Journal of Experimental Psychology:
Learning, Memory, and Cognition
2001, Vol. 27, No. 4, 907–912

0278-7393/01/$5.00 DOI: 10.1037//0278-7393.27.4.907

Metacognition in Motor Learning

Dominic A. Simon
McMaster University

Robert A. Bjork
University of California, Los Angeles

Research on judgments of verbal learning has demonstrated that participants' judgments are unreliable and often overconfident. The authors studied judgments of perceptual–motor learning. Participants learned 3 keystroke patterns on the number pad of a computer, each requiring that a different sequence of keys be struck in a different total movement time. Practice trials on each pattern were either blocked or randomly interleaved with trials on the other patterns, and each participant was asked, periodically, to predict his or her performance on a 24-hr test. Consistent with earlier findings, blocked practice enhanced acquisition but harmed retention. Participants, though, predicted better performance given blocked practice. These results augment research on judgments of verbal learning and suggest that humans, at their peril, interpret current ease of access to a perceptual–motor skill as a valid index of learning.

Research into the subject of metacognition, or what we know about what we know, has received increasing attention over recent years (see, e.g., Metcalfe & Shimamura, 1994; Nelson, 1992). Research focused on people's abilities to predict their own future recall or recognition of studied material has demonstrated that such predictions are frequently less than accurate, sometimes very wrong, and often overconfident. Such studies, however, have tended to focus on verbal–conceptual learning, rather than on the acquisition of motor skills. The current research concerns whether metacognitive predictions of one's later ability to perform a to-be-learned motor skill are similarly open to error, or, given differences between motor skills and verbal learning, protected from such errors.

To address this question, we chose a manipulation of the conditions of skill acquisition, blocked versus random practice, which has been shown to result in a dissociation between performance during training and performance at a delay (e.g., Shea & Morgan, 1979). Blocked practice, in which all trials of a particular to-be-learned pattern are completed before practice is begun on another pattern, is compared with random practice, in which trials of the to-be-learned tasks are interleaved in a semirandom fashion. Blocked practice usually leads to smaller error during practice than does random practice, but on retention tests the opposite pattern is generally obtained.

To the extent, then, that acquisition performance during blocked and random practice is an imperfect index of future performance, the blocked-versus-random manipulation represents a challenging and potentially diagnostic paradigm for the purposes of examining learners' metacognitive assessments of the degree to which they have, or have not, mastered a cognitive–motor skill. By comparing predictions made during practice with actual retention performance, it should be possible not only to measure the relative accuracy of learners' metacognitive judgments about their state of skill acquisition, but also, via the patterns of over- and underconfidence, to gain some idea as to the objective and subjective indexes of performance that are informing or misinforming such judgments (see Bjork, 1999).

Subjective Assessments of Learning

Often people's metacognitive assessments match well with their actual performances (e.g., Hart, 1965), suggesting that they are good at judging their own level of knowledge. However, there are a number of situations in which people's metacognitive evaluations have been demonstrated to be not only unreliable, but also negatively correlated with their own later performance.

Baddeley and Longman (1978), in training postal workers to type, implemented several practice groups and varied the relative spacing of typing practice. They found, consistent with much laboratory research, that spacing the training sessions in time enhanced learning. It is interesting though that when the postal workers were asked to rate their training experiences, the trainees in the most spaced condition—objectively the best condition for learning—reported the least satisfaction with their training schedule. Indeed, some of those workers indicated that they would refuse to take part if asked to undergo further training on that same schedule. Moreover, those with the most massed practice schedule—objectively the worst for learning—reported the greatest satisfaction, and the majority indicated that, given the choice, they would choose that same schedule again. In this situation, trainees' subjective impressions clearly did not accord with the more objective performance measures observed between different conditions of practice.

Subjective assessments about the relative ease of performance during learning can yield false impressions of true competence, or memorability (see, e.g., Benjamin, Bjork, & Schwartz, 1998; Jacoby & Kelley, 1987).

Dominic A. Simon, Department of Kinesiology, McMaster University, Ontario, Canada; Robert A. Bjork, Department of Psychology, University of California, Los Angeles.

We thank Derek McAuley for helping to gather the data and Aaron Benjamin for comments on a preliminary draft.

Correspondence concerning this article should be addressed to Dominic A. Simon, Department of Kinesiology, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada. Electronic mail may be sent to simond@mcmail.cis.mcmaster.ca.

