The Journal of General Psychology, 2005, *132*(4), 335–346

Second-Order Judgments About Judgments of Learning

JOHN DUNLOSKY
MICHAEL J. SERRA
GREG MATVEY
Department of Psychology
University of North Carolina at Greensboro

KATHERINE A. RAWSON
Department of Psychology
University of Colorado at Boulder

ABSTRACT. The authors explored the relations between predictions of the likelihood of recalling studied items (called judgments of learning, or JOLs) and second-order judgments (SOJs), in which one rates confidence in the accuracy of each JOL. Each participant studied paired–associate items and made JOLs. A given JOL was either immediate or delayed and was followed immediately by an SOJ. After all items were studied and judged, paired–associate recall occurred. The incorporation of SOJs into this standard method yielded numerous outcomes relevant to theory of metacognitive judgments. SOJs were greater for extreme JOLs (0, 100) than for intermediate JOLs (40, 50). Also, JOL accuracy was greater for delayed than for immediate JOLs, and, reflecting this delayed-JOL effect, SOJs were greater for delayed than for immediate JOLs. These and other outcomes support 2-process hypotheses of how people make JOLs and uncover some pitfalls in interpreting poor judgment accuracy.

Key words: judgments of learning, metacognition, metamemory, second-order judgments

IN THIS ARTICLE, we offer an advance for metacognitive research by introducing second-order judgments (SOJs), which provide a new perspective for exploring the processes and accuracy of metacognitive judgments in general and judgments of learning in particular. *Judgments of learning* (JOLs) are a person's predictions about the likelihood of correctly retrieving recently studied items on an upcoming test. These judgments have become one of the most intensely investigated of the metacognitive judgments (Koriat, 1997; Schwartz, 1994), partly because making them accurately can support the effective regulation of self-paced study (Thiede, 1999). *Second-order judgments* (SOJs) pertain to an individual's confidence in the JOLs themselves and have not yet been investigated in the literature. To demonstrate the potential of SOJs for contributing to the development of theories about metacognitive monitoring, we incorporated them into a standard method used to explore a robust effect from the JOL literature—the *delayed-JOL effect* (Nelson & Dunlosky, 1991). We begin with a brief discussion of this effect and then describe the present approach involving SOJs.

The delayed-JOL effect is demonstrated by having participants study paired associates. For half of the pairs, a JOL is made immediately after each pair has been studied, and for the other half, the JOL is typically delayed for several minutes after study (for an exception, see Kelemen & Weaver, 1997). To obtain a JOL for a pair, such as "dog–spoon," a participant is presented the stimulus alone (dog– ?) and is asked to predict the likelihood of later recalling the response; for example, from 0 (*certain will not recall*) to 100 (*certain will recall*). After this study–judgment phase, each stimulus is presented again, and the participant is asked to recall the corresponding response. The *relative accuracy* of an individual's judgments is computed by correlating his or her JOLs with recall performance across items. Relative accuracy is substantially greater for delayed than for immediate JOLs, an effect that has been independently replicated under numerous conditions and for numerous populations (for a recent review, see Nelson, Narens, & Dunlosky, 2004).

In our present approach, immediately after participants made a given JOL, they made an SOJ by stating their confidence that the JOL was accurate. The incorporation of SOJs into standard methods may yield new answers to various questions pertaining to theory of metacognitive judgments. We focus on two here. First, how should we interpret the accuracy of JOLs, especially when it is poor? Second, what processes contribute to an individual's construction of a JOL? We discuss the relevance of SOJs to each of these questions in turn.

With regard to the accuracy of JOLs, the prevailing interpretation of the low accuracy of immediate JOLs (and other metacognitive judgments; Maki, 1998) is that people often do not know what they know. A provocative possibility is that the same people who demonstrate relatively inaccurate monitoring of memory (as measured by JOLs) may also show quite accurate second-order judgments. To foreshadow, SOJs do demonstrate above-chance accuracy, which itself holds intriguing implications for how to interpret JOL accuracy. We pursue this point again in the Discussion section of this article.

