note, however, that the absolute value of this correlation was significantly greater for RCJs than for JOLs [$t(60) = 3.34, p < .01$], suggesting that RCJs are more closely tied to retrievability than are JOLs. Thus, longer RTs were associated with lower magnitude judgments (especially among RCJs), which is to be expected if judgments are based on retrievability (as proposed by Barnes et al., 1999).

Of particular interest in examining the relationship between judgment magnitude and latency of recall were analyses that examined correct and incorrect prejudgment recall separately, as shown in Table 1. Of these correlations, the only one not statistically greater than zero was the correlation between JOL and latency of prejudgment recall for *correctly* recalled prejudgment items [$t(60) = 1.49, p > .10$]. The correlation between JOL magnitude and latency of retrieval was significantly different from zero only for *incorrectly* recalled items [$t(48) = 4.98, p < .01$]. Pairwise t tests using a Bonferroni adjustment ($\alpha' = \alpha/4 = .0125$) indicated that there was no difference between correct and incorrect prejudgment recall in how well RCJs were correlated with latency of prejudgment recall [$t(54) = 2.09, p = .04$]. However, there was a significant difference between correct and incorrect prejudgment recall for JOLs [$t(48) = 3.77, p < .01$]. Thus, although both judgments were related to speed of retrieval, JOLs were related to speed of retrieval only for incorrectly recalled prejudgment items.⁴ There was no difference between how well JOLs and RCJs correlated with latency of incorrectly recalled items [$t(48) = 0.71, p > .10$], but RCJs showed a higher correlation than did JOLs with latency for correctly recalled items [$t(48) = 3.77, p < .01$]. The finding that RCJs were more highly correlated with latency of prejudgment retrieval among correctly recalled items suggests that JOLs and RCJs relied on different information primarily when prejudgment recall was successful, or that the RCJ intervening between recall and the JOL diluted the relationship between recall and magnitude of JOL.

One way to examine more closely the differences between JOLs and RCJs is to plot the distribution of confidence judgments conditionalized on whether prejudgment recall was correct or incorrect. These graphs are shown in Figure 2 separately for JOLs (top panel) and RCJs (bottom panel). Two aspects of the graphs are particularly revealing. First, there are no noteworthy differences between JOLs and RCJs in the distributions for the incorrectly recalled items. This suggests that when prejudgment recall fails, RCJs and JOLs are based on the same processes or information. In contrast, when prejudgment recall is successful, the participants utilized the middle of the confidence scale more often when making JOLs. The majority of the RCJs were given a rating of 100%. Thus,

when prejudgment recall was correct, the participants either relied on different information when making the two judgments or utilized the confidence scale differently.

A third aspect of the data suggesting that the two judgments tap different processes was judgment magnitude. Table 2 shows the mean judgments for RCJs and JOLs. There were main effects of study time [$F(1,61) = 165.79, p < .01$] and judgment type [$F(1,61) = 49.18, p < .01$] on the magnitude of metacognitive judgment, as well as a study time \times judgment type interaction [$F(1,61) = 14.02, p < .01$] (see Table 1 for means). However, post hoc tests using Bonferroni t tests ($\alpha' = .0125$) revealed that both RCJs and JOLs were higher for items studied for 12 sec than for items studied for 3 sec and that RCJs were higher than JOLs for both the 12-sec and 3-sec conditions, suggesting that the interaction may have been scale dependent. In any case, the finding that RCJ magnitude was consistently higher than JOL magnitude lends converging evidence for the idea that the JOLs and RCJs are based on different information.

Are JOLs Better Than RCJs at Predicting Future Recall?

Were JOLs better or worse than RCJs at predicting future recall? The mean gamma correlations between metacognitive judgment and final recall are presented in Figure 3. As can be seen, RCJs were more accurate than JOLs in predicting future recall, as evidenced by the main effect of judgment type on the correlation between judgment and final recall [$F(1,56) = 23.30, p < .01$]. Equally interesting is the finding that study time affected the accuracy of JOLs more than RCJs, as evidenced by the significant judgment type \times study time interaction [$F(1,56) = 4.04, p < .05$]. Post hoc tests using a Bonferroni adjustment ($\alpha' = .0125$) revealed that JOLs were significantly less accurate than RCJs in both the 3-sec [$t(58) = 3.18, p < .01$] and 12-sec [$t(58) = 3.72, p < .01$] study conditions. The decrease in accuracy from 3 to 12 sec was significant for JOLs [$t(58) = 2.25, p = .029$], but nonsignificant for RCJs [$t(56) = .02, p > .10$].

Are the Differences Between RCJs and JOLs Due to Differences in Scale Use?

One explanation of the differences observed between JOLs and RCJs is that participants use the confidence

Table 1
Mean Gamma Correlations Between Latency of Prejudgment Recall and Magnitude of Judgment Computed Separately for Correct Versus Incorrect Prejudgment Recall

	Experiment 1				Experiment 2			
	Correct Prejudgment Recall		Incorrect Prejudgment Recall		Correct Prejudgment Recall		Incorrect Prejudgment Recall	
	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>
Gamma (latency, JOL)	-.05	.03	-.17	.04	-.08	.03	-.34	.06
Gamma (latency, RCJ)	-.35	.07	-.31	.05	-.23	.04	-.17	.05

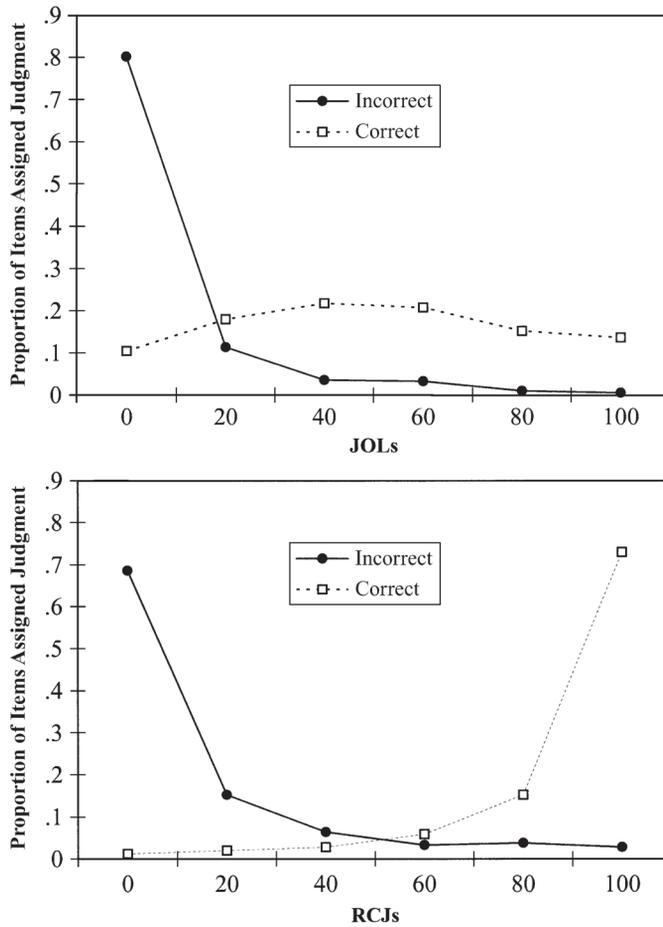


Figure 2. Distribution of responses for JOLs (top panel) and RCJs (bottom panel) conditionalized on whether pre-judgment recall was correct or incorrect for Experiment 1.

scale differently for the two types of judgments without relying on differences underlying cognitive processes. One way to assess this possibility is to examine part-scale gamma correlations between judgment magnitude and pre-judgment recall accuracy, between judgment magnitude and latency of pre-judgment recall, and between judgment magnitude and final recall accuracy. Part-scale

gammas were computed by grouping pairs of items for which one item received a confidence rating of J and the other item a confidence rating of K. For example, the first pair of points in Figure 4 shows the mean gamma correlation between judgment magnitude and pre-judgment recall accuracy for the subset of each participant's judgments where one item (word pair) received a rating of 0 and the

Table 2
Mean Gamma Correlations Between Judgment and Prejudgment Recall, Judgment and Final Recall, Mean Judgments, and Mean Correlations Between Prejudgment Recall and Final Recall

	Presentation Rate: Experiment 1								Presentation Rate: Experiment 2							
	3 sec				12 sec				3 sec				12 sec			
	JOLs		RCJs		JOLs		RCJs		JOLs		RCJs		JOLs		RCJs	
	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>
Mean magnitude of judgment	.32	.02	.58	.03	.39	.03	.67	.03	.46	.03	.57	.03	.54	.03	.65	.03
Gamma between judgment and pre-judgment recall	.92	.03	.97	.01	.87	.05	.98	.01	.92	.02	.96	.01	.93	.02	.96	.01
Gamma between judgment and final recall	.79	.04	.86	.04	.63	.06	.85	.04	.76	.04	.84	.04	.68	.06	.84	.04
Gamma between pre-recall and final recall		.99		.01		.97		.02		.99		.00		.99		.00