A particularly important metacognitive index that has been used quite widely is the judgment of learning (JOL). Typically, learners make an assessment of how well they have learned some material that they are studying by predicting their own future ability to recall that information at a later time. Someone studying foreign language vocabulary items, for example, might be asked to predict how likely they would be to recall the native equivalent of a given foreign word when presented with that word after some specified delay. It is worth noting that judgments of learning are an extremely common mental operation in real-world contexts. In any learning context where there is a self-pacing of study or practice, for example, we allot our time and efforts on the basis of our subjective impressions of how well we have mastered different aspects of the studied material.

Research on verbal learning has demonstrated that JOLs are often poorly correlated with actual learning (e.g., Vesonder & Voss, 1985), although the accuracy of such judgments has been shown to vary with when the JOL is made and with the nature of the cue used to prompt the JOL (e.g., Dunlosky & Nelson, 1992, 1994; Nelson & Dunlosky, 1991). In their Experiment 2, Dunlosky and Nelson (1994), found that the JOLs for studied word pairs varied as a function of both distribution of practice (massed vs. distributed) and timing of the JOLs (immediate vs. delayed). Notably, immediate JOLs were poorly related to actual recall performance: Confidence did not differ for massed and distributed items. In contrast, delayed JOLs more accurately reflected relative test performance: For distributed study items, cued recall was better than for massed items. In the immediate case, the ease of accessibility of the massed study conditions apparently served to foster overconfidence in the degree of learning. Because standard JOL predictions are often imperfectly related to actual performance, those conditions that lead people to make high-confidence JOLs may foster what Jacoby, Bjork, and Kelley (1994) referred to as "illusions of competence."

Metacognition in Skilled Behavior

In the domain of skilled behavior, relatively little research has been focused on the metacognitive processes that govern learners' self-assessments (although Harvey, 1994, provided a good review of much of what has been done). Cohen, Dearnley, and Hansel (1956) studied confidence levels in bus drivers, asking them to judge whether they would be able to successfully negotiate a bus through a narrow gap. The width of the gap was manipulated and the judgments, and then performances, of experienced and inexperienced drivers were recorded. The results indicated that the drivers were generally overconfident and that even though experienced drivers were, understandably, more competent in their performance, they were no better at assessing their chances of success than were the less experienced drivers.

Other findings also suggest that self-judgments of competence can be misleading. Marteau, Wynne, Kaye, and Evans (1990), for example, studied junior doctors and found that self-confidence in their ability to resuscitate a patient was positively related to the number of cardiac arrests they had attended, but their actual resuscitation skills did not show a commensurate improvement with experience. Similarly, Harvey, Garwood, and Palencia (1987) had singers listen to a musical interval and then sing it. The participants then rated their confidence in having successfully reproduced the interval. Confidence was positively correlated with amount of practice at the task, but confidence did not correlate with performance.

West and Stanovich (1997) actually compared overconfidence levels on a motor task with that for a knowledge task. They found limited evidence for a positive correlation between overconfidence on the two tasks. One of their research goals was to test the notion that overconfidence on some tasks emerges because of artifacts in the response scale. By using such different tasks, and thus different response scales, they thought it highly unlikely that correlations would occur between overconfidence levels due simply to such artifacts. On the basis of their results, they suggest that overconfidence in an individual may generalize across task domains. On the motor task itself, in which participants had to slide a coin along a smooth table-top surface to stop as close as possible to a target zone, a few introductory trials were followed by two blocks of performance. Both blocks were preceded by predictions of performance, and although the degree of overconfidence was attenuated on the second block, the participants showed significant overconfidence.

Goals of the Present Research

In the Cohen et al. (1956) and West and Stanovich (1997) studies, the participants made a judgment about a task that they were about to attempt in the immediate future. An aspect of such judgments that has received little attention within the learning of skills is how well learners are able to assess their current state of learning. That is, how well can they predict, during practice, their performance at some delay after practice ends? The accuracy of such judgments is clearly important because it can influence both the extent of practice itself, where amount of practice is under the control of the learner, as well as the learner's expectations of his or her own ability to perform at some later time. We know that practice performance is often a poor indicator of learning as assessed by delayed tests of retention and transfer (see reviews by Chamberlin & Lee, 1994; Christina & Bjork, 1991; Ghodsian, Bjork, & Benjamin, 1996; Schmidt & Bjork, 1992). Learners in many study situations, such as in contextual-interference manipulations, may therefore be particularly prone to making inaccurate estimates of their own later competence if they use immediate performance as an indicator of learning.