With regard to the construction of a JOL, single-process and two-process hypotheses provide somewhat different answers about how individuals make JOLs. To date, research on how people make JOLs has largely examined whether a particular kind of evidence (or cue) influences the judgments (e.g., Begg, Duft,

John Dunlosky, Katherine A. Rawson, and Michael J. Serra are now at Kent State University. Greg Matvey is now at the University of Maryland.

Address correspondence to John Dunlosky, P.O. Box 5190, Department of Psychology, Kent State University, Kent, OH 44242; jdunlosk@kent.edu (e-mail).

Lalonde, Melnick, & Sanvito, 1989; Benjamin, Bjork, & Schwartz, 1998; Dunlosky & Nelson, 1994; Koriat, 1997; Mazzoni & Nelson, 1995; Schwartz, 1994). By contrast, single- and two-process hypotheses pertain to how such evidence is used in making a JOL. According to a single-process hypothesis, when one makes a JOL, one searches for evidence that indicates that an item will later be remembered. As more evidence is accumulated, the JOL is incremented from 0 (which indicates no evidence has accumulated) to higher values. Thus, the amount of evidence is directly translated into a particular rating on the JOL scale. Assuming that obtaining more evidence increases one's confidence in one's JOLs, a prediction from this single-process hypothesis is that the magnitude of SOJs will increase monotonically with the JOLs.

According to two-process hypotheses, when one makes a JOL, one uses evidence (a) to decide whether a to-be-judged item will or will not be remembered, and then (b) to estimate confidence about the decision. Thus, an initial decision plus a confidence estimate are used in selecting a particular rating on the JOL scale. Consider a specific instantiation of a two-process hypothesis. The *dual-counter hypothesis* was initially described by Nelson and Narens (1990) to account for retrospective confidence judgments, and we adapted it here for JOLs.[1] According to the dual-counter hypothesis, when one makes a JOL, one searches for evidence that an item will later be remembered and for evidence that an item will not later be remembered, and the two kinds of evidence are accumulated into different counters. After the search for evidence is terminated, the individual decides if the item will or will not be recalled (based on which counter has accumulated the most evidence), and the JOL is adjusted based on the amount of evidence. Such adjustments are presumably made from an anchor near the intermediate values of the rating scale (Connor, Dunlosky, & Hertzog, 1997) because "the laboratory setting creates an expectation of an intermediate level of difficulty" (Keren, 1991, p. 255). The evidence is used to adjust the JOL upward from this anchor if the initial decision was that the item would be remembered, and it is adjusted downward if the decision was that the item would not be remembered. Thus, as more evidence is accumulated in support of a decision, JOLs will tend toward the extremes of the rating scale. Because obtaining more evidence for a decision will tend to yield greater confidence in the decision, a prediction from the dual-counter hypothesis—and two-process hypotheses in general—is that the SOJs will form a U-shaped function when plotted against JOLs, with relatively low SOJs for the intermediate JOL ratings and higher SOJs for the extreme JOLs.

In summary, a major goal of this article is to highlight some potential uses of SOJ methodology in metacognitive research. In doing so, we sought to obtain evidence relevant to two issues that have yet to be investigated in this literature: (a) whether people's SOJs demonstrate any accuracy, which would have implications for interpreting JOL accuracy, and (b) whether single- or two-process hypotheses better describe how people make JOLs.

Method

Participants, Design, and Materials

One hundred introductory psychology students from the University of North Carolina at Greensboro participated for course credit. Informed consent was obtained from each. The within-participant variable was immediate versus delayed JOLs. The materials consisted of 66 word pairs, each composed of unrelated nouns. A list was constructed so that half of the items received immediate JOLs, and half received delayed JOLs at least 30 s after a given item had been studied. For details about the construction of the list, see Nelson and Dunlosky (1991). We used Macintosh computers to present all the instructions and items and to collect all responses. The participants were tested individually.