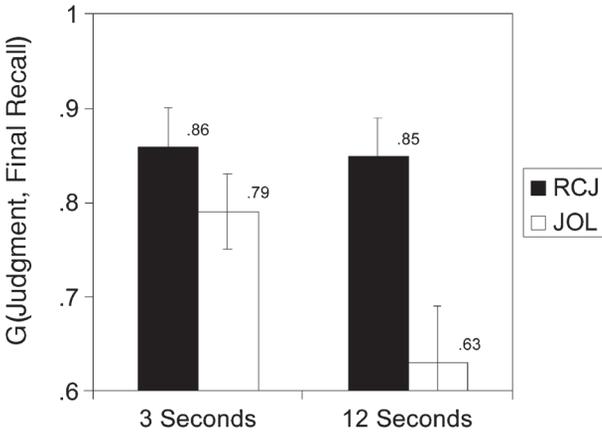


Figure 3. Mean overall gammas between final recall and magnitude of the judgment for Experiment 1. Error bars represent standard error of the mean.

other a rating of 20. The second pair of points shows the mean gamma for observations where one item was given a 0 and the other a 40, and so forth. These part-scale gammas were computed for individual participants, collapsing across study time.⁵ Note that by comparing RCJs and JOLs within parts of the scale, we control for differential scale use. Thus, if differences between RCJs and JOLs exist after controlling for scale use, it would strengthen the case that the two judgments are based on different information or/and on different utilizations of information. Moreover, if RCJs were still more highly correlated with final recall, it would rule out the possibility that the RCJ advantage was an artifact of scale use.

Figure 4 shows the part-scale gammas between judgment magnitude and prejudgment recall accuracy. The

vertical line separates the dyads in which one member of the dyad received a zero judgment from the dyads in which both members received nonzero judgments. Two findings are of interest. First, in 12 out of 15 of the comparisons of JOL and RCJ for a given pair of magnitudes of judgments, RCJs were more highly correlated than JOLs were with prejudgment recall accuracy. This difference is significant by a sign test ($p < .02$). Second, in 9 out of 10 comparisons in which both members of the dyad received a confidence judgment greater than zero, RCJs were more highly correlated with prejudgment recall than JOLs were, which is also significant by a sign test ($p < .02$). The pattern of results for dyads in which one member received a zero did not consistently favor RCJs or JOLs. Note that the pattern of results in Figure 4 qualifies the interpretation of the overall gammas presented in Figure 1, which showed that higher overall correlations occurred between RCJs and prejudgment recall than between JOLs and prejudgment recall. That is, the part-scale gamma correlations show that RCJs are better predictors of prejudgment recall than are JOLs only when both items in the dyad are given nonzero judgments.

Figure 5 shows the part-scale gamma between judgment magnitude and latency of prejudgment recall. As can be seen, in 11 out of the 15 possible comparisons, RCJs were more highly correlated with latency of prejudgment recall than were JOLs. By a sign test, the probability of 11 observations favoring RCJs by chance is .059. Consistent with the gamma correlations between prejudgment recall and judgment shown in Figure 4, the advantage for RCJs over JOLs was consistent for only nonzero dyads (right of the vertical line): RCJs were more highly correlated with latency of prejudgment recall than were JOLs for 8 of the 10 comparisons ($p = .055$) in which both members of the dyad had judgments greater than zero, but for only 3 out

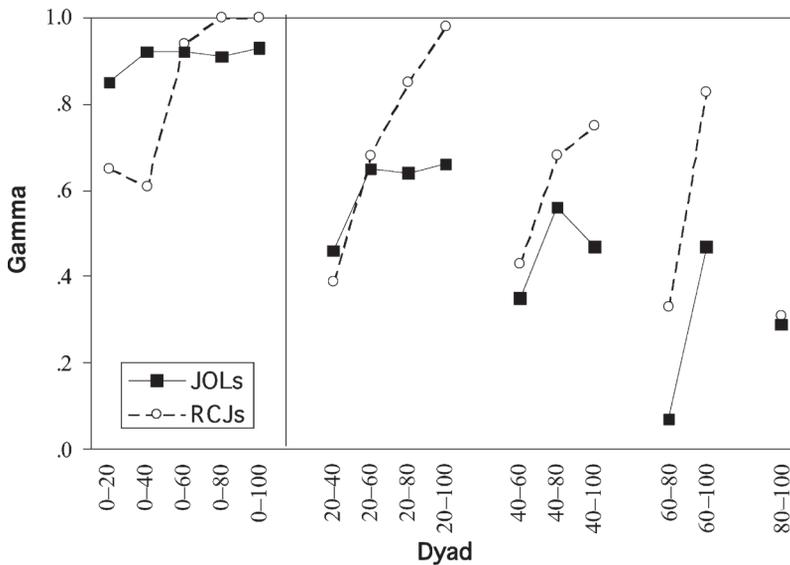


Figure 4. Part-scale gammas between prejudgment recall and magnitude of the judgment for Experiment 1.

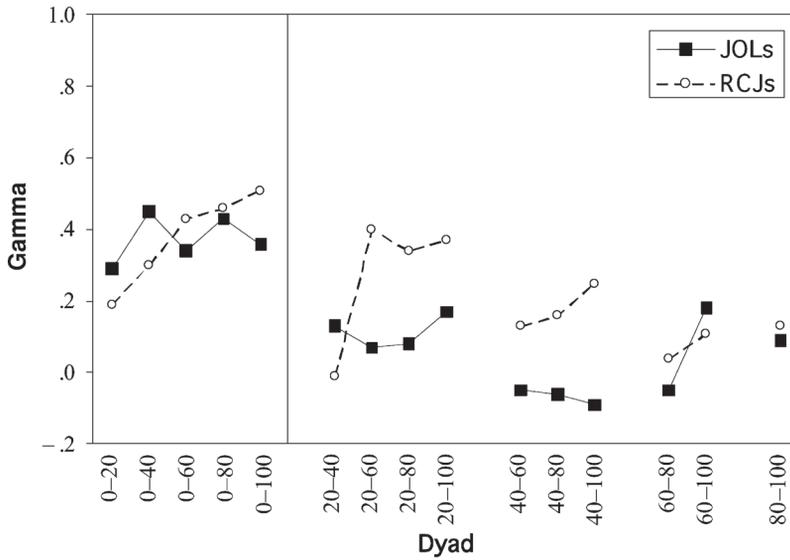


Figure 5. Part-scale gammas between latency of prejudgment recall and magnitude of the judgment for Experiment 1.

of 5 dyads that included judgments of zero. This pattern of results, coupled with those in Figure 4, suggests that RCJs and JOLs rely on prejudgment recall and latency of prejudgment recall to different extents among items receiving nonzero judgments. However, whether this is the case for items that receive judgments of zero is unclear.

Analyses of the overall gamma correlations revealed that RCJs were better predictors than JOLs were of final recall accuracy. However, as can be seen in Figure 6, when scale use is controlled for by using part-scale gamma correlations, the pattern is more complicated. For example, RCJs were more highly correlated with final recall accuracy than were JOLs in only 9 out of the 15 possible

comparisons ($p > .10$) and in only 6 out of 10 of the comparisons ($p > .10$) in which both members of the dyad were nonzero. Nevertheless, these results indicate that RCJs are at least as good as JOLs in predicting future recall, a finding that is somewhat surprising because participants presumably form RCJs without consideration of future recall. The part-scale gammas also indicate the importance of exercising caution in drawing conclusions from overall gammas (as in Figure 3) because different use of the judgment scale can be a modulating factor that should be taken into consideration.