In line with a long history of experiments on contextual interference effects (see Magill & Hall, 1990, for a review), we expected that random practice would lead to inferior acquisition performance, but superior retention performance, relative to the blocked practice condition. Judgments of learning made during practice were expected to reflect immediate performance. Because blocked and random practice were expected to have opposite effects in practice and retention, we therefore wanted to examine whether participants would be fooled by their own current performance, so to speak, or would—instead—be able to discount the temporary benefits of blocked practice in predicting their own posttraining retention performance.

Method

Participants

The participants were 48 undergraduate students (28 women, 20 men: *M* age = 19.7 years) from the University of California, Los Angeles, who took part for credit in a psychology course. These participants were randomly assigned to one of two equal-sized groups for the acquisition phase of the experiment: blocked practice or random practice.

Task and Apparatus

The main experimental procedures were administered using a desktop computer. A customized software package was used to administer the experimental trials, collect participants' responses, and record data for later analysis. The basic task involved learning to press particular five-key sequences on the number pad of the computer keyboard so that the sequences were executed as close as possible to predetermined goal movement times (MTs). This task has been used previously by Lee, Wishart, Cunningham, and Carnahan (1997), as well as by Levy and Bjork (2000). Each of three to-be-learned sequences had a distinct set of keys to be pressed, was presented in a different color on the computer screen, and had a unique goal MT: 900, 1,200, and 1,500 ms. These movement times were chosen to be well within the response-time range of the average person, given some practice.

Acquisition Phase

Instructions. Participants were instructed that the experimental task involved pressing specified sequences of keys on the number pad of the computer keyboard as close as possible to a specified goal MT. They were asked to make all key presses with the index finger of the dominant hand—keyboard placement was adjusted according to handedness. They were also instructed not to press any keys until the pattern to be performed had appeared on the screen and a beep had sounded. The form of feedback was explained, and a printout of the feedback from a sample trial was presented to them, which they then had to interpret verbally (to the satisfaction of the experimenter). The nature of the prediction task was also explained, and an example of the prompt that would appear for each prediction was displayed. The instructions stressed the fact that the participants were to estimate how close, in milliseconds, they thought their performance would be to the relevant goal MT when they were given a retention test the next day. Any questions that the participants raised were answered before testing began.

Experimental tasks. In order to give the participants an opportunity to get used to the task, practice began with five successful trials of a pattern different from the three that were to be learned in the main part of the experiment. Here, and throughout the experiment, for a trial to be deemed successful it was necessary only that the correct keys be pressed in the correct order. Once these five trials of the sample task were completed, the participants were asked if they had any further questions and then they proceeded with the main acquisition procedures.

These three to-be-learned patterns, as well as their goal MTs and presentation colors, are presented in Figure 1. On each learning trial, the pattern to be performed was presented on the computer screen. The first key of the sequence was marked with an "S" for start, and a line indicated the order in which the remaining keys were to be pressed. As shown in Figure 1, the green pattern required striking the keys *9-5-1-2-3* in a total of 900 ms; the red pattern required striking the keys *3-6-5-8-4* in 1,200 ms; the white pattern required striking *4-2-5-8-9* in 1,500 ms.

Before each of the first three trials of each pattern, the goal MT was presented on the screen for 4 s. The to-be-performed pattern then appeared on the screen, and the participants were free to respond, as soon as they were ready, by pressing the appropriate keys. The pattern remained visible on the screen until 2 s after the response was completed. After this 2-s delay, feedback was provided about the trial. This feedback, or knowledge of results, consisted of three pieces of information: whether or not the keys pressed had been correct; the actual MT for that trial in milliseconds; and the amount, in milliseconds, by which that MT was faster or slower than the goal MT. The feedback remained on the screen for 5 s, and then the screen went blank for 5 s before the next trial began.

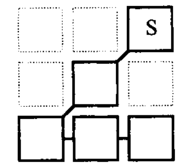
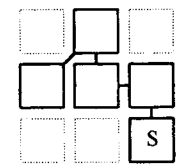
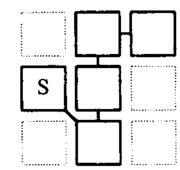
Keys	Color	Goal MT	Keyboard Pattern
9-5-1-2-3	Green	900 ms	
3-6-5-8-4	Red	1200 ms	
4-2-5-8-9	White	1500 ms	

Figure 1. To-be-learned keystroke sequences. The key labeled "S" was the first in each sequence, and participants had to follow the line to press the subsequent keys in the correct order. MT = movement time.

