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

9

Memory and Metamemory Considerations in the Training of Human Beings

Robert A. Bjork

The conventialities of portrait painting are only tolerable in one who is a good painter — if he is only a good portrait painter he is nobody. Try to become a painter first and then apply your knowledge to a special branch — but do not begin by learning what is required for a special branch, or you will become a mannerist.

John Singer Sargent

The mistake we pop stars fall into is stating the obvious. "War is bad. Starvation is bad. Don't chop down the rain forest." It's boring. It's much better to hide it, to fold the meaning into some sort of metaphor or maze, if you like, and for the listener to have a journey to find it.

Sting

In recent papers, Christina and Bjork (1991) and Schmidt and Bjork (1992) have argued that training programs are often much less effective than they could be. A central part of the argument is that individuals responsible for training are often misled as to what are, and are not, effective conditions of practice. Conditions that enhance performance during training are assumed, implicitly or explicitly, to be the conditions of choice with respect to enhancing the goal of training: namely, long-term posttraining performance. That assumption, however, is frequently questionable and sometimes dramatically wrong. Manipulations that speed the rate of acquisition during training can fail to support long-term posttraining performance, while other manipulations that appear to introduce difficulties for the learner during training can enhance posttraining performance.

The goal of the present chapter is to examine two other contributors to nonoptimal training: (1) the learner's own misreading of his or her progress and current state of knowledge during training, and (2) nonoptimal relationships between the conditions of training and the conditions that can be expected to prevail in the posttraining real-world environment.

Memory Considerations

The Goals of Training

The principal goals of a typical training program are to produce optimal transfer of that training to an anticipated posttraining environment of some kind. With rare exceptions, then, the goals of training are long-term goals. We would like the knowledge and skills acquired during training to be durable, not only in the sense of surviving from the end of training to a later time when that knowledge or skill is demanded in a real-world setting, but also in the sense of surviving periods of disuse in the posttraining environment itself.

An equally important long-term goal of training is to produce a mental representation of the knowledge or skill in question that allows for flexible access to that knowledge or skill. We would like the learner to be able to generalize appropriately, that is, to be able to draw on what was learned during training in order to perform adequately in real-world conditions that differ from the conditions of training. Verifying that some individual has ready access to critical skills and knowledge in some standard situation does not, unfortunately, ensure that individual will perform adequately in a different situation, or on altered versions of the task in question. Even superficial changes can disrupt performance markedly. Perceived similarity, or the lack thereof, of new tasks to old tasks is a critical factor in the transfer of training (see, e.g., Gick & Holyoak, 1980). To the extent feasible, a training program should provide a learned representation that permits the learner to recognize when the knowledge and skills acquired during training are and are not applicable to new problems.

Stated in terms of human memory, then, we would like a training program not only to produce a stored representation of the targeted knowledge in long-term memory, but also to yield a representation that remains accessible (recallable) as time passes and contextual cues change. In general, it is explicit or conceptually driven processing of information that we want to optimize, not implicit or stimulus-driven processing (for discussions of the distinction, see Richardson-Klavehn & Bjork, 1988; Roediger & Blaxton, 1987; Roediger & McDermott, 1993; Schacter, 1987; Shimamura, 1986), and we want to optimize the ability to access knowledge and skills, not the ability to judge whether knowledge or skills produced by someone else seem appropriate to the situation. Such distinctions are discussed further in the metamemory section of this chapter.

Relevant Peculiarities of the Human as a Memory Device

Toward achieving the goals of training, it is important to remind ourselves of some of the ways that humans differ from man-made recording devices. We do not, for example, store information in our long-term memories by making any kind of literal recording of that information, but, rather, by relating that new information to what we already know — that is, to the information that already exists in our memories. The process is fundamentally semantic in nature; we store information in terms of its meaning to us, defined by its associations and relationships to other information in our memories. For all practical purposes, our capacity for such storage is essentially unlimited — storing information, rather than using up memory capacity, appears to create opportunities for additional storage. It also appears that once new information is successfully mapped on to existing knowledge in long-term memory, it remains stored, if not necessarily accessible, for an indefinitely long period of time.

The process of accessing stored information given certain cues also does not correspond to the "playback" of a typical recording device. The retrieval of stored information is a fallible, probabilistic process that is more inferential and reconstructive than literal. Information that is readily accessible at one point in time, or in a given situation, may be impossible to recall at another point in time, or in another situation. The information in our long-term memories that is, and is not, accessible at a given point in time is heavily dependent on the cues available to us, not only on cues that explicitly guide the search for the information in question, but also on environmental, interpersonal, mood-state, and body-state cues.