Experiment 1 revealed several important findings. In terms of the basis for JOLs and RCJs, although both judg-

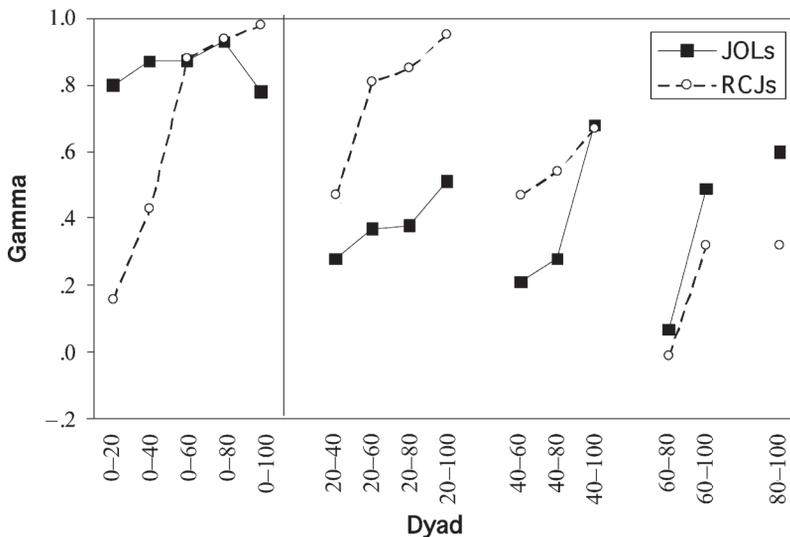


Figure 6. Part-scale gammas between final recall and magnitude of the judgment for Experiment 1.

ments were related to prejudgment recall accuracy, three findings indicate that the two types of judgments were based on different information. One piece of evidence was the finding that the correlations between RCJs and final recall and between JOLs and final recall were differentially affected by the study time manipulation: The relative accuracy of JOLs was affected more than the relative accuracy of RCJs was. This suggests that participants utilized the study time cue differently when forming the two types of judgments. That study time figured more prominently in the formation of JOLs concurs with Busey et al.'s (2000) finding that study time had a relatively larger effect on JOLs than it did on RCJs. Moreover, the fact that JOLs were more poorly correlated with final recall for items studied for 12 sec as opposed to 3 sec suggests that participants either overweighed the importance of longer study time or failed to optimally integrate the cues.⁶

A second finding suggests that JOLs and RCJs were based on different information as well: Among items correctly recalled during prejudgment recall, only RCJs were correlated with latency of prejudgment recall. This suggests that when prejudgment recall was correct, retrieval fluency (speed of prejudgment recall) figured more prominently in the formation of RCJs than it did in the formation of JOLs. This finding—coupled with the finding that there was no difference in the correlation between RCJ and latency of retrieval versus JOLs and latency of retrieval among incorrect prejudgment recall—suggests that RCJs and JOLs were based on different information only when prejudgment recall was successful.

Finally, also consistent with the idea that RCJs and JOLs are based on different information was the finding that the distribution of judgments for RCJs and JOLs differed considerably among items that were correct during prejudgment recall.

Regarding whether RCJs or JOLs were more predictive of future recall, Experiment 1 revealed that RCJs were more highly correlated with final recall than were JOLs. However, this may have been due at least partly to different use of the judgment scale. The possibility that RCJs are at least as predictive of future recall as are JOLs is counterintuitive: RCJs, by their very nature, do not include cues other than retrievability that would be diagnostic of future recall, whereas JOLs, by their nature, should utilize cues that are predictive of future recall.

The findings above give rise to several questions. First, are the findings of Experiment 1 contingent on the participants' making both RCJs and JOLs for the same items? For example, it is possible that participants interpreted our prompt to make both RCJs and JOLs as an indication that we wanted them to use different information for the two kinds of judgments. Such an interpretation would be consistent with Grician conversation norms (Grice, 1975) insofar as participants would not expect to be asked to make both JOLs and RCJs unless the two judgments were alleged to be qualitatively different. Perhaps participants' natural inclination would have been to use the same infor-

mation in the formation of RCJs and JOLs, but our prompt enticed them to produce differences between the two judgments. This could have had the unintended effects of both (1) producing differences between JOLs and RCJs when such differences might not occur in a between-participants design, and (2) inducing participants to utilize the study-time cue differently when making JOLs. For example, it might be the case that differences between how strongly RCJs and JOLs correlated with prejudgment recall and latency of prejudgment recall might disappear if participants did not make both an RCJ and a JOL for each item. Also, a between-participants design would be regarded as a better test of our hypotheses and would eliminate the potential for carryover effects and/or eliminate the potential demand effect inherent in having participants make both judgments. Consequently, Experiment 2 was a replication and extension of Experiment 1. Rather than manipulating judgment type within participants, metacognitive judgment (JOL vs. RCJ) was a between-participants variable. In addition, Experiment 2 included a no-judgment control condition in which the participants did not make a metacognitive judgment. This enabled us to test whether the probability of final recall is affected by making a metacognitive judgment, as had been found previously (e.g., Kelemen & Weaver, 1997; Kimball & Metcalfe, 2003; Spellman & Bjork, 1992) and whether the effect of making JOLs on recall extends to other kinds of metacognitive judgments.

EXPERIMENT 2

Method

Design and Procedure

Experiment 2 was an extension of Experiment 1. The major difference between Experiments 1 and 2 was that judgment type was a between-participants variable, with approximately one third of the participants making JOLs following prejudgment recall and one third making RCJs following prejudgment recall. The remaining one third of the participants did not make a judgment following prejudgment recall (the no-judgment control condition). Thus, the experimental design was a 3 (dummy items: 0, 1, 3) \times 3 (judgment type: JOL, RCJ, no-judgment) \times 2 (study time: 3 sec, 12 sec) mixed factorial with judgment type as a between-participants variable. All other aspects of Experiment 2 were identical to Experiment 1.

Participants

The participants were 184 undergraduate students enrolled in introductory psychology at the University of Maryland; they participated to fulfill course requirements. Participants were randomly assigned to the RCJ condition ($n = 62$), JOL condition ($n = 62$), or no-judgment condition ($n = 60$).

Results and Discussion

Are RCJs and JOLs Based on the Same Information?

Figure 7 shows the mean correlations between metacognitive judgment and prejudgment recall. As in Experiment 1, the correlation between metacognitive judgment and prejudgment recall was quite high, with all gammas above .92. This suggests that prejudgment recall was used greatly when making both RCJs and JOLs. However,

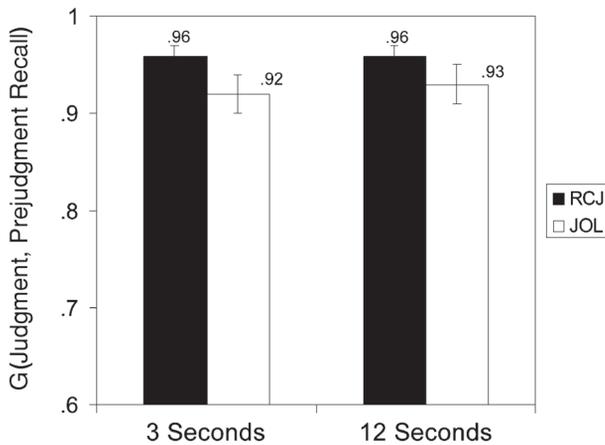


Figure 7. Mean overall gammas between prejudgment recall and magnitude of the judgment for Experiment 2. Error bars represent standard error of the mean.

the pattern of correlations between metacognitive judgment and prejudgment recall again suggested that RCJs and JOLs are not based on the same cues, for there was a significant effect of judgment type on the correlation between prejudgment recall and judgment [$F(1,113) = 3.97, p < .05$], with RCJs yielding higher correlations than did JOLs.

Also as in Experiment 1, there were main effects of judgment type [$F(1,122) = 8.89, p < .01$] and study time [$F(1,122) = 60.29, p < .01$] on metacognitive judgment magnitude (see Table 2), with RCJs greater than JOLs and the 12-sec study condition greater than the 3-sec condition. The judgment type \times study time interaction was nonsignificant [$F(1,122) = 0.37, p > .10$]. Post hoc tests revealed that the effect of study time on judgment magnitude was significant for both JOLs [$t(59) = 6.49, p < .01$] and RCJs [$t(63) = 4.73, p < .01$]. Thus, as in Experiment 1, study time played a role in the formation of both types of metacognitive judgments. However, unlike in Experiment 1, RCJs and JOLs were not differentially affected by the study time manipulation, perhaps because RCJs did not intervene between study and JOL in Experiment 2.

Table 1 shows the correlations between the latency of both correct and incorrect prejudgment recall and judgment. Collapsing across both correct and incorrect prejudgment recall revealed that both RCJs and JOLs were correlated with latency of prejudgment recall: The correlation between metacognitive judgment and latency of prejudgment recall was significantly greater than zero for both RCJs [$M = -.33; t(63) = 11.78, p < .01$] and JOLs [$M = -.29; t(58) = 14.76, p < .01$]. However, unlike in Experiment 1, there was no significant difference between RCJs and JOLs in the magnitude of the correlation [$t(121) = 1.09, p > .10$].