The acquisition phase continued, for both blocked- (e.g., 900 ms, 900 ms, 900 ms . . . , 1,200 ms, 1,200 ms, 1,200 ms . . . , 1,500 ms, 1,500 ms, 1,500 ms . . .) and random-practice conditions (e.g., 900 ms, 1,200 ms, 900 ms, 1,500 ms, 1,200 ms . . .), until each participant had performed 30 successful (correct key strokes) trials of each of the three patterns. If the wrong keys were pressed, this fact was indicated in the feedback and that trial was repeated at the end of the current block. Order of pattern performance was counterbalanced across subjects in both the blocked-practice and random-practice conditions.

Subjective JOLs. After every fifth successful trial on each of the three patterns, a question appeared on the computer screen asking the participants to make a judgment about how well they thought that they had learned the pattern that they had just executed. Specifically, they were asked to predict their own performance on the test scheduled for the next day, assuming that they would get no more practice trials on the pattern in question. They were asked to estimate how close they thought they would be able to come to the target time on that test, then enter that number of milliseconds and press the enter key. Smaller numbers on this subjective judgment were thus indicative of greater confidence on the forthcoming test. As part of the prompt to make a judgment, participants were reminded that on the test to be administered in 24 hr they would see only a diagram of the pattern to be performed, not its goal MT.

Retention Phase

On the day after the acquisition phase of the experiment, the participants returned to the experimental room for tests of retention. After a brief paper-and-pencil test of their recall of the patterns and associated MTs from the previous day, they made a prediction as to their expected performance on each of the patterns, in the same form as the predictions they had made throughout the learning phase.

Performance tests. Three trials of each pattern were presented to the participants under a blocked order and 3 under a random order, making a total of 18 retention test trials. The order of blocked- and random-test-trial presentations was counterbalanced across participants. Both blocked and random orders were used in order to detect any specificity of practice effects (cf. Shea & Morgan, 1979). On these retention-test trials, the diagram of the pattern was presented on the screen—explicit memory of the patterns was therefore not strictly necessary for performance on this test—and the participants were to respond as soon as they were ready. They were asked to execute a given keystroke pattern in as close as possible to what they remembered the appropriate target time to be. Two seconds after the pattern was executed, a message was displayed on the computer screen informing them whether or not the keys pressed were correct. If the wrong keys were pressed, that trial was repeated in a manner concordant with the current test schedule. No information about MTs was provided, either about the goal MTs or as feedback about actual performance. The order of presentation of the patterns within the blocked and random test trials for each participant matched the ordering that had occurred in the acquisition phase. At the end of this retention session, the participants were thanked, debriefed, and excused.[1]

Results

The first section below reports the participants' predictions and actual performance during the acquisition phase; the second section reports their predictions and performance at the time of the final test.

Acquisition Phase

In the analyses of acquisition performance, the trials in the blocked practice conditions were reordered so as to match those of the random conditions. That is, for both types of schedule, the first five correct trials on each pattern were combined to form the first acquisition block, the second five correct trials on each pattern were combined to form the second acquisition block, and so on for all of the acquisition data.

Performance measures. The primary performance measure of interest is absolute constant error (|CE|), a measure reflecting the average deviation, in either direction, of a participant's MTs from the target MTs.[2] To facilitate comparisons between the different practice patterns, this measure was converted to percent absolute constant error (%|CE|) by dividing |CE| by the corresponding target movement time and multiplying by 100.

The mean %|CE| scores for the two conditions across acquisition trial blocks are presented in the top panel of Figure 2. Those scores were subjected to a 2 × 3 × 6 (Practice Schedule × Pattern × Trial Block) mixed univariate analysis of variance (ANOVA), where Pattern and Trial Block are within-subjects factors.