Procedure

The procedure to obtain JOLs was identical to that used by Nelson and Dunlosky (1991), except for the inclusion of SOJs. The participants were instructed to learn each of the 66 word pairs, which were individually presented at 8 s/pair. The order of items was randomized for each participant, and the first 6 items comprised a primacy buffer and were excluded from analyses. After studying a given item (either immediately or after intervening items, depending on the condition), the stimulus was presented alone and the participant made a JOL in which he or she predicted the likelihood of recalling the correct response on the criterion test. The JOLs ranged from 0 (*0% chance of recall*) to 100 (*100% chance of recall*) in increments of 10. Immediately after making a given JOL, an SOJ was collected by presenting the stimulus and having a participant rate the confidence in the accuracy of the JOL made for that item, from 0 (*definitely is not accurate*) to 100 (*definitely is accurate*), in increments of 10. After the participant had studied and judged all the items, the order of the items was randomized, and each stimulus was presented individually for paired-associate recall. The participants were encouraged to respond but were allowed to omit responses.

Results

We first describe recall performance and judgment magnitude and then judgment accuracy. In each section, we begin with an analysis of JOLs to establish replication of standard effects, and then we proceed to an analysis of SOJs. All differences declared reliable had ps < .05.

Recall

For each participant, we computed the percentage of items correctly recalled. A small but reliable difference was found in the mean values of recall for items slated with delayed JOLs (*M* = 39.7, *SEM* = 3.0) and immediate JOLs (*M* = 36.7, *SEM* = 2.0), *t*(99) = 2.20.

Judgment Magnitude

For each participant, we computed the median JOL separately for the two kinds of JOL. Across participants, JOL magnitude was reliably less for delayed JOLs (*M* = 35.9, *SEM* = 3.5) than for immediate JOLs (*M* = 42.2, *SEM* = 2.2), *t*(99) = 2.08.

The median of each individual's SOJs was computed. Across participants, the mean of these values was reliably greater for delayed JOLs (*M* = 77.7, *SEM* = 2.8) than for immediate JOLs (*M* = 61.8, *SEM* = 2.2), *t*(99) = 6.69. One should note that this effect was in the opposite direction of the JOL magnitude, yielding a crossover interaction that indicates the two kinds of metacognitive judgment are not functionally equivalent.

In Figure 1, SOJs are plotted as a function of JOLs. For inferential analysis, the majority of participants did not use every one of the JOL ratings, and hence, each participant had some missing values. Given that the design involved repeated measures (i.e., kind of judgment by JOL ratings), an omnibus ANOVA would have excluded all participants and was not computable. Even so, major outcomes are evident by mere inspection of Figure 1 and were substantiated with post hoc analyses.

First, the nadir of both curves is near the intermediate values of the JOL rating scale, with SOJs increasing as the JOLs shifted toward the extremes of the scale. Consistent with these observations, statistical comparisons revealed that SOJs were reliably higher for JOLs of 0 than 50 and for JOLs of 100 than 50, for both immediate and delayed JOLs (all *t*s > 3.50, *p*s < .002). Second, SOJs were lower for immediate than for delayed JOLs across 10 of the 11 ratings, *p* = .012, which substantiates that SOJs are not merely a direct function of the specific JOL rating provided for an item. If the latter were the case, then the same JOL rating should give rise to the same SOJ regardless of the condition. Finally, the curves are not symmetrical in that people were less confident in their low JOLs than in their high JOLs. This trend is evident both for immediate JOLs (*p* = .035, sign test across individuals comparing SOJs for 0 versus 100 JOLs) and for delayed JOLs (*p* = .009, sign test across individuals).

