Commentary

WHEN PREDICTIONS CREATE REALITY: Judgments of Learning May Alter What They Are Intended to Assess

Barbara A. Spellman and Robert A. Bjork
University of California, Los Angeles

Abstract—*Nelson and Dunlosky* (Psychological Science, *July 1991) reported that subjects making judgments of learning (JOLs) can be extremely accurate at predicting subsequent recall performance on a paired-associate task when the JOL task is delayed for a short while after study. They argued that this result is surprising given the results of earlier research, as well as their own current experiment, indicating that JOLs are quite inaccurate when made immediately after study. We note that the delayed-JOL procedure used by Nelson and Dunlosky invited covert recall practice (which was reported by their subjects). Retrieval practice is a well-known determinant of subsequent recall. Accordingly, Nelson and Dunlosky's findings can be explained by the simple assumption that people base delayed JOLs on an assessment of retrieval success, which in turn influences their retrieval success on the subsequent recall test.*

Nelson and Dunlosky (1991) reported that they have identified a particular circumstance in which subjects' judgments of learning (JOLs) are extremely accurate—a finding that they view as surprising given the generally inaccurate character of such judgments as reported in the literature. We argue here, however, that the specific JOL task used by Nelson and Dunlosky invited subjects to employ a strategy that, in effect, turned their predictions of future performance into a self-fulfilling prophecy.

Subjects in Nelson and Dunlosky's experiment were told to study word pairs so that later they could recall the second word (response) when cued with the first word (stimulus). In addition to these study trials were trials in which subjects were asked to judge their state of learning with respect to particular pairs presented earlier. On such JOL trials, subjects were shown the stimulus alone and were asked: "How confident are you that in about ten minutes from now you will be able to recall the second word of the item when prompted with the first?" (p. 268). For half of the pairs, the JOL was made immediately after the study trial for that pair; for the other half of the pairs, the JOL was delayed by at least 10 intervening mixed study and JOL trials. Nelson and Dunlosky found that JOL accuracy increased dramatically with delay; in fact, on delayed JOLs, subjects were almost perfect at predicting their own subsequent recall probability. The researchers expressed considerable amazement at the delayed-JOL performance: "*Every* subject's accuracy on delayed JOL was greater than the mean of those same subjects' accuracy on immediate JOL!" (p. 269). They also noted that their finding is in opposition to the "nearly universal finding in the literature" (p. 267) that JOLs are quite inaccurate.

Given their procedure, however, Nelson and Dunlosky should not have been so surprised: Their results are consistent with the body of previous research on the effects of retrieval practice, generation, and spacing. It has long been known that retrieval practice, based on unreinforced recall attempts, enhances long-term retention (e.g., Allen, Mahler, & Estes, 1969). One strategy for making a delayed JOL is to use the presented stimulus as a cue to try to recall the response item, and to base the JOL on whether recall is successful. Given the known effect of such retrieval practice, successful covert recall during the JOL task will in turn increase the likelihood that the subject will successfully recall that item on the later overt recall test. (Conversely, unsuccessful retrieval practice strongly predicts failure on the later recall test.) Thus, if delayed JOLs are based on the ability to recall the response, and final recall is also based on the ability to recall the response, it follows that delayed JOLs and final recall will necessarily be correlated, just as Nelson and Dunlosky found. Learning due to JOLs could have been assessed by examining the final recall performance on items that did not receive JOLs, but no such control was included in the experimental design.

In accord with the foregoing argument, Nelson and Dunlosky reported that their subjects did in fact use the attempted-recall strategy to make delayed JOLs: "In end-of-session interviews . . . , 19 of 20 subjects reported having attempted to recall the correct response to the stimulus cue during delayed JOL . . ." (p. 270, note 4). Whether subjects used their covert recall success as a basis for predicting final recall performance was not reported. However, using a procedure in which some subjects had to make overt responses to a stimulus cue (i.e., they had test trials) and then a JOL (given both stimulus and response), King, Zechmeister, and Shaughnessy (1980) found that subjects who received test trials were better at JOLs than subjects who did not receive test trials. King et al. concluded that tested subjects were

Address correspondence to Barbara A. Spellman, Department of Psychology, University of California, Los Angeles, 405 Hilgard Ave., Los Angeles, CA 90024-1563; e-mail: bobbie@cognet.ucla.edu.

more accurate because they used their performance on the test trials as a basis for their JOLs.

An example might help make our argument more concrete. Assume a subject has learned pairs A and B to some intermediate level, with A learned slightly better than B. Given the probabilistic nature of recall processes, it is quite possible that at the time of the JOL, the subject will succeed in covertly recalling the response to B but fail in the attempt to covertly recall the response to A. A subject using the strategy outlined above will then falsely conclude that B has been learned better than A and will make JOLs reflecting that false belief. The subject will appear, however, to have assessed his or her knowledge state correctly because—due to the successful covert retrieval of B during the JOL task—the subject will now be more likely to recall B on the final test. Thus, although the subject will have been incorrect at assessing the initial degree of learning, the JOLs will seem to be accurate.

One question remains: Even if the JOL and final recall are not independent, why were subjects so much more accurate in the delayed condition than in the immediate condition? It is a well-known finding that the long-term effect of retrieval practice on final recall depends on the delay between the original presentation of the material and the retrieval practice. One study, for example, found that as the delay between presentation and retrieval practice increased, performance on the practice task decreased (indicating normal forgetting), but performance on the final recall task improved (Whitten & Bjork, 1977). This unexpected result was possible because as delay increased, the conditional probability of an item being correctly recalled on the final test—given that it was correctly recalled on the practice—increased at a rate that more than offset the decreasing performance on the practice task. Therefore, at a longer delay, success on a practice task becomes both a better predictor of success on a final recall test and a more potent cause of success on the recall task. Similar effects have been observed using an episodic generation task. Jacoby (1978) found that a generation task occurring immediately after an initial presentation of paired associates did not produce better long-term recall than an immediate second presentation; in contrast, items in a delayed-generation condition were better recalled than both immediate-generation items and items receiving a delayed second presentation.

Nelson and Dunlosky (1991) suggested that subjects may have been more accurate for delayed-JOL items than for immediate-JOL items because they used the attempted-recall strategy more often for the former item type. Based on evidence concerning the relationship of spacing to retrieval practice and generation effects, however, it is likely that whether or not subjects used the attempted-recall strategy in the immediate condition was irrelevant to their JOL accuracy. The expected degree of dependence between JOLs and final recall is a function of timing—only after a delay will covert retrieval practice evoked by the JOL task have a strong influence on final recall.

In our view, Nelson and Dunlosky's findings reflect a psychological analog of the Heisenberg Uncertainty Principle: Any effort to take a reading of a subject's current state of knowledge may alter that state of knowledge. In this specific instance, when subjects measure their own degree of learning after a delay by making covert recall attempts, they alter their degree of learning. The delayed JOL, in effect, creates its own reality; in such happy circumstances, the accuracy of the measurement is assured.

Acknowledgments—We would like to thank Todd Gross and Keith Holyoak for helpful comments on an earlier draft.

REFERENCES

Allen, G.A., Mahler, W.A., & Estes, W.K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior, 8*, 463–470.

Jacoby, L.J. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17*, 649–667.

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.

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.

Whitten, W.B., & Bjork, R.A. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning and Verbal Behavior, 16*, 465–478.