A final relevant peculiarity of human memory is that the act of retrieving information is itself a potent learning event. Rather than being left in the same state it was in prior to being recalled, the retrieved information becomes more recallable in the future than it would have been without having been accessed. In that sense, the act of retrieval is a "memory modifier" (Bjork, 1975). As a learning event, in fact, it appears that a successful retrieval can be considerably more potent than an additional study opportunity, particularly in terms of facilitating long-term recall (see, e.g., Gates, 1917, Hogan & Kintsch, 1971; Landauer & Bjork, 1978). Though not as relevant to the concerns of this chapter, there is also evidence that such positive effects of prior recall on the later recall of the retrieved items can be accompanied by impaired retrieval of competing information, that is, of other items associated to the same cue or set of cues as the retrieved items (for discussions of such retrieval dynamics, see Anderson & Bjork, 1993; Bjork & Bjork, 1992).

In a very general way, then, creating durable and flexible access to critical information in memory is partly a matter of achieving a certain type of encoding of that information, and partly a matter of practicing the retrieval process. On the encoding side, we would like the learner to achieve, for lack of a better word, an *understanding* of the knowledge in question, defined as an encoding that is part of a broader framework of interrelated concepts and ideas. Critical information needs to be multiply encoded, not bound to single sets of semantic or situational cues. On the retrieval side, practicing the actual production of the knowledge and procedures that are the target of training is essential: One chance to actually put on, fasten, and inflate an inflatable life vest, for example, would be of more value — in terms of the likelihood that one could actually perform that procedure correctly in an emergency — than the multitude of times any frequent flier has sat on an airplane and been shown the process by a steward or stewardess. Similar to the argument for multiple encoding, it is also desirable to induce successful access to knowledge and procedures in a variety of situations that differ in the cues they do and do not provide.

The Need to Introduce Difficulties for the Learner

What specific manipulations of training, then, are best able to foster the long-term goals of training, whether stated in terms of measures of posttraining performance or in terms of underlying memory representations? Attempting to answer that question in any detail would involve prescribing a mixture of desirable manipulations, and there would clearly be some disagreement among researchers as to what set of manipulations constitute the optimal mixture. Any such prescription would also need to be tailored to the specifics of a given training mission. Whatever the exact mixture of manipulations that might turn out to be optimal, however, one general characteristic of that mixture seems clear: It would introduce many more difficulties and challenges for the learner than are present in typical training routines. Recent surveys of the relevant research literatures (see, e.g., Christina & Bjork, 1991; Farr, 1987; Reder & Klatzky, 1993; Schmidt & Bjork, 1992) leave no doubt that many of the most effective manipulations of training — in terms of post-training retention and transfer — share the property that they introduce difficulties for the learner. Some of the clearest examples of such manipulations are the following.

Varying the Conditions of Practice

It has now been demonstrated in a variety of ways, and with a variety of motor, verbal, and problem-solving tasks, that introducing variation and/or unpredictability in the training environment causes difficulty for the learner but enhances long-term performance — particularly the ability to transfer training to novel but related task environments. Where several differing motor-movement tasks are to be learned, for example, scheduling the practice trials on those tasks in random fashion, rather than blocking the trials by task type, has been shown to impair performance during training but enhance long-term performance (Shea & Morgan, 1979; Hall, Domingues, & Cavazos, 1992). Analogous results have been obtained with problem-solving tasks (e.g., Reder, Charney, & Morgan, 1986). Similarly, varying the parameters of a to-be-learned task — by, for example, varying the speed or distance of a target — impairs performance during training but enhances posttraining performance (e.g., Catalano & Kleiner, 1984; Kerr & Booth, 1978). And the effects of increasing the variety, types, or range of exercises or problems (e.g., Carson & Wiegand, 1979; Gick & Holyoak, 1983; Homa & Cultice, 1984) tend to exhibit the same general pattern. Even varying the incidental environmental context in which learning sessions are situated has been shown to enhance long-term retention (Smith, Glenberg, & Bjork, 1978; Smith & Rothkopf, 1984).

Providing Contextual Interference

Such ways of making the task environment more variable or unpredictable can be considered one set of a broader category of manipulations that produce "contextual interference" (Battig, 1979). Other examples of contextual interference include designing or interleaving materials to be learned in a way that creates, at least temporarily, interference for the learner (e.g., Mannes & Kintsch, 1987), and adding to the task demands (e.g., Battig, 1956; Langley & Zelaznik, 1984). In Mannes and Kintsch's experiment, for example, subjects had to learn the content of a technical article (on industrial uses of microbes) after having first studied an outline that was either consistent with the organization of the article or inconsistent with that organization (but provided the same information in either case). The inconsistent condition impaired subjects' verbatim recall and recognition of the article's content (compared to the consistent condition), but facilitated performance on tests that required subjects to infer answers or solve problems based on their general understanding of the article's content.