The pattern that seems apparent in the top panel of Figure 2—that is, that blocked practice yielded better performance during acquisition than did random practice, but that the difference decreased across blocks—is supported by the statistical analysis. The main effect of practice schedule is significant (M = 5.68% and M = 8.85% for blocked and random practice, respectively), $F(1, 46) = 10.41$, $MSE = 209.203$, $p < .05$, as is—not surprisingly—the main effect of block, $F(5, 230) = 19.90$, $MSE = 41.65$, $p < .01$. The Schedule × Trial Block interaction is also significant, $F(5, 230) = 4.208$, $MSE = 41.654$, $p < .01$. Follow-up analyses indicated that whereas the blocked condition yielded lower error scores than the random condition for all blocks of practice, the difference was only significant for the first three blocks. That is,

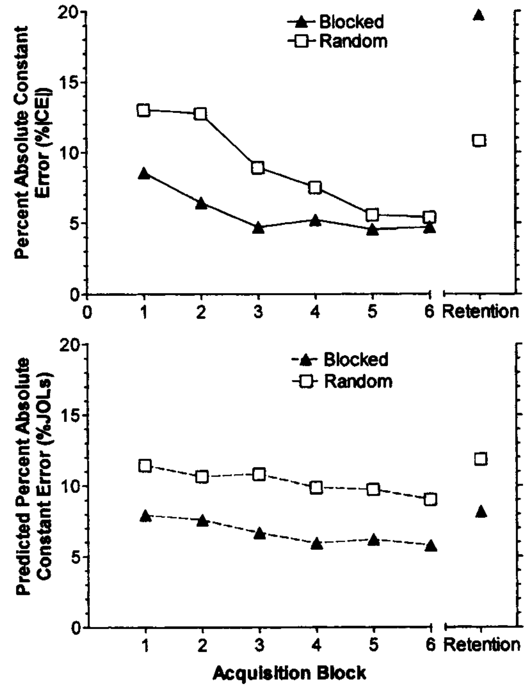


Figure 2. Top panel: Percentage absolute constant error during acquisition and on 24-hr retention test as a function of blocked or random practice during acquisition. Bottom panel: Learner-made predictions (during acquisition and immediately prior to the 24-hr test) of performance on the 24-hr retention test as a function of blocked or random practice.

the advantage of blocked over random practice decreased across trials. There was also a main effect of pattern but the Pattern × Practice Schedule interaction did not reach significance, and so this effect is of secondary interest.[3]

Metacognitive measures. JOLs, predictions about how close the participants thought they would be able to get to the target time for a given movement pattern after a delay of 24 hr (i.e., on the retention test scheduled for the next day), were converted to percentages by dividing the predicted deviations by the corre-

[1] Before they left, participants filled out brief questionnaires designed to get their impressions of their learning experience. Ratings of their practice schedule effectiveness and their own learning were made on likert-type scales. Unfortunately, analysis of these responses revealed no interesting results.

[2] Absolute constant error is the absolute mean deviation from the target for a set of trials for a given participant {|CE| = absolute value of: $(\Sigma(x_i - T)/n)$, where x_i is the time for trial i, T is the target time, and n is the number of trials in the set} (see, e.g., Schmidt & Lee, 1999). Absolute scores are used so that averaging across participants does not lead to an artificially low mean accuracy score. Percent variable errors were also compared, but effects during acquisition were of limited interest and there were no effects in retention. Hence, those analyses have been omitted.

[3] Other dependent measures were collected, and analyzed. No interesting effects emerged for percent variable error (%VE), a measure of variability in movement times; preresponse delay (PRD), a measure of the time between presentation of the to-be-executed pattern and the beginning of the response; or execution errors, errors made such as pressing the wrong sequence of keys.

sponding target MTs and then multiplying by 100. This step facilitated comparisons across the different target times. To be clear, lower %JOL scores indicated greater confidence in the ability to perform close to the target time.

Participants' predictions of their own future (retention) performance are shown in the bottom panel of Figure 2. Their %JOL scores were analyzed via a 2 × 3 × 6 (Practice Schedule × Pattern × Prediction Number) ANOVA with Pattern and Prediction Number as within-subjects factors. The ANOVA supports what seems apparent in Figure 2: The participants in the blocked condition predicted better retention-test performance than did the participants in the random condition, and that difference remained relatively constant across acquisition blocks. The main effect of acquisition condition (M = 6.69% for blocked; M = 10.26% for random) is significant, $F(1, 46) = 15.11$, $MSE = 182.55$, $p < .001$, as is the main effect of block, $F(5, 230) = 6.36$, $MSE = 17.05$, $p < .0001$, but the interaction is not significant, $F < 1$. Thus, unlike the case for %|CE| measures, where the performance of the blocked-practice and random-practice participants converged across the acquisition blocks, the participants' %JOL scores exhibited no such convergence.