Although JOLs showed a significant correlation with latency of retrieval for correctly recalled items [$M_{JOL} = -.08, t(57) = 3.21, p < .01$], this correlation was attenuated compared to the correlation between JOLs and

latency for incorrectly recalled items [$M_{JOL} = -.34, t(46) = 5.46, p < .01$]. As in Experiment 1, Bonferroni adjusted t tests ($\alpha' = \alpha/4 = .0125$) revealed that the mean correlation between JOL and latency was higher for incorrectly recalled items than for correctly recalled items [$t(45) = 3.80, p < .01$]. Also consistent with Experiment 1 was the finding that RCJs were significantly correlated with latency of retrieval for both correct [$M = -.23, t(61) = 6.16, p < .01$] and incorrect [$M = -.17, t(56) = 3.32, p < .01$] prejudgment recall, with no difference in the magnitude of the correlation between correctly and incorrectly recalled items [$t(54) = 1.14, p > .10$]. Thus, although RCJs were equally well correlated with latency of prejudgment retrieval for both correct and incorrect prejudgment recall, JOLs were more highly correlated with latency of retrieval for the incorrect prejudgment recall than for correct prejudgment recall. Also consistent with Experiment 1 was the finding that RCJs were more highly correlated than JOLs with latency of retrieval for correct prejudgment recall [$t(118) = 3.14, p < .01$], but not for incorrect prejudgment recall [$t(102) = 2.08, p = .04$].

The preceding analyses examining the relationships between latency of retrieval and judgment for correct and incorrect prejudgment recall suggest that latency of prejudgment retrieval plays a more prominent role in RCJs than in JOLs. Note that the magnitude of the correlation between JOLs and latency of prejudgment recall decreased from $M = -.34$ for incorrectly recalled items to $M = -.08$ for correctly recalled items. This finding is interesting because it suggests that the degree to which participants utilized retrieval fluency depended on whether prejudgment recall was correct or incorrect. If prejudgment recall was correct, the participants might have utilized cues in addition to retrieval fluency in forming their JOLs. This would account for the relatively small correlation between the latency of prejudgment recall and JOLs. The relatively high correlation between JOL and the time that participants took to enter a prejudgment recall response among incorrectly recalled items might reflect long search times following guessing (cf. Barnes et al., 1999).

Figure 8 shows the distribution of JOLs (top panel) and RCJs (bottom panel) conditionalized on whether prejudgment recall was correct or incorrect. As in Experiment 1, the distribution of responses among incorrect prejudgment recall was virtually the same for JOLs and RCJs. However, among correct prejudgment recall, participants utilized more midpoints on the scale when forming JOLs than when forming RCJs. Again, this suggests that participants utilized different information when forming the two types of judgments, but that the differences between the two judgments were localized within how they responded to items recalled correctly at prejudgment recall. However, as mentioned above, the differences in response distributions might reflect differences in scale use rather than in process. We address this possibility below.

A final piece of evidence suggesting that RCJs and JOLs were based on different information is the latency for making the judgment. One might speculate that if

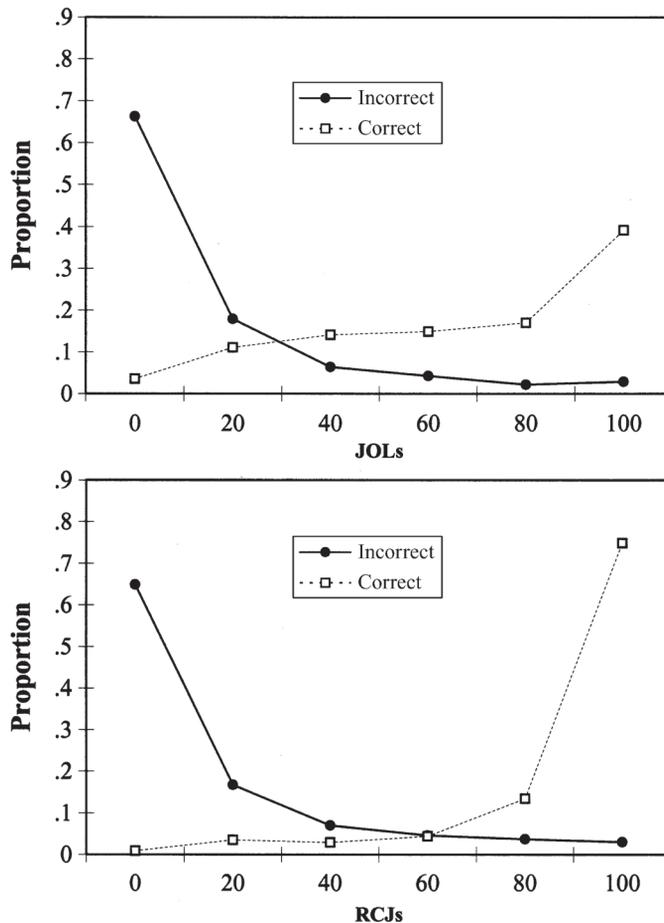


Figure 8. Response distributions for JOLs (top) and RCJs (bottom) for Experiment 2.

RCJs and JOLs were both based on retrievability, then the time to make the two kinds of judgments should be identical. However, if JOLs were based on retrievability plus additional cues, participants should take longer to form JOLs than to form RCJs. Indeed, there was a significant effect of judgment type on the latency of metacognitive judgment [$F(1,122) = 3.85, p = .05$], with participants taking longer to form JOLs ($M = 3.02, SE = .12$) than to form RCJs ($M = 2.63, SE = .10$).⁷ Had the same processes been used to make both types of judgments, the judgment RTs should have been equivalent.

Are JOLs Better Than RCJs at Predicting Future Recall?

Figure 9 shows the correlations between metacognitive judgment and final recall. There were main effects of judgment type [$F(1,110) = 7.80, p < .01$] and study time [$F(1,110) = 3.96, p < .05$], with no significant judgment type \times study time interaction [$F(1,110) = 0.51, p > .10$]. Importantly, consistent with the overall correlations observed in Experiment 1, RCJs were more accurate than JOLs at predicting future recall.

Are the Differences Between RCJs and JOLs Due To Differences in Scale Use?

Figure 10 shows the part-scale gamma correlations between judgment magnitude and prejudgment recall accuracy. There was no consistent pattern showing that RCJs were more highly correlated with prejudgment recall because the correlation between judgment and prejudgment recall was higher for RCJs in only 8 out of the 15 cases ($p > .20$) and for only 7 out of the 10 sets of nonzero dyads. Although the majority of the nonzero dyads favored RCJs, this did not reach significance with the binomial test. Thus, we cannot conclude that RCJs were more heavily influenced by prejudgment recall than were JOLs. Thus, the data displayed in Figure 10 failed to rule out the possibility that differences in scale use account for the differences between how well RCJs and JOLs correlate with prejudgment recall. Although this finding is inconsistent with Experiment 1, note that in Experiment 1 the participants made both JOLs and RCJs for each studied item, whereas in Experiment 2, the participants made only one or the other. As noted in the results and discussion section of Experiment 1, it is possible that the within-participants

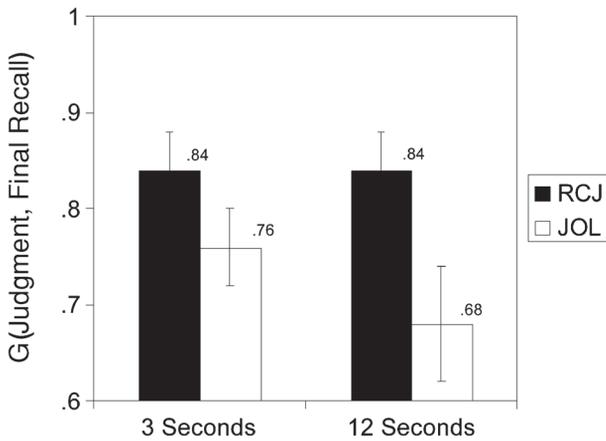


Figure 9. Mean overall gammas between final recall and the magnitude of the judgment for Experiment 2. Error bars represent standard error of the mean.

manipulation of judgment type enticed the participants to treat RCJs and JOLs as more distinct than they might have in a between-participants design. Alternatively, the higher correlations for RCJs (than for JOLs) in Experiment 1 may have been because the RCJs (but not the JOLs) occurred immediately after prejudgment recall in Experiment 1. In Experiment 2, both RCJs and JOLs occurred immediately after prejudgment recall and showed similar effects.