Distributing Practice on a Given Task

In general, compared to distributing practice sessions on a given task over time, massing practice or study sessions on to-be-learned procedures or information produces better short-term performance or recall of that procedure or information, but markedly inferior long-term performance or recall. The long-term advantages of distributing practice sessions over time have been demonstrated repeatedly for more than a century, tracing back over the entire history of controlled research on human memory (for modern reviews, see Dempster, 1990; Glenberg, 1992; Lee & Genovese, 1988). The differing short-term and long-term consequences of distributing practice sessions are nicely illustrated by the results of an experiment by Bahrick (1979). The subjects' basic task was to learn the Spanish translations of a list of 50 English words. During each of several training sessions on the list, an alternating series of study and test trials were presented until a given subject had responded correctly in Spanish to every English word on the list (once a given Spanish word was given correctly that English-Spanish word pairing was dropped out of the next study trial). Successive training sessions were separated by 0, 1, or 30 days, and at the start of every training session after the first subjects were tested on their memory for all 50 words. Looking at those tests alone, performance was clearly poorest with the 30-day separation, was better with the 1-day separation, and better yet when the several training sessions were all on a single day (the 0-day separation). On a test of long-term retention, however, administered 30 days after the last training session, the levels of recall were dramatically reversed, with the 30-day spacing of training sessions yielding clearly superior recall (72% after three training sessions, versus 33% and 64% in the 0-day and 1-day conditions, respectively).

Reducing Feedback to the Learner

Until recently, a common generalization about motor skills was that providing external feedback to the learner facilitates the acquisition of skills, and that any means of improving such augmented feedback — by, for example, making it more immediate, more frequent, or more accurate — helps learning and performance. Recently, however, Richard Schmidt and his collaborators (see, e.g., Schmidt, 1991; Schmidt, Young, Swinnen, & Shapiro, 1989; Winstein & Schmidt, 1990) have found that — as in the case of the other manipulations summarized in this section — reducing the frequency of feedback makes life more difficult for the learner *during* training, but can enhance posttraining performance. They have demonstrated that providing summary feedback to subjects (after every 5 or 15 trials, for example), or "fading" the frequency of feedback over trials, impedes acquisition of simple motor skills but enhances long-term retention of those skills.

Using Tests as Learning Events

Such effects of reducing the frequency of feedback during the learning of motor skills are broadly consistent with a large verbal-memory literature on tests as learning events. As mentioned earlier, there is abundant evidence that the act of retrieval induced by a recall test can be considerably more potent than a study opportunity in facilitating future recall. Prior testing also appears to increase the learning that takes place on subsequent study trials (e.g., Izawa, 1970). Once again, however, using tests rather than study trials as learning events, or increasing the difficulty of such tests, may appear to be counterproductive *during* training. Hogan and Kintsch (1971), for example, found that study trials produced better recall at the end of an experimental session than did test trials, but that test trials produced better recall after a 48-hour delay. And Landauer and Bjork (1978; see also Rea & Modigliani, 1985) found that "expanding retrieval practice," in which successive recall tests are made progressively more difficult by increasing the time and intervening events prior to each next test of some target information, facilitates long-term recall substantially — compared to the same number of tests administered at constant (and easier) delays.

It is not the mission of the present chapter to put forth any detailed conjectures as to why each of the foregoing manipulations induces desirable encoding and/or retrieval operations. In a general way, it seems safe to say that in responding to the difficulties and challenges induced by such manipulations the learner is forced into more elaborate encoding processes and more substantial and varied retrieval processes. As Battig (1979) argued with respect to contextual interference, and Schmidt and Bjork (1992) have argued more broadly, such manipulations are likely to induce more "transfer appropriate processing" (Bransford, Franks, Morris, & Stein, 1979; Morris, Bransford, & Franks, 1977), that is, processing that will transfer to the posttraining environment. For present purposes, however, the central point is that the research picture is unambiguous: A variety of manipulations that impede performance during training facilitate performance on the long term.