Retention Phase

Upon returning to the experimental room on Day 2 of the experiment, the participants were asked to recall the keystroke sequence composing each pattern (i.e., draw it) and the corresponding target MT. Notably, richer memory for the to-be-learned task information was fostered by random than by blocked practice. For both the patterns and their MTs, average recall was better for the random- than the blocked-practice condition (M pattern accuracy: random = 50.0%, blocked = 16.7%; M timing accuracy: random = 83.3%, blocked = 43.1%; $ps < .005$).

After attempting to recall each pattern and its target MT, the participants were asked, again, to predict how close they thought they would be able to get to the target MT for each pattern on the upcoming test. They were then tested in the manner described earlier.

The participants' predictions at the time of the final retention test are shown in the bottom panel of Figure 2; their actual retention-test performance is shown in the top panel of Figure 2. As is apparent in Figure 2, the participants who had received blocked practice during acquisition continued to predict that they would perform better on the retention test than did the participants who had received random practice, but they followed those predictions by performing substantially worse than did the random-practice participants. Both main effects are significant, but in the opposite direction. The mean prediction for the blocked condition (M = 8.19%) was lower than that for the random condition (M = 11.87%), $F(1, 46) = 4.59$, $MSE = 106.39$, $p < .05$; whereas the actual performance in the blocked condition (M = 19.68%) was worse than performance in the random condition (M = 10.80%), $F(1, 46) = 22.29$, $MSE = 254.69$, $p < .001$.

It is also clear in Figure 2 that the discrepancy between predicted and actual performance at the time of the final test comes about because the participants in the blocked practice condition were overconfident to a striking degree. The participants who had had random practice 24 hr earlier were quite accurate in predicting how well they would perform, whereas the participants who had had blocked practice performed at a much poorer level than they thought they would.

Discussion

The random-practice effect is one instantiation of a broader *contextual-interference* principle—namely, that arranging to-be-learned materials or tasks in a way that introduces interference between those materials or tasks typically impedes acquisition performance, but frequently enhances posttraining retention and transfer. Looking only at the performance measures for this experiment, the basic contextual-interference effect (e.g., Shea & Morgan, 1979) was replicated: Blocked practice resulted in superior performance during acquisition but poorer learning, as indexed by performance on a 24-hr retention test. The principal goal of the present experiment was not, however, to replicate the contextual-interference effect, but, instead, to examine the ability of learners to assess, during practice, their "true" level of learning, as distinct from their performance at the time of making such assessments.

It seems clear that the participants in the blocked condition were fooled, so to speak, by their own performance during acquisition. They were unable, apparently, to appreciate that their current level of performance was artificial, propped up, as Bjork (1994, 1999) has argued, by massed practice, which can serve as a kind of crutch during training. Participants in the random-practice conditions also appear to have used their current performance as the primary basis for predicting their future performance, but the fact that those participants were not swayed by the improvements made late in practice is suggestive and interesting. It is possible that the random group's predictions did not move with performance because they were subject to some kind of anchoring effect, but perhaps their learning experience gave them insight into the fact that some of their improvement was likely to be a temporary change. Certainly, random practice led to more accurate predictions, suggesting that under these conditions learners may have had a more realistic idea of their actual learning state.

A particularly striking aspect of the present results is that blocked-practice participants, even immediately prior to the retention test, continued to be so overconfident about their upcoming performance. In advance, one might have expected that participants in that condition would have had a better sense, after being away from the task for 24 hr, of how well the movement patterns had (or had not) been learned (cf. Dunlosky & Nelson, 1994). It is possible that in making these delayed JOLs, participants were influenced by their own predictions from the previous day. That is, instead of making new predictions per se, they may have been recalling their earlier predictions from acquisition. What makes their overconfidence all the more remarkable, though, is that participants in both groups predicted their performance on the final test only after attempting to recall the key sequences and MTs composing each of the earlier practiced patterns. One might have expected that the blocked-practice participants, by virtue of their own (poor) performance on that task, would have been alerted to their less than exemplary learning of those movement patterns.

The present findings have obvious and important practical implications. Overall, participants' judgments of learning were quite good under random-practice conditions and quite poor under blocked-practice conditions. Thus, in real-world contexts, overconfidence in one's ability to perform well at a delay is less likely to arise during conditions of training and instruction that are analogous to the random, interleaved condition in the present research, that is, conditions of training that induce spacing, variation, or unpredictability. One important speculation that arises from the present results is that learners who train under such conditions would be less likely to terminate practice before achieving the level of learning that is the goal of such practice and less prone to attempt a task for which they are unprepared.