Judgment Accuracy

For each participant, the relative accuracy of JOLs was operationalized by a Goodman–Kruskal gamma correlation computed between JOLs and recall performance (Nelson, 1984). The mean correlation across individuals was reliably greater for delayed JOLs (*M* = .91, *SEM* = .01) than for immediate JOLs (*M* = .42, *SEM* = .03), *t*(93) = 14.22, which demonstrates the delayed-JOL effect.

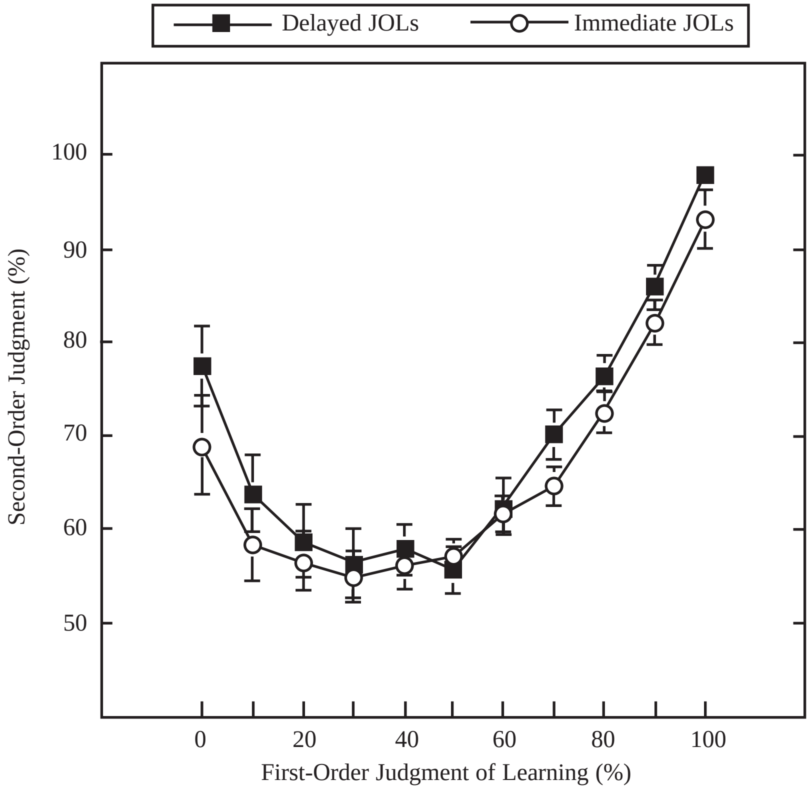

FIGURE 1. Mean of the median second-order judgments (SOJs)plotted (with standard errors) as a function of the first-order judgments of learning (JOLs).

Because the SOJs involve judging the accuracy of JOLs and are themselves accurate, then as the SOJs increase, the relative accuracy of JOLs should also increase. Thus, to assess the accuracy of SOJs, we computed relative accuracy of the JOLs for subsets of items that had received a given level of SOJ or higher. For instance, we selected items that had received an SOJ of 50 or greater and computed JOL accuracy for this subset. We did the same for items that had received an SOJ of 80 or greater, of 90 or greater, and of 100. These values are reported in Table 1. Because SOJs of 100 occurred infrequently for immediate JOLs, the estimated level of JOL accuracy for this subset of items is based on data from only 18 participants and hence was not included in the analysis. For archival purposes, however, we report the corresponding values in Table 1. Because the accuracy of delayed JOLs was consistently on the ceiling, we also limited the analysis to immediate JOLs.

Most important, the accuracy of immediate JOLs increased as a function of SOJs, which is evident in the pattern of both means and medians. Consistent with this trend, sign tests revealed reliably greater levels of accuracy for JOLs corresponding to SOJs of 90 (or greater) than for overall JOL accuracy and for JOLs corresponding to SOJs of 50 or greater, *p*s < .05.