Misperceptions of the Trainer

If the research picture is so clear, why then are massed practice, excessive feedback, fixed conditions of training, and limited opportunities for retrieval practice — among other nonproductive manipulations — such common features of real-world training programs? It is tempting to argue that there should be more venues for interaction, and vehicles of communication, between researchers and practitioners, and that might be true. More important than any underexposure to relevant research findings, however, is the fact that the typical trainer is overexposed, so to speak, to the day-to-day performance and evaluative reactions of his or her trainees. A trainer, in effect, is vulnerable to a type of operant conditioning, where the reinforcing events are improvements in the performance and/or happiness of trainees. Such a conditioning process, over time, can act to shift the trainer toward manipulations that increase the rate of correct responding — that make the trainee's life easier, so to speak. Doing that, of course, will move the trainer away from introducing the types of desirable difficulties summarized in the preceding section.

The tendency for instructors to be pushed toward training programs that maximize the performance or evaluative reaction of their trainees *during* is exacerbated by certain institutional characteristics that are common in real-world organizations. First, those responsible for training are often themselves evaluated in terms of the performance and satisfaction of their trainees during training, or at the end of training. Second, individuals with the day-to-day responsibility for training often do not get a chance to observe the posttraining performance of the people they have trained; a trainee's later successes and failures tend to occur in settings that are far removed from the original training environment, and from the trainer himself or herself. It is also rarely the case that systematic measurements of posttraining on-the-job performance are even collected, let alone provided to a trainer as a guide to what manipulations do and do not achieve the posttraining goals of training. And, finally, where refresher or retraining programs exist, they are typically the concern of individuals other than those responsible for the original training.

Metamemory Considerations

A second consideration in the training of human beings, arguably as important as the actual learning produced by a training program, is the extent to which trainees gain a valid assessment of their own state of learning or competence. Individuals who have illusions of comprehension or competence pose a greater hazard to themselves and others than do individuals who correctly assess that they lack some requisite information or skill. The reading we take of our own state of knowledge determines whether we seek further study or practice, whether we volunteer for certain jobs, whether we instill confidence in others, and so forth. In general, then, as argued by Jacoby, Bjork, and Kelley (1993), it is as important to educate subjective experience as it is to educate objective experience.

As it turns out, it is not just those individuals responsible for training who are susceptible to being fooled by the level of performance of trainees during training. Recent research suggests that the learner himself or herself is susceptible to the same type of inferential error. Rapid progress in the form of improved performance is reassuring to the learner, even though little learning may be taking place, whereas struggling and making errors are distressing, even though substantial learning may be taking place. Such a misreading of one's progress, together with the other types of misassessments discussed below, can lead trainees to prefer less effective training over more effective training. Baddeley and Longman (1978), for example, found that British postal workers who were taught a keyboard skill under massed-practice (and less efficient) conditions actually were more satisfied with their training than were workers taught under spaced-practice (and more efficient) conditions.

Relevant Peculiarities of the Human as a Memory Device

At the root of such problems is our misunderstanding of the complexities of our own memories. Human memory is multidimensional and multifaceted in ways that we apparently do not come to realize on the basis of the trials and errors of everyday experience alone. We seem to persist in holding to a kind of implicit assumption that what we can and cannot recall or recognize is governed by memory

traces that vary on a unidimensional strength continuum — that past experiences of differing duration and intensity leave impressions or traces in the brain that are like footprints of differing depths in the sand. And such traces or footprints are subject to blurring over time, becoming harder to read as a function of retention interval and intervening events.

From a research standpoint, any such unidimensional idea, if ever plausible, is now preposterous. During the last decade particularly, the research of behavioral scientists, neuroscientists, and clinicians, employing subject populations ranging from animals and children to amnesic patients and normal adults has yielded a picture of human memory that is remarkably multifaceted. In response to an array of evidence of various types that implicate differing processes and types of memories, researchers have proposed a bewildering assortment of overlapping and nonoverlapping distinctions: short-term versus long-term memory, semantic versus episodic knowledge, declarative versus procedural knowledge, stimulus-driven versus conceptually driven knowledge, explicit versus implicit memories, controlled versus automatic processing, and memory as a tool versus memory as an object, to name a few.

Whatever the resolution of the current terminological turmoil, the important point for present purposes is that one subjective or objective measure of the "strength" of a memory representation may not correlate with the "strength" of a different subjective or objective measure. The research literature is now replete, for example, with a variety of dramatic interactions of encoding condition and test condition on performance. Encoding conditions or processes that yield good short-term performance can fail to support long-term performance as stressed above. Encoding conditions/processes that facilitate later recognition may not support later recall, and vice versa. And initial conditions of exposure that do and do not prime performance on indirect measures of performance, such as perceptual identification or word-fragment completion, can differ markedly from the conditions that facilitate performance on direct measures, such as recall and recognition (for some striking examples, see Roediger & Blaxton, 1987; for a review, see Richardson-Klavehn & Bjork, 1988).