Figure 11 shows the part-scale correlations between latency of prejudgment recall and judgment magnitude for both RCJs and JOLs. Across the 15 comparisons, RCJs were more highly correlated than JOLs with prejudgment recall latency in only 8 comparisons ($p > .20$). However, in 8 of the 10 comparisons among the nonzero dyads, RCJs were more highly correlated than JOLs were with latency of prejudgment recall ($p = .055$). This finding is consistent with those of Experiment 1 and suggests that

RCJs are more closely tied than JOLs to prejudgment recall latency. Contrarily, JOLs were more highly correlated with latency of prejudgment recall for all six dyads ($p = .015$) in which one member received a zero. These opposite results for zero and nonzero dyads indicate that the two judgments are based on different information. One possibility is that JOLs are particularly sensitive to retrieval failure; that is, long memory searches followed by guessing likely is diagnostic of future recall failure. Thus, when one is making a JOL, retrieval failure might be heavily weighted. It seems unlikely that someone would provide a JOL > 0 when prejudgment recall fails. However, when one is making an RCJ following a prejudgment recall guess, participants may still allow for the possibility that the guessing response is correct and assign an RCJ > 0 . If this were the case, JOLs should be unaffected by the interval width (i.e., by whether the correlation is based on 0–20 dyads or 0–100 dyads). In contrast, RCJs should show increasing gamma correlations as the interval width increases (i.e., the gamma for 0–20 pairs should be lower than for 0–40 pairs, which should be lower than for 0–60 pairs, etc.).

Figure 12 shows the part-scale correlations between judgment magnitude and final recall accuracy. Two results are especially important. First, for six out of six dyads in which one member received a zero judgment, JOLs were more predictive of future recall than were RCJs. Second, in 9 out of 10 cases in which both members of the dyads were nonzero, RCJs were more predictive of future recall than were JOLs. These opposite patterns suggest that RCJs were better able than JOLs were to discriminate between which items are more likely to be recalled in the future when both items are given judgments greater than 0. However, JOLs are better able than RCJs to discriminate between items that received a judgment of 0 versus items that received a judgment greater than 0. This pattern of results indicates that the interpretation of the overall

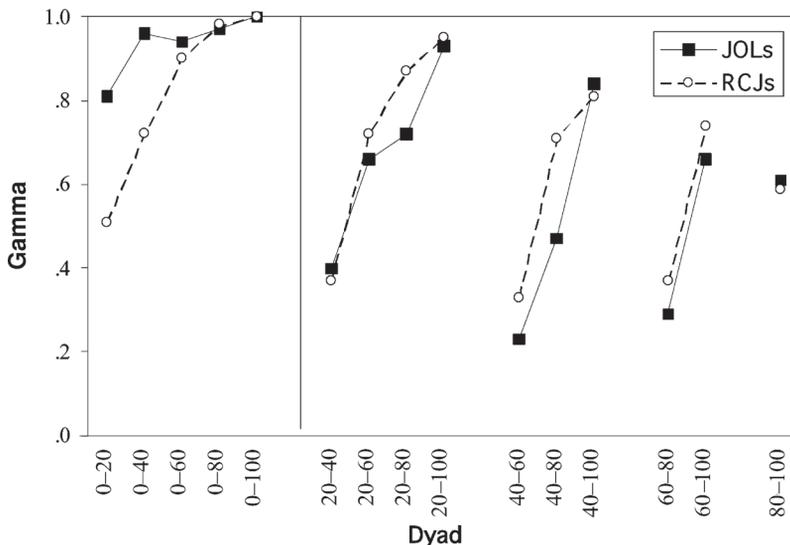


Figure 10. Part-scale gammas between prejudgment recall and magnitude of the judgment for Experiment 2.

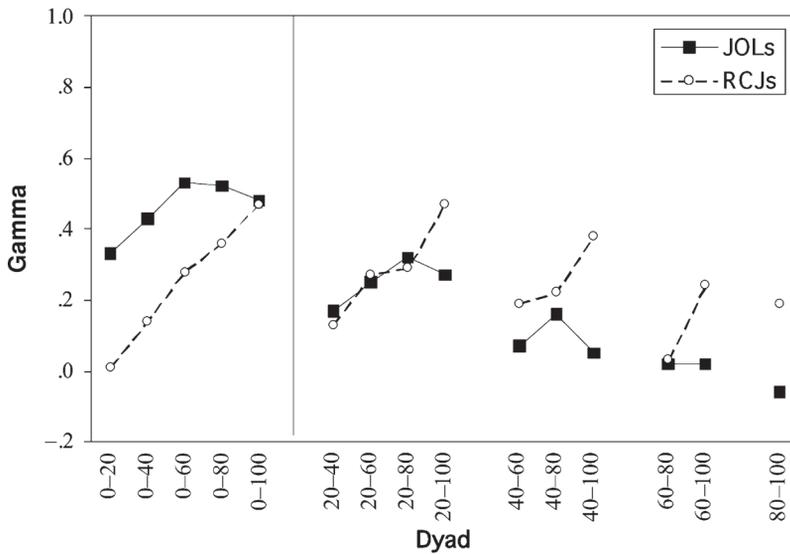


Figure 11. Part-scale gammas between latency of prejudgment recall and magnitude of the judgment for Experiment 2.

gamma correlations in Figure 9 has to be qualified: RCJs are relatively more accurate than JOLs among nonzero dyads, but the opposite is true among dyads containing an item receiving a judgment of 0.

Does Having Made a Metacognitive Judgment Affect Recall Performance?

Spellman and Bjork (1992) and Kelemen and Weaver (1997) theorized that having participants making a JOL in a delayed-JOL task (without prejudgment recall) influenced the state of the item in a way that affected JOL accuracy. Spellman and Bjork argued that the improvement in JOL accuracy was due to an increase in final recall

that resulted from participants' engaging in covert recall when making their JOLs. This hypothesis was supported by Kelemen and Weaver.

Our Experiment 2 tested whether metacognitive judgments improve recall performance above and beyond that which would be expected by successful initial recall. Because all participants were required to engage in both prejudgment recall and final recall, any differences in recall performance among the no-judgment, RCJ, and JOL conditions would not be attributable to the requirement to engage in prejudgment recall. Thus, a third issue addressed by Experiment 2 was whether the process of making a metacognitive judgment affects subsequent recall

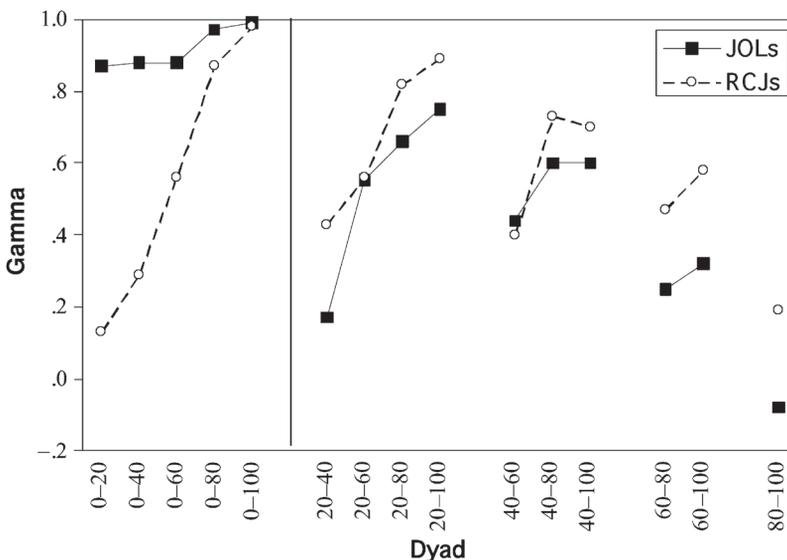


Figure 12. Part-scale gammas between final recall and magnitude of the judgment for Experiment 2.

performance. One interesting hypothesis is that making a metacognitive judgment forces participants to process the to-be-remembered item more thoroughly than they would if no judgment was made. Thus, the act of making a metacognitive judgment may affect how well the item is stored in memory.

We examined the effect of judgment condition (no-judgment, RCJ, JOL) on the probability of correct prejudgment recall, the probability of correct final recall, and the probability of correct final recall given correct prejudgment recall. The mean recall rates are presented in Table 3. In all cases, participants in the JOL condition had higher recall rates than did participants in the no-judgment and RCJ conditions. Statistical analyses using a 3 (judgment type: none, RCJ, JOL) \times 2 (study time: 3 vs. 12 sec) mixed ANOVA revealed that there were main effects of judgment type and study time on the probability of prejudgment recall [judgment type, $F(2,179) = 2.43, p = .09$; study time, $F(1,179) = 87.98, p < .01$]; final recall [judgment type, $F(2,179) = 5.01, p < .01$; study time, $F(1,179) = 108.34, p < .01$] and probability of final recall given correct prejudgment recall [judgment type, $F(2,179) = 6.23, p < .01$; study time, $F(1,179) = 21.09, p < .01$]. None of the interactions approached significance (all $ps > .30$). Dunnett's multiple comparisons using the no-judgment condition as the control revealed that the participants in the JOL condition had reliably higher recall rates than did the participants in the no-judgment condition for probability of prejudgment recall, probability of final recall, and probability of final recall given correct prejudgment recall.⁸ Moreover, this result held in both the 3- and 12-sec study conditions. In contrast, there were no reliable differences between the RCJ and no-judgment control conditions on any of the three recall measures.