Discussion

By incorporating SOJs into a standard method used in metacognitive research, we obtained some effects of theoretical significance. Namely, the function relating SOJs to JOLs (Figure 1) was U-shaped and asymmetrical, with SOJs increasing with extreme JOLs and being greater for items that had received high JOLs (80 and above) than low JOLs (20 and below). The relative accuracy of immediate JOLs also increased as people's SOJs increased (Table 1). We attempted to replicate these effects in a recent investigation in which 120 college students made immediate and delayed JOLs as well as SOJs. Although this experiment used a slightly different procedure that involved a different presentation rate and fewer items than the present experiment, the pattern of effects reported here was replicated. In this Discussion, we highlight effects that are not only relevant to issues described in the introduction to this article but are also relevant to the relationship between SOJs and JOLs.

TABLE 1. Relative Judgment-of-Learning (JOL) Accuracy as a Function of Second-Order Judgments (SOJs)

	SOJ				
Kind of JOL	0	50	80	90	100
Immediate					
N	96	88	67	41	18
M	.42	.40	.48	.55	.68
Mdn	.43	.42	.56	.84	1.0
SEM	.03	.04	.05	.09	.16
Delayed					
N	96	91	80	74	51
M	.91	.89	.91	.92	.87
Mdn	.96	.96	.98	1.0	1.0
SEM	.01	.02	.02	.02	.06

Note. Each correlation was computed on items that had the specified second-order judgment or higher. JOLs ranged from 0 (*0% chance of recall*) to 100 (*100% chance of recall*) in increments of 10. *SEM* = standard error of the mean; *N* refers to the number of participants contributing a gamma correlation to each cell value.

Theory of Judgments of Learning

Of particular importance to theory of metacognitive judgments, SOJs demonstrated a U-shaped function with JOLs. As we have argued, such a U-shaped function provides more competitive support for two-process hypotheses for JOLs than for a single-process hypothesis. Although we chose to use the dual-counter hypothesis to illustrate predictions, our results would also be consistent with other instantiations of two-process hypotheses. One way in which two-process hypotheses differ from one another is in how much evidence is accrued before making a decision about whether an item will be remembered. For the dual-counter hypothesis described in the Introduction (cf. Nelson & Narens, 1990), the accrual of evidence is completed before making a decision. By contrast, individuals may make a decision about whether an item will be remembered well before all the evidence is in. For instance, familiarity with the judged item may result in an initial and rapid decision about whether an item will or will not be remembered. After this decision is made, the individual would attempt to find further evidence for the particular decision, and hence only one counter (but two processes) would be required. Analogous hypotheses have recently met with success in describing how individuals make other metacognitive judgments (e.g., Koriat & Levy-Sadot, 2001; Vernon & Usher, 2003).

Two-process hypotheses also provide an intuitive explanation for the present dissociation between JOLs and SOJs in which the participants' JOLs were greater for immediate than for delayed JOLs, whereas SOJs were less for immediate than for delayed JOLs. When one makes immediate JOLs, in many circumstances, relevant evidence may not be available (Koriat, 1997), so that the magnitude of the JOLs would tend to be intermediate and the SOJs would be relatively low. By contrast, when a participant is presented with the prompt for a delayed JOL, covertly retrieving a response provides evidence that it may later be recalled (and would result in a shift toward 100), whereas not retrieving a response provides evidence that it will likely not be recalled (and would result in a shift toward 0). In this case, SOJs will be relatively high because the prompt for delayed JOLs will elicit relevant evidence, even when the magnitude of delayed JOLs is quite low, because the accessed evidence indicates that those items will not be recalled.

What Do Second-Order Judgments Measure?