Concluding Comment

Benjamin et al. (1998; see also Dunlosky & Nelson, 1992; Kelley & Lindsay, 1993), using verbal materials, have demonstrated that retrieval fluency, that is, how readily something "comes to mind," can be misleading as a subjective–metacognitive index. The present findings support, but also extend, that conclusion. Motor skills, too, are subject to being misassessed by the learner, and objective indexes, such as one's current performance, can also be an unreliable foundation for confidence in one's future performance.

References

Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics, 21,* 627–635.

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

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 435–459). Cambridge, MA: MIT Press.

Chamberlin, C., & Lee, T. D. (1994). Arranging practice conditions and designing instruction. In R. N. Singer, M. Murphey, & L. K. Tennant (Eds.), *Handbook of research on sport psychology* (pp. 213–241). New York: Macmillan.

Christina, R. W., & Bjork, R. A. (1991). Optimizing long-term retention and transfer. In D. Druckman & R. A. Bjork (Eds.), *In the mind's eye: Enhancing human performance* (pp. 23–56). Washington DC: National Academy Press.

Cohen, J., Dearnaley, E. J., & Hansel, C. E. M. (1956). Risk and hazard: Influence of training on the performance of bus drivers. *Operational Research Quarterly, 7,* 67–82.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20,* 374–380.

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.

Ghodsian, D., Bjork, R. A., & Benjamin, A. S. (1997). Evaluating training during training: Obstacles and opportunities. In M. A. Quiñones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications of psychological research* (pp. 63–98). Washington, DC: American Psychological Association.

Hart, J. T. (1965). Memory and the feeling of knowing experience. *Journal of Educational Psychology, 56,* 208–216.

Harvey, N. (1994). Relations between confidence and skilled performance. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 321–352). Chichester, England: Wiley.

Harvey, N., Garwood, J., & Palencia, M. (1987). Vocal matching of pitch intervals: Learning and transfer effects. *Psychology of Music, 15,* 90–106.

Jacoby, L. L., Bjork, R. A., & Kelley, C. M. (1994). Illusions of comprehension, competence, and remembering. In D. Druckman & R. A. Bjork (Eds.), *Learning, remembering, believing: Enhancing human performance* (pp. 57–80). Washington, DC: National Academy Press.

Jacoby, L. L., & Kelley, C. M. (1987). Unconscious influences of memory for a prior event. *Personality and Social Psychology Bulletin, 13,* 314–336.

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 and Language, 32,* 1–24.

Lee, T. D., Wishart, L., Cunningham, S., & Carnahan, H. (1997). Modeled timing information during random practice eliminates the contextual interference effect. *Research Quarterly for Exercise and Sport, 68,* 100–105.

Levy, D., & Bjork, R. A. (2000). *Disentangling learning and performance: The potential of training-based tests of transfer.* Unpublished manuscript.

Magill, R. A., & Hall, K. G. (1990). A review of the contextual interference effect in motor skill acquisition. *Human Movement Science, 9,* 241–289.

Marteau, T. M., Wynne, G., Kaye, W., & Evans, T. R. (1990). Resuscitation: Experience without feedback increases confidence but not skill. *British Medical Journal, 300,* 849–850.

Metcalfe, J., & Shimamura, A. P. (Eds.). (1994). *Metacognition: Knowing about knowing.* Cambridge, MA: MIT Press.

Nelson, T. O. (Ed.). (1992). *Metacognition: Core readings.* Boston: Allyn & Bacon.

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.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3,* 207–217.

Schmidt, R. A., & Lee, T. D. (1999). *Motor control and learning: A behavioral emphasis* (3rd ed.). Champaign, IL: Human Kinetics.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 179–187.

Vesonder, G. T., & Voss, J. F. (1985). On the ability to predict one's own responses while learning. *Journal of Memory and Language, 24,* 363–376.

West, R. F., & Stanovich, K. E. (1997). The domain specificity and generality of overconfidence: Individual differences in performance estimation bias. *Psychonomic Bulletin and Review, 4,* 382–392.

Received March 14, 2000
Revision received November 8, 2000
Accepted December 18, 2000 ■