Misperceptions of the Learner: Using One Index to Predict Another

Failing to understand the multifaceted nature of human memory opens the learner to a variety of misassessments of his or her state of knowledge during training. As mentioned already, the learner may be fooled by his or her own successes during training. Manipulations such as blocking practice by subtask, providing continuous feedback during training, and fixing the conditions of practice act like crutches that artificially support performance during training. When those crutches are absent in the posttraining environment, performance collapses. The learner, however, will typically lack the perspective and experience to realize that he or she has not yet achieved the level of learning demanded by the posttraining environment. Conversely, the errors and confusion caused during training by spaced practice, infrequent feedback, and variations in the task or task environment can lead trainees to underestimate their own state of learning and comprehension.

At a somewhat oversimplified level of analysis, such misassessments arise as a function of trainees observing their own objective performance during training. They then assume, implicitly or explicitly, that successes predict future successes and failures predict future failures. In effect, the learner relies too heavily on an unreliable index — the current ease of access to a correct answer or procedure — as a measure of the extent to which learning in a broader sense has been achieved.

As an overall generalization from all of our past experiences, of course, ease of retrieving some procedure or information *does* provide a measure of how well that procedure or information is registered in memory. The problem is that there are multiple determinants of speed or ease of retrieval, only some of which are commensurate with degree of learning. The type of training "crutches" mentioned above increases the speed and probability of retrieval via such mechanisms as constraining the possible responses, multiplying retrieval cues, and tapping short-term memory — that is, processes different from those that might truly build the long-term representation of some procedure or knowledge. Apparently, however, we lack the type of understanding of our own memories that

would permit us to distinguish between the different sources of retrieval speed or probability.

A recent experiment by Kelley and Lindsay (1993) serves as a good illustration of that point. Using a general knowledge test, Kelley and Lindsay found, not surprisingly, that subject's confidence in the correctness of a given answer increased as a function of how rapidly that answer was given. They also found that having subjects read a list of answers prior to being given the general knowledge test increased the speed with which those answers were given, and the subjects' confidence in those answers — whether those answers were right or wrong. That is, if a closely related but incorrect answer (e.g., Hickock) to a given question (What was Buffalo Bill's last name?) had been read earlier, subjects gained an illusion of knowing: Such studied incorrect answers were not only given more frequently, they were given more confidently.

Speed or ease of retrieval access is only one type of index or measure that is subject to misinterpretation. A wealth of recent experimental evidence from several research paradigms suggests that the sense of familiarity or fluency during the encoding of retrieval cues can also be a source of illusions of knowing or comprehending. Reder (1987, 1988), for example, found that she could alter subjects' feeling-of-knowing judgments simply by making certain words in a general-information question more familiar. When key words in a question (such as "golf " and "par" in the question, "What is the term in golf for scoring one under par"?) were prefamiliarized by virtue of having appeared on an earlier experimental task, subjects were then more likely to judge the question as answerable.

Schwartz and Metcalfe (1992) and Reder and Ritter (1992) have demonstrated that not only is cue familiarity a factor in subjects' feeling-of-knowing judgments, it may be a more important factor than target familiarity. Schwartz and Metcalfe had subjects study a list of unrelated cue-target word pairs (such as OAK TURTLE) and then later tested subjects' cued recall of the target words. When subjects were unable to recall a given target (such as "TURTLE") in response to its cue ("OAK"), they were asked to give a feeling of knowing judgment, which took the form of rating their likelihood of being able to later recognize the correct target from among several alternatives. With certain types of general-information questions such judgments can be quite accurate (e.g., Hart, 1967a), though subjects tend to be overconfident (for a review see Nelson & Narens, 1990), and one theory is that it is recall of partial information — a first letter, for example, or whether the word is short or long — that is the basis for such judgments. Schwartz and Metcalfe found, however, that prefamiliarizing cue words increased subjects' feeling-of-knowing judgments without increasing the likelihood of recall of the targets associated with those cues, or having an effect on the accuracy of such judgments, whereas prefamiliarizing the target words had no effect on subjects' feeling of knowing. Consistent with that pattern, Reder and Ritter (1992) found that subjects' speeded judgments of whether they knew the answer to a given arithmetic problem (such as 13 times 27) was more heavily influenced by the frequency of prior exposures to the terms of the problem than by the actual degree of learning (as indexed by the frequency of prior exposures to the intact problem itself).