One possible criticism of the SOJ method is that these judgments are based on the same processes that give rise to JOLs. At the extreme, this criticism proclaims that SOJs and JOLs are functionally equivalent. This criticism, however, is countered by the outcomes shown in Figure 1. First, the same JOL rating does not always give rise to the identical SOJ. For instance, SOJs for immediate JOLs were consistently lower than the SOJs for delayed JOLs, even for the same JOL rating. Second, one might argue that the U-shaped function—higher SOJs at extreme JOLs—is not surprising at all, because to make extreme JOLs is equivalent to stating extreme confidence in the recall outcome. However, because the SOJ–JOL curve is asymmetrical, the level of confidence in the recall outcome cannot be observed in JOLs without obtaining SOJs. Put differently, a researcher cannot infer a participant's confidence in his or her JOLs from simply examining the JOLs themselves.

If SOJs and JOLs are not functionally equivalent, then what is their relationship? On the surface, SOJs are judgments about JOLs, so an SOJ can be literally viewed as a *meta-monitoring* judgment. More conceptually, however, our preferred answer to this question is informed by two-process hypotheses in which JOLs are constructed from (a) an initial decision about whether a response will be recalled, and (b) an assessment of confidence in this decision. Our proposal is that these two processes are conflated in JOLs and cannot be definitively separated by analyzing JOLs alone, except by the intuition that the highest and lowest JOLs are expressions of strong confidence that the corresponding items will or will not be remembered. SOJs provide a more specific measure of the latter component—that is, confidence in one's initial decision about whether each item will or will not be remembered. Accordingly, although SOJs and JOLs are not functionally equivalent, they also are not entirely distinct. Instead, SOJs provide a useful method for disentangling the two processes involved in JOLs by isolating the product (i.e., confidence) that arises from the second process.[2] As is evident in Figure 1, using SOJs to inspect this component of JOLs can be quite informative. The asymmetry alone suggests that people are less confident in whether their memory will fail (low JOLs) than whether their memory will succeed (high JOLs). Understanding why individuals show such unbalanced confidence in not knowing versus knowing awaits future research.

Judgment Accuracy

SOJs demonstrated above-chance accuracy. The participants were more confident in their delayed JOLs than in their immediate JOLs, which appropriately reflects the fact that accuracy was greater for delayed than immediate JOLs. SOJs were also accurate at the level of individual items. In particular, as participants' SOJs increased across items, so did the accuracy of immediate JOLs for those items: Mean JOL accuracy was .42 (*Mdn* = .43) when computed across all JOLs but rose to .55 (*Mdn* = .84) when computed for those JOLs in which participants showed extreme confidence. Given the relationship between the SOJs and JOLs shown in Figure 1, these effects may seem relatively obvious, at least in retrospect. Even so, these effects, along with the SOJ–JOL function, hold important implications for interpreting the accuracy of JOLs—implications that are evidently not obvious, given that they have not influenced some researchers' (often cavalier) interpretation of predictive accuracy (the present authors not excepted). We briefly present two implications here—one concerning the interpretation of relative accuracy and the other concerning the interpretation of absolute accuracy.

Relative accuracy pertains to the degree to which a person's JOLs are predictive of criterion performance for one item relative to others, and it is commonly measured by correlating an individual's judgment with his or her own test performance (Nelson, 1984). For interpreting the relative accuracy of immediate JOLs, an analogy is informative. Suppose a friend asks you to predict the likelihood that the Carolina Panthers—a professional football team—will win each of its home games. Assuming you are not an expert, you would have little certainty in your predictions, and if given the chance, you may refuse to even guess. Nevertheless, your friend forces you to make predictions, and at the end of the season, he finds that your clairvoyance for football outcomes is unexceptional. Of course, this news would be quite uninteresting. In the same vein, the participants are forced to make JOLs in a context in which they likely have little experience and one in which they are often unsure of their judgment. This contrived experimental context may partly contribute to the low accuracy of immediate JOLs. Moreover, the use of intermediate JOLs and the overall low confidence in immediate JOLs suggests that participants do know that their predictions are quite poor. Perhaps ironically, even though the relative accuracy of immediate JOLs is often low, participants are demonstrating quite good accuracy about their first-order immediate JOLs when they report having low confidence in them.