The pattern of results illustrated above is consistent with Spellman and Bjork (1992) and Kelemen and Weaver (1997) insofar as JOLs appear to alter what they assess. However, the effect of making a delayed JOL on recall cannot be due solely to the enticement for participants to engage in prejudgment recall because all participants in our study engaged in prejudgment recall. Because the effect of making JOLs showed up at prejudgment recall, it suggests that the requirement to make JOLs actually alters

how well participants are learning the to-be-recalled items at study. Perhaps participants who make JOLs implement a more effective study strategy than participants making RCJs because the JOL task forces them to focus on future retrieval. Interestingly, the finding that there was no effect of making RCJs on any of the recall measures suggests that there is something special about making JOLs that improves learning. That is, the improvement in recall is not merely the result of increased processing that might be associated with requiring participants to make a judgment about the to-be-recalled item: The judgment must be a JOL for this to occur. Note, however, that $p(\text{final recall}|\text{correct prejudgment recall})$ was also higher for the JOL condition than for the no-judgment control condition. This suggests that the boost in recall exhibited by the JOL condition was not entirely due to factors occurring prior to prejudgment recall (e.g., study strategies), otherwise $p(\text{final recall}|\text{correct prejudgment recall})$ should have been equivalent across the conditions. One possible explanation is that the judgment phase is serving as a second "study" opportunity. As presented above, participants in the JOL condition ($M = 3.02$ sec) took significantly longer than participants in the RCJ condition ($M = 2.63$ sec) to make their judgments. Thus, the higher recall rate for the JOL condition might have been due to longer exposure durations in the JOL condition. However, note that participants in the JOL condition were less than half of a second slower than those in the RCJ condition in making their judgments, so that the JOL condition had less than .5 sec more exposure time to the item recalled at prejudgment recall. In contrast, in comparison with those in the no-judgment control condition, participants in the RCJ condition had more than 2.5 sec more exposure time to the item recalled at prejudgment recall and yet did not show a significant increase in $p(\text{final recall}|\text{correct prejudgment recall})$ in comparison with the no-judgment control. Taken together, the results above indicate that whatever processes participants utilize when forming a JOL improves later recall for that item above and beyond the prejudgment recall attempt. In addition, the finding that prejudgment recall was higher in the JOL condition, relative to the no-judgment control, suggests that participants in this condition learned the to-be-remembered items better to begin with.

Table 3
Mean $P(\text{Prejudgment Recall})$, $P(\text{Final Recall})$, and $P(\text{Final Recall}|\text{Correct Prejudgment Recall})$

	Experiment 1				Experiment 2											
	3 sec		12 sec		3 sec				12 sec							
	RCJ & JOL		RCJ & JOL		JOL		RCJ		No Judgment		JOL		RCJ		No Judgment	
	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>
Mean $P(\text{prejudgment recall})$.60	.03	.68	.03	.61	.03	.56	.03	.51	.03	.70	.03	.64	.03	.61	.03
Mean $P(\text{final recall})$.43	.04	.53	.04	.46	.03	.37	.03	.33	.03	.57	.03	.45	.04	.42	.03
Mean $P(\text{final recall} \text{correct prejudgment recall})$.64	.04	.72	.03	.70	.03	.58	.04	.59	.03	.78	.02	.63	.03	.64	.03

GENERAL DISCUSSION

The present experiments addressed two main questions. First, are RCJs and JOLs based on the same underlying information? Second, which are better at predicting future recall? In addition, Experiment 2 examined whether judgments affect the probability of recall.

Are RCJs and JOLs Based on the Same Information?

Prior research on both RCJs and JOLs suggested that both types of judgments were based on some aspect of retrievability. For example, both types of judgments are related to retrieval fluency (Connor et al., 1997; Costermans et al., 1992; Kelley & Lindsay, 1993; Matvey et al., 2001), and both arguably are based on the same processes that are responsible for recognition memory (Sommer et al., 1995). However, other studies have shown that the two types of judgments are at least somewhat distinct. For example, JOLs and RCJs are differentially affected by variables that manipulate study and test conditions (Busey et al., 2000) and appear to be dissociated in Korsakoff (Shimamura & Squire, 1986, 1988) and Alzheimer's (Pappas et al., 1992) patients. Moreover, whereas RCJs might be solely based on aspects of retrievability, participants apparently utilize additional cues, such as encoding fluency (Hertzog, Dunlosky, Robinson, & Kidder, 2003), when forming JOLs.

The present research is consistent with these prior findings. Thus, the answer to the first question is that both JOLs and RCJs seem based on some aspect of retrievability. However, the participants appeared to utilize other information in addition to retrievability when making JOLs. The primary finding supporting the idea that JOLs and RCJs both utilize retrievability comes from the correlations between the judgment magnitude and prejudgment recall. In both experiments, both JOLs and RCJs were highly correlated with prejudgment recall accuracy. However, several findings support the idea that JOLs and RCJs tap somewhat different aspects of memory. For example, the correlation between prejudgment recall latency and the magnitude of the judgment suggests that only RCJs were correlated with latency of retrieval for both correct and incorrect prejudgment recalls. Moreover, there was no difference between correct and incorrect recalls in the magnitude of these correlations for RCJs. In contrast, in both experiments, the correlation between latency of retrieval and judgment magnitude for only items recalled correctly during prejudgment recall was greater for RCJs than for JOLs, and both experiments revealed that the correlation between latency and JOL was higher among incorrectly recalled items. The differences in the patterns of correlations between judgment magnitude and latency of prejudgment recall indicate that participants utilize retrievability differently when making JOLs and RCJs.

A second finding supporting the idea that JOLs and RCJs are based on different information lies in differences in the response distributions. Although there were no noticeable differences between the JOL and RCJ response

distributions for incorrect prejudgment recall, there were differences for correct prejudgment recall. Minimally, this suggests that when prejudgment recall was correct, participants utilized different processes in making JOLs than they did in making RCJs. Although one might suppose that the differences in response distributions were due to different scale use or response mapping, this supposition merely opens up additional questions: First, why do participants utilize the confidence scale differently when making JOLs as opposed to RCJs? Perhaps participants are using different processes to accomplish this. Second, why are differences in scale use limited to correct prejudgment recall? Perhaps participants treat the two judgments differently only when the to-be-judged items are recalled at prejudgment recall.

A third finding suggesting that JOLs and RCJs are based on different information is that, as revealed in Experiment 2, participants take longer to make JOLs than to make RCJs. Had the participants been relying on the same processes to make the two types of judgments, we would have expected no differences in response latency. However, if the participants were basing JOLs on more information than just retrievability, we would expect them to take longer to make JOLs than to make RCJs.

A fourth finding that indicates that participants treat JOLs differently than RCJs was the finding in Experiment 1 that the correlation between JOL magnitude and final recall accuracy was affected by the study time manipulation, whereas there was no effect of study time on the correlation between RCJ and final recall accuracy.

One way to conceptualize the differences between JOLs and RCJs is to view JOLs as based on the same information as are RCJs with variation added to the inference process: $JOLs = RCJs + \text{variation}$. This variation may be *random noise* due to the inference process or possibly the mapping of a covert feeling of confidence onto the scale (Erev, Wallsten, & Budescu, 1994). Alternatively, the variation may be *systematic* and related to the consistency with which participants use or integrate the various cues, or perhaps it is due to the use of poor cues.

The idea that JOLs are RCJs perturbed by random error can account for at least a portion of our data. For example, adding random noise to the RCJs would produce JOL response distributions that were more regressive than the RCJ distributions (see Figures 2 and 8). The response curves for RCJs are highly skewed (most correctly recalled items are given an RCJ of 100), so adding variation to RCJs can only perturb the JOLs downward. Moreover, the addition of random noise would also lead to JOLs that were less well correlated with recall. Although random noise may play a role in explaining our data, it does not account for the full range of observed effects. For example, the random noise explanation cannot account for the tendency of JOLs to be more predictive than RCJs for some subsets of dyads (i.e., dyads involving one zero and one nonzero judgment; see Figure 12). Nevertheless, future research should attempt to separate the degree to which differences between RCJs and JOLs are due to random or systematic variation.

Are JOLs Better Than RCJs at Predicting Future Recall?