In terms of their real-world implications, a possible concern about the foregoing results is that the experimental tasks employed may be too artificial and/or simple to be compared to the types of tasks that are the typical objects of training. However, an impressive series of experiments by Arthur Glenberg, William Epstein, and their collaborators (Epstein, Glenberg, & Bradley, 1984; Glenberg & Epstein, 1985, 1987; Glenberg, Sanocki, Epstein, & Morris, 1987; Glenberg, Wilkinson, & Epstein, 1982) does much to allay that concern. The basic paradigm involves having subjects read expository text covering relatively technical content and then rate their comprehension of that material — in terms of the likelihood that they will later be able to answer questions on that material. In general, the subjects were poorly calibrated: The correlations of their judged comprehension and their later actual ability to answer correctly were surprisingly low. Consistent with the work of Reder (1987, 1988; Reder & Ritter, 1992) and Schwartz and Metcalfe (1992; see also Metcalfe, Schwartz, & Joaquim, 1993), subjects appear to be vulnerable to illusions of comprehension based on the general familiarity of the domain in question. Glenberg and Epstein (1987), for example, found that subjects' judgments were apparently more influenced by their self-classification of their own level of expertise than by their actual comprehension of the specific content of a text passage. Within a given domain,

such as physics or music, level of expertise was actually inversely related to the calibration of comprehension! With a different paradigm, Costermans, Lories, and Ansay (1992) also obtained results consistent with the idea that subjects use one index, their general familiarity with a knowledge domain, to predict another, the degree to which the answer to a specific question exists in their memories.

A final important point, closely related to misreading the meaning of subjective familiarity, is that the learner is subject to hindsight biases (Fischhoff, 1975). Once an answer is provided or a solution is demonstrated, we appear unable to correctly assess the likelihood that we could have provided that answer or solved that problem ourselves. More specifically, we are subject to an "I knew it all along" effect. Given the nature of real-world instruction and training, the implications of the hindsight effect are profound. In a variety of ways we are put in the position of judging our level of comprehension on the basis of an exposure to the information or problem-solving procedure in question. As a student, for example, we make judgments of what we know and do not know (and, hence, how we should allocate our study time) based on reading a text or listening to an instructor. Such judgments, however, contaminated as they are by familiarity effects, hindsight biases, and other factors — such as the ease of following a "well polished" lecture — are a poor basis for judging one's ability to produce an answer or solve a problem.

The Need to Introduce Difficulties for the Learner

One implication of such misperceptions of the learner is that the conditions of training should provide meaningful rather than misleading subjective experiences. In designing training programs we are at risk of denying trainees the opportunity for certain types of feedback that are essential to their achieving a valid assessment of their current state of knowledge.

We can, in effect, inadvertently ruin the learner's subjective experience. Experiments by Jacoby and Kelley (1987) and Dunlosky and Nelson (1992; see also Nelson & Dunlosky, 1991) illustrate that point. Jacoby and Kelley presented a number of anagrams to subjects and asked the subjects to rate the difficulty of each anagram in terms of the likelihood that other people could solve it. In one condition, subjects had to first solve the anagram (e.g., FSCAR ?????), and in another condition the anagram was presented together with its solution (FSCAR SCARF). Subjects ratings in the former condition, presumably based largely or entirely on their own subjective solution experience, were considerably more accurate than subjects' ratings in the latter condition. Being given the solution to a given anagram apparently ruined a subject's opportunity to experience the solution process, which then forced them to use some less-predictive "theory" of what makes anagrams more or less difficult to solve.

Dunlosky and Nelson (1992) had subjects study a series of unrelated cue-target word pairs (e.g., WEED JURY). Interleaved among the study trials were judgments-of-learning (JOL) trials on which subjects were to judge their degree of learning of a particular pair presented earlier. Such JOL trials were immediate or delayed in terms of when they followed the study trial of the pair to be judged, and they consisted of the cue alone (WEED ????) or the intact cue target pair (WEED JURY). Subjects were asked to predict the likelihood they would be able, 10 minutes later, to recall the target when given the cue. Such predictions were unreliable for either type of JOL trial administered immediately, were not much better on delayed cue-target JOL trials, and were very good on delayed cue-alone JOL trials. One interpretation is that it is only on the delayed cue-alone trials that subjects get any kind of valid subjective experience as to their state of learning of a given pair (for an expansion of that argument, see Spellman & Bjork, 1992). On the immediate cue-alone JOL trials subjects can interpret ease of access from short-term memory as evidence of learning; on cue-target JOL trials, either immediate or delayed, subjects are vulnerable to the effects of familiarity and hindsight discussed in the preceding section.