SOJs are also relevant to interpreting the absolute accuracy of JOLs. *Absolute accuracy* pertains to the degree to which the percentage predicted on the JOL scale corresponds to the actual percentage of recall. Some pitfalls of interpreting absolute accuracy have been noted (e.g., Keren, 1991; Wallsten, 1996), and the present evidence from SOJs suggests that even further caution should be observed in interpreting measures of absolute accuracy. Consider immediate JOLs, in which the intermediate ratings (e.g., a JOL of 40, 50) are the most often used (e.g., Dunlosky & Nelson, 1994; Weaver & Kelemen, 1997) and also receive the lowest levels of confidence (Figure 1). When calculating a calibration index or a difference score (both are common measures of absolute accuracy), one assumes a JOL of 40 or 50 means that an individual is 40 or 50% confident that an item will be remembered. Put differently, computing absolute accuracy is founded on an (often implicit) assumption that participants' judgments are based on a frequentist view of the JOL rating scale. When an individual makes a prediction of 40%, he or she is presumably stating that "in the long run, I will recall about 40% of the items that have the same characteristics as the item I am currently judging." Instead, such intermediate JOLs mean that an individual is unsure of whether an item will be remembered (cf. Connor et al., 1997), presumably because little evidence was available concerning the uncertain decision. Thus, even though JOLs and recall can be measured on the same percentage scales, the relatively low confidence that people show for some JOL ratings suggests that the scales are not necessarily psychologically comparable, and hence, measures of absolute accuracy may often have limited relevance to theoretical issues pertaining to predictive accuracy.

In summary, SOJs may provide new insight into the processes and accuracy of JOLs. We argued that although JOLs and SOJs are not functionally identical, people's SOJs are informative because they allow investigators to isolate one process of JOLs involved in developing confidence in one's decision about whether a particular response will be recalled. Although SOJs were exploited here in the context of JOLs, they could also be used profitably in exploring other kinds of metacognitive judgment and higher order relations among them.

NOTES

1. The dual-counter hypothesis is described here to illustrate a prediction about the relation between SOJs and JOLs from a specific two-process model. The pattern predicted by this hypothesis, however, can also be accounted for by other two-factor models of JOLs—one of which we describe in the Discussion section. Most important, our approach here was to demonstrate how SOJs could potentially bear on hypotheses relevant to how people make JOLs. And although the data we highlight in the present article may be instantiated by a more general theory of judgments, they also provide converging evidence about the relative merits of single- versus two-process accounts of metacognitive judgments.

2. Another method to disentangle these two components of JOLs would be to first have participants state whether a given item would be remembered. Next, they would rate their confidence in the decision (remember versus not remember) that was made. Although this technique has been used to explore judgment dynamics in other domains, it presupposes a two-process account of standard JOLs, which are typically made on a predicted-likelihood scale. For this reason, we offer SOJs as an alternative method to explore the dynamics of JOLs that does not demand changing the JOL format and potentially the corresponding processes that give rise to JOLs. A fruitful approach for future research may also include contrasting outcomes based on the SOJ method with those based on an initial binary decision with a subsequent confidence judgment.

REFERENCES

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, *28,* 610–632.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127,* 55–68.

Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging, 12,* 50–71.

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

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

Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica, 77,* 217–273.

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

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

Maki, R. H. (1998). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117–144). Mahwah, NJ: Erlbaum.

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

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

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

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–173). San Diego, CA: Academic Press.

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

Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review, 1,* 357–375.

Thiede, K. W. (1999). The importance of monitoring and self-regulation during multitrial learning. *Psychonomic Bulletin & Review, 6,* 662–667.

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

Wallsten, T. S. (1996). An analysis of judgment research analyses. *Organizational Behavior and Human Decision Processes, 65,* 220–226.

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

Manuscript accepted January 2005