The second question addressed by the present research concerned the relative accuracy of RCJs and JOLs in predicting future recall. Although conventional wisdom might suggest that JOLs should be more accurate than RCJs at predicting future recall, our results indicate that RCJs are at least as good as, and sometimes better than, JOLs at predicting future recall. This finding is surprising because one would expect predictions of future recall to be more predictive of future recall than an assessment of past recall would be. However, this finding raises the question of why JOLs are not better than RCJs at predicting final recall.

There are at least three possible reasons JOLs are not more predictive than RCJs at predicting final recall. One (see Figures 2 and 8) is that participants used the confidence scale differently when making JOLs than when making RCJs. However, analyses that controlled for scale use by comparing part-scale gamma correlations showed that RCJs usually were more predictive than JOLs at predicting future recall, especially for dyads that consisted of nonzero items. Combining across experiments, the relative advantage for RCJs in predicting future recall for nonzero dyads held for 15 out of 20 comparisons (6 out of 10 in Experiment 1, and 9 out of 10 in Experiment 2), which is significant by a sign test ($p < .03$). A different pattern occurred on dyads that included zero-confidence items, where for 7 out of 10 comparisons (2 out of 5 in Experiment 1, and 6 out of 6 in Experiment 2), JOLs were more accurate than RCJs in predicting future recall. Because the investigation of judgment type was less clear-cut in Experiment 1 (insofar as participants always made an RCJ followed by a JOL), Experiment 2 offers a more straightforward comparison of the relative predictability of RCJs and JOLs. In Experiment 2, 9 out of 10 comparisons among the nonzero dyads favored RCJs over JOLs, whereas 6 out of 6 comparisons for the zero dyads favored JOLs over RCJs.

That JOLs tend to be better than RCJs for dyads that include zero-confidence items suggests that JOLs are better than RCJs at discriminating between items that will not be recalled (or will be recalled with very low probability) than between items that have a greater chance of being recalled. This conclusion is supported by the relatively high gamma correlations among dyads consisting of one nonzero and one zero judgment. Note that in Figures 6 (Experiment 1) and 12 (Experiment 2), the gamma correlations between JOLs and final recall for the dyads in which one item received a JOL of zero are nearly perfect and unaffected by the interval width. For example, in Experiment 2, the gamma correlations for the 0–20 dyads was the same as the gamma correlation for the 0–40 and 0–60 dyads and not much less than the 0–80 and 0–100 dyads. Thus, participants were quite good at discriminating between items that would not be recalled (and hence received a JOL = 0) and items that had a chance of being recalled (and hence received a JOL > 0). In contrast, the

gamma correlations between RCJs and final recall illustrate that participants' RCJs become increasingly more accurate as the interval width increased from, say, 0–20 to 0–100. Thus, the discrimination of RCJs in predicting which items would be recalled at final recall is affected by interval width, whereas that of JOLs is not.

Why were RCJs better predictors than JOLs on dyads in which neither item received a judgment of zero? One possibility is that whatever cues participants use in addition to prejudgment recall when forming their JOLs are relatively poorer predictors of final recall than is prejudgment recall accuracy. Either the participants do not use an optimal weighting scheme to integrate the various cues when making JOLs, or the use of the additional cues contaminates JOL accuracy. For instance, perhaps participants utilize cues that are less positively correlated (or possibly negatively correlated) with recall performance. The use of these cues would reduce the correlation between judgment and recall.

Why Do JOLs Improve Recall Performance but RCJs Do Not?

One unexpected finding was that having the participants make metacognitive judgments improved recall, but only when the metacognitive judgment was a JOL.⁹ The finding that delayed JOLs lead to better final recall performance than do immediate JOLs has been reported previously (Kelemen & Weaver, 1997; Kimball & Metcalfe, 2003; Spellman & Bjork, 1992). No prior research, though, has investigated whether the improvement in recall is specific to JOLs or whether other kinds of metacognitive judgments might also improve recall. Moreover, most explanations of the improvement in final recall associated with making delayed JOLs (as opposed to immediate JOLs) have focused on the possible role of retrieval practice that is assumed to arise as a result of making a JOL (i.e., JOL formation is assumed to encourage participants to engage in a retrieval process, which after a delay may operate as spaced retrieval practice). However, in our Experiment 2, all the participants were required to engage in prejudgment recall. If the only reason for the improvement in final recall accompanying delayed JOLs was retrieval practice, we would have expected differences in recall to be eliminated when everyone engaged in prejudgment recall. However, we found that JOLs still led to better overall recall than in the no-judgment control condition when both groups were engaging in delayed retrieval. More importantly, our finding that RCJs failed to yield better recall than the no-judgment condition suggests that there is something special about the JOL task that changes how well participants initially learn the to-be-recalled items. Thus, the retrieval practice hypothesis notwithstanding, JOLs appear to alter retrieval for more reasons than mere retrieval practice: In some way they also appear to alter how well participants initially learn the to-be-recalled items.

There are at least two ways in which the increase in recall for participants in the JOL condition might arise.

One possibility is that encoding differences arise as a by-product of the JOL. That is, it is possible that the process of making a JOL encourages participants to process different (or more) aspects of the to-be-recalled word (e.g., the intrinsic properties of the word) than they do when making an RCJ. Alternatively, it is also possible that participants in the JOL condition adopt a more effective study strategy than do participants in the RCJ condition. This might occur, for example, if the JOL task (but not the RCJ task) served as a reminder that recall would be tested again in the future. The possibility that participants in the JOL condition are using a more effective study strategy is consistent with the finding that the JOL recall advantage occurs in prejudgment recall accuracy, which took place before participants made a judgment about the to-be-remembered items. Had the improvement in final recall been due to the extra processing that items receive as a result of making a JOL, we would not expect enhanced prejudgment recall in the JOL condition in comparison with the no-judgment control.

However, the improvement in final recall in the JOL condition relative to the no-judgment control condition cannot be solely attributed to how well the items were studied in the study phase, because this account would not explain why $p(\text{final recall}|\text{correct prejudgment recall})$ also was higher in the JOL condition than in the no-judgment control (see Table 3). That $p(\text{final recall}|\text{correct prejudgment recall})$ was higher in the JOL condition suggests that whatever participants do when making their JOL after prejudgment recall improves their memory for those items above and beyond the benefit provided by having engaged in prejudgment recall. Thus, in addition to the JOL task's leading to better initial learning of the to-be-recalled items (as evidenced by higher prejudgment recall rates relative to the no-judgment control), the process of making a JOL appears to affect the recall performance. Again, it is interesting that only participants in the JOL condition showed higher recall rates than those in the no-judgment control for $p(\text{prejudgment recall})$, $p(\text{final recall})$, and $p(\text{final recall}|\text{correct prejudgment recall})$: There was no benefit of making RCJs on any of these three recall measures. Thus, the benefit of making a metacognitive judgment on recall performance was limited to JOLs.

Implications of Our Findings

The present research has several implications. First, our results support the idea that RCJs and JOLs tap memory differently. However, the finding that both types of judgments rely to some extent on retrievability suggests that theories of RCJs and JOLs might be subsumed under the same general theoretical framework. For example, within the context of Koriat's (1997) cue utilization model, one might think of RCJs and JOLs as being based on different weightings of intrinsic, extrinsic, and mnemonic cues, where retrievability is primarily a mnemonic cue.

One way to conceptualize the differences between JOLs and RCJs is that JOLs are based on the same information as are RCJs with variation added to the inference

process: $\text{JOLs} = \text{RCJs} + \text{variation}$. This variation might be *random noise* due to the inference process or possibly the mapping of a covert feeling of confidence onto the scale (Erev et al., 1994). Alternatively, the variation might be *systematic*, related to the consistency with which participants use or integrate the various cues, or perhaps due to the use of poor cues. For example, JOLs (in contrast to the RCJs) might contain both a systematic component and a random component, such as when two items have been recalled correctly and have received $\text{RCJ} = 100\%$ but differ on subsequent JOLs insofar as the person perturbs each of the confidences down by some amount because of anticipated forgetting. This perturbation could be somewhat noisy (e.g., involving perhaps a random component) but at the same time somewhat systematic (e.g., involving an assessment of the memorial aspects of the retrieval such as ease of retrieval). One way to conceptualize this would be to use what Luce formalized as a semiorder. Applied here, the semiorder (on the scale of difference in confidence between the two items) contains a random component when the items are not distinguishable by one or more just noticeable difference (JND) but probabilistically become distinguishable when the difference exceeds such a threshold, such as when one of the items is judged to be certainly not recallable in the future (i.e., one of the items receives $\text{JOL} = 0$). This would explain why JOLs are not as good as RCJs for much of the scale of JOLs (wherein the random component on JOLs is adding in a $G = 0$ to whatever the G is for the systematically valid differences in JOLs) but why JOLs are better than RCJs when exactly one of the two JOLs for a dyad of items is $\text{JOL} = 0$ (as reported above). This post hoc explanation may be worth investigating in future research.