In general, then, a major goal of training should be to inform the learner's own subjective experience. People need to experience the type of testing to which they will later be subjected (see Glenberg & Epstein, 1987), and, to the extent possible, questions embedded in training need to be phrased such that the processes tapped in answering those questions are the same processes that support long-term retention (see Begg, Duft, Lalonde, Melnick, & Sanvito, 1989). Stated more broadly, the conditions of training need to be constructed to reveal to the subject what knowledge and procedures are, and are not, truly accessible under the types of conditions that can be expected to prevail in the posttraining environment. Some of the best ways to achieve that goal involve making life seem more difficult for the learner. Manipulations such as varying the conditions of training, inducing contextual interference, distributing practice, reducing the frequency of augmented feedback, and using tests as learning events share the property that they act to better educate the learner's subjective experience.

It may be necessary, however, to educate the learner in another respect as well. For people to be receptive to the types of manipulations of training suggested herein, institutional and individual attitudes toward the meaning of errors and mistakes must change. People learn by making and correcting mistakes. We have known at least since an influential paper by Estes (1955; see also Cuddy & Jacoby, 1982) that it may be necessary to induce forgetting during training to enhance learning. Training conditions that prevent certain mistakes from happening (and give trainees a false optimism about their level of comprehension and competence) can defer those mistakes to a posttraining setting where they really matter. That is an especially important consideration in certain job contexts, such as police work, air-traffic control, and nuclear-plant operation, where society cannot afford the kind of on-the-job learning such mistakes might entail. Stated most strongly, when embarked on any substantial learning enterprise we should probably find the absence, not the presence, of errors, mistakes, and difficulties to be distressing — a sign that we are not exposing ourselves to the kinds of conditions that most facilitate our learning, and our self-assessment of that learning.

Should the Posttraining Environment be Simulated during Training?

A broad implication of the foregoing analyses is that training is frequently nonoptimal because it fails to incorporate the variability, delays, uncertainties, and other challenges the learner can be expected to face in a real-world job setting of some kind. It would seem, then, that optimizing training may be a simple matter — in principle, if not in practice — of simulating the posttraining environment during training. Such an assumption is clearly one rationale for spending massive amounts of money on high fidelity simulators in the aircraft industry and elsewhere.

At one level, it seems incontestable that the learner should experience conditions during training that are analogous or identical to those expected in the posttraining environment. But to what degree is it necessary to simulate the physical and social details of real-world settings in order to achieve that end? A strong position on that issue is staked out by advocates of the "situated learning" approach (see, e.g., Greeno, Smith, & Moore, 1993; Lave & Wenger, 1991). In that theoretical framework, it is critical to situate the learner in the context of application. The argument is that learning processes cannot be separated from contextual determinants of performance, particularly social aspects of context, and that learning by abstraction — as in a classroom — is ineffectual. That extreme position is the topic of considerable current debate among social scientists and educators (for an excellent review of the issues and relevant data, see Reder & Klatzky, 1993).

But is it really necessary to simulate the posttraining environment to induce processing that will transfer to that environment? It is an intriguing possibility that the conditions of learning should, in a sense, go beyond situated learning. That is, it may be optimal, from both a memory and metamemory standpoint, to introduce difficulties of certain types that are *not* anticipated in the real-world environment. Introducing more variability than one expects to be present in the real world, for example, or reducing the anticipated frequency of augmented feedback, may result in a more elaborated and internalized representation of knowledge and skills.

Such a possibility is suggested by the results of certain of the experiments on induced variability of practice cited earlier. Shea and Morgan (1979), for example, found that a random schedule of practice on several different motor-movement patterns — as opposed to blocked practice on those patterns — not only produced much superior transfer to a posttraining test carried out under random conditions, but also produced better transfer to a post-training test

carried out under blocked conditions. That Shea and Morgan's results — which were obtained using relatively simple motor tasks in a laboratory environment — may well generalize to real-world settings is suggested by the results of a recent experiment by Hall et al. (1992). With the cooperation of the coaches of the varsity baseball team at the California Polytechnic University, San Luis Obispo, they arranged for extra batting practice to be given under either blocked or random conditions. Twice a week for 6 weeks, two matched subsets of players were thrown 45 pitches—15 fast balls, 15 curve balls, and 15 change-ups — under blocked or random conditions. Players in the blocked condition got those pitches blocked by type, whereas successive pitches in the random condition were determined by a random schedule. At the end of those 6 weeks, two transfer tests were administered, the first under random conditions and the second under blocked conditions. As in Shea and Morgan's experiment, random practice produced better transfer to blocked as well as random conditions than did blocked practice.