The second implication concerns the possibility of improving people's metamemory accuracy. Our research suggests a heuristic mechanism for improving the accuracy of JOLs: JOLs should be made by assessing one's confidence in past retrieval, especially when these confidences are greater than 0. That is, prospective judgments should be made exclusively by assessing confidence in retrieval without taking into account prospective cues. As our research and previous research has shown (e.g., Estes et al., 1960), past recall performance is highly predictive of future recall performance. Thus, basing predictions of future recall on past recall should lead to more accurate predictions of future recall. This process, however, may lead to a cost in learning, which leads us to the third implication of our result.

The third implication of our research is that having participants make JOLs is beneficial to the participant in two ways. First, because delayed JOLs are highly correlated with the probability of final recall, participants can use their JOLs to determine which items need to be studied more; for example, items with lower JOLs should receive a higher percentage of study effort than that devoted to high-confidence items (but see Son & Metcalfe, 2000). Thus, one benefit of making JOLs is that they can be used to allocate study time more effectively. The second ben-

efit of making JOLs is that whatever processes people use in making JOLs provides a boost to overall recall performance. Whereas relative judgment accuracy may be somewhat lower for JOLs than for RCJs, this decrease in accuracy can be traded off for increases in the likelihood of correct recall.

REFERENCES

- BARNES, A. E., NELSON, T. O., & DUNLOSKEY, J. (1999). An integrative system of metamemory components involved in retrieval. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII. Cognitive regulation of performance: Interaction of theory and application* (pp. 287-313). Cambridge, MA: MIT Press.
- BAUER, R. H., KYAW, D., & KILBEY, M. M. (1984). Metamemory of alcoholic Korsakoff patients. *Society for Neuroscience Abstracts*, **10**, 318.
- BUSEY, T. A., TUNNICLIFF, J., LOFTUS, G. R., & LOFTUS, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review*, **7**, 26-48.
- CLARK, S. E. (1997). A familiarity-based account of confidence-accuracy inversions in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 232-238.
- CONNOR, L. T., DUNLOSKEY, J., & HERTZOG, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology & Aging*, **12**, 50-71.
- 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.
- DOUGHERTY, M. R. P. (2001). Integration of the ecological and error models of overconfidence using a multiple-trace memory model. *Journal of Experimental Psychology: General*, **130**, 579-599.
- DOUGHERTY, M. R. P., GETTYS, C. F., & OGDEN, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, **106**, 180-209.
- EREV, I., WALLSTEN, T. S., & BUDESCU, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review*, **106**, 180-209.
- ESTES, W. K., HOPKINS, B. L., & CROTHERS, E. J. (1960). All-or-none and conservation effects in the learning and retention of paired associates. *Journal of Experimental Psychology*, **60**, 329-339.
- GIGERENZER, G., & GOLDSTEIN, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, **103**, 650-669.
- GIGERENZER, G., HOFFRAGE, U., & KLEINBÖLTING, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, **98**, 506-528.
- GOLDSTEIN, D. G., & GIGERENZER, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, **109**, 75-90.
- GRICE, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: III. Speech acts* (pp. 41-58). New York: Academic Press.
- HERTZOG, C., DUNLOSKEY, J., ROBINSON, A., & KIDDER, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 22-34.
- HINTZMAN, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, **95**, 528-551.
- KELEMEN, W. L., & WEAVER, C. A. III (1997). Enhanced memory 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.
- KING, J. F., ZECHMEISTER, E. B., & SHAUGHNESSY, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *American Journal of Psychology*, **93**, 329-343.
- KORAIAT, 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.
- LOVELACE, E. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 756-766.
- MATVEY, G., DUNLOSKEY, 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.
- NELSON, T. O. (1996). Consciousness and metacognition. *American Psychologist*, **51**, 102-116.
- NELSON, T. O., & DUNLOSKEY, 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., & DUNLOSKEY, J. (1992). How shall we explain the delayed-judgment-of-learning effect? *Psychological Science*, **3**, 317-318.
- NELSON, T. O., & NARENS, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning & Motivation*, **26**, 125-173.
- NELSON, T. O., NARENS, L., & DUNLOSKEY, J. (2004). A revised methodology for research on metamemory: Prejudgment recall and monitoring (PRAM). *Psychological Methods*, **9**, 53-69.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, **76**(1, Pt. 2), 1-25.
- PAPPAS, B. A., SUNDERLAND, T., WEINGARTNER, H. M., VITIELLO, B., MARTINSON, H., & PUTNAM, K. (1992). Alzheimer's disease and feeling-of-knowing for knowledge and episodic memory. *Journals of Gerontology: Psychological Sciences*, **47**, P159-P164.
- SHIMAMURA, A. P., & SQUIRE, L. R. (1986). Memory and metamemory: A study of the feeling-of-knowing phenomenon in amnesic patients. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 452-460.
- SHIMAMURA, A. P., & SQUIRE, L. R. (1988). Long-term memory in amnesia: Cued recall, recognition memory, and confidence ratings. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 763-770.
- SOMMER, W., HEINZ, A., LEUTHOLD, H., MATT, J., & SCHWEINBERGER, S. R. (1995). Metamemory, distinctiveness, and event-related potentials in recognition memory for faces. *Memory & Cognition*, **23**, 1-11.
- 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.

NOTES

1. We use the term *retrievability* as a general term to refer to the ease with which items can be brought to mind, and treat concepts such as fluency, familiarity, speed of retrieval, ease of processing, and trace strength as specific instances of retrievability. We recognize that each of these later concepts is treated somewhat differently in the literature.
2. One might conceptualize the cues in PMM theory as intrinsic cues in Koriat's (1997) theory in the sense that the environmental cues in PMM are assumed to be associated with (and therefore can be thought of as an intrinsic property of) the to-be-judged stimuli.
3. One possible alternative explanation of the differences between JOLs and RCJs is that the two types of judgments prompt participants to use the confidence scale differently. Thus, differences in gamma correlations might arise because of differential scale use even when the underlying psychological processes are identical. We present additional analyses below, using part-scale gamma correlations, which unconfound scale use

and process. Conclusions based on the part-scale gammas that differ from the overall gammas are noted in the section labeled "Are the Differences Between RCJs and JOLs Due to Differences in Scale Use?"

4. We also examined the latency of the metacognitive judgment. Participants were slightly faster at making their JOLs than at making their RCJs. Analysis of the raw RTs failed to reveal a significant difference between JOLs ($M = 2.75$) and RCJs ($M = 2.54$) [$F(1,61) = 2.69, p > .10$]. Transforming the RTs to $1/RT$ (which is done to reduce the skewness of the RT distribution) revealed that JOLs ($M = .44$) were made significantly more quickly than RCJs ($M = .29$) [$F(1,61) = 18.93, p < .01$]. However, given that the participants always made the RCJ prior to the JOL in Experiment 1, the time it took them to form their JOLs likely was influenced by having made the RCJ. Given the confound that JOLs always followed RCJs in Experiment 1, these RTs are not informative regarding whether JOLs and RCJs are based on the same or different information. Experiment 2 provided a better test of whether judgment latency was affected by judgment type.

5. Separating the part-scale gammas by study time resulted in the mean of part-scale gammas being based on too few participants. For example, if a participant did not provide a confidence of "20," then he or she would not contribute to any part-scale gammas that included "20."

6. An alternative explanation of the lower judgment accuracy found for the 12-sec study condition is that the study time manipulation could

also have changed the response distribution of the confidence judgments. A decrease in gamma correlations could occur, for example, if participants made finer grained distinctions (i.e., used more values on the confidence scale) in the 12-sec condition relative to the 3-sec condition. However, even this interpretation begs the question as to what led participants to make finer grained distinctions. One possibility is that participants utilized the study-time cue differently, which affected their response distribution. Another possibility is that other aspects of the stimuli or study became more salient.

7. Analyses using the inverse of RT ($1/RT$) were consistent with the analyses using the raw RTs.

8. We did separate multiple comparisons for the 3- and 12-sec study conditions for each of the recall measures. Alpha was controlled family-wise at .05 within each set of comparisons and using one-tailed tests.

9. Examination of the recall performance in Experiment 1 indicates that having made a metacognitive judgment in that experiment also led to improved memory performance compared to the no-judgment control condition of Experiment 2 ($p < .05$, Dunnett's test). However, in Experiment 1, participants made both JOLs and RCJs for each item.

(Manuscript received February 24, 2004;
revision accepted for publication October 20, 2004.)