Using 8-year-old and 12-year-old children as subjects, Kerr and Booth (1978) obtained analogous results with a somewhat different paradigm. The task involved throwing miniature beanbags underhanded at a 4 inch by 4 inch target on the floor. In the case of the 8-year-old children, one group was given training at a fixed distance (3 ft), while another group was given the same number of training trials, half at 2 ft and half at 4 ft (but mixed across trials). On a posttraining transfer test carried out at a 3 ft distance, the group that practiced at 2 and 4 ft, but never at 3 ft, performed better than did the group that practiced *at* the criterion distance! With the same procedure, but with the distance increased by a foot, the outcome was the same for 12-year-old subjects.

Results of the foregoing type suggests that certain benefits that accrue from contending with variation and unpredictability may outweigh the benefits of having an exact match of the training and posttraining task environments. Another important consideration may argue against constraining the training environment to be the same as the anticipated posttraining environment: It may not be optimal to "contextualize" the learning process, even within the context that is the target of training. The problem is twofold. On the one hand, fixing environmental and task conditions during training, whether those conditions correspond to the posttraining target context or not, may reduce the frequency of the types of desirable processing induced by variation. On the other hand, the environmental, social, and task characteristics of any given job environment are not all that predictable. Equipment, physical settings, procedures, and co-workers usually differ across locations, or change with time. And emergencies and other unusual events are, almost by definition, hard to predict. It is in such special circumstances that the risk of having contextualized training may be greatest. If we want people to respond optimally to unanticipated novel conditions, such as emergencies and/or unique conditions of some other type, the evidence summarized in this chapter suggests that we do not want to have trained those people under fixed conditions.

Such issues are obviously crucial in the complex business of optimizing the design of simulators. Comparisons of high-fidelity (and high cost) simulators to simpler (and lower cost) simulators have often failed to demonstrate that high fidelity facilitates learning. It has been argued that high fidelity can even be detrimental early in learning by providing cues and complexities that are confusing in the early stages of learning (Andrews, 1988). Consistent with the theme of this chapter, it could also be argued that there are some benefits of *not* providing every bell and whistle present in the real-world apparatus. Simulators that require the learner to substitute imagery for external cues as a means of keeping track of the state of the system, for example, might facilitate higher levels of learning.

To argue that high fidelity is never necessary in a simulator is clearly unwarranted theoretically and empirically. Research with aircraft simulators has demonstrated that high fidelity can be very important for certain aspects of performance. But overall, as Patrick (1992) has argued, the most important determiner of transfer is likely to be psychological fidelity, not engineering fidelity. An extension of that argument may be the best single answer to the question raised at the start of this section. It is not the nominal overlap of the training and real-world environments that really matters, but, rather, the functional overlap. Our goal should be to best exercise during training the types of processing that performing at a high level in the posttraining environment will demand.

Concluding Comments

One implication of the considerations summarized in this chapter is that intuition and standard practice are poor guides to training. The body of research on human cognitive processes, though far from fully developed, has grown to the point where it provides a far better guide. A second implication is that, as a guide to training, research on the learner's metacognitive processes is as important, and inseparable from, research on the objective consequences of training.

Acknowledgment

This chapter was written while the author, as a Visiting Scholar, enjoyed the support and hospitality of the Department of Psychology, Dartmouth College.

Acknowledgment

This chapter was written while the author, as a Visiting Scholar, enjoyed the support and hospitality of the Department of Psychology, Dartmouth College.

REFERENCES

Anderson, M. C., & Bjork, R. A. (In press). Retrieval inhibition in long-term memory: Description or explanation? In D. Dagenbach & T. Carr (Eds.) Inhibition in Attention, Memory, and Language. New York: Academic Press.

Andrews, D. H. (1988). Relationships among simulators, training devices, and learning: A behavioral view. Educational Technology, January, 48-54.

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.

Bahrick, H.P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. Journal of Experimental Psychology: General, 108, 296-308.

Battig, W.F. (1956). Facilitation and interference. In E.A. Bilodeau (Ed.), Acquisition of skill (pp. 215-244). New York: Academic Press.

Battig, W. F. (1979). The flexibility of human memory. In L. S. Cermak & F.I.M. Craik (Eds.), Levels of processing in human memory (pp. 23-44). Hillsdale, N. J.: Erlbaum.

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.

Bjork, R.A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), Information processing and cognition: The Loyola Symposium (pp. 123-144). Hillsdale, NJ: Laurence Erlbaum.

Bjork, R.A., & Bjork, E.L. (1992). A new theory of disuse and an old theory of stimulation fluctuation. In A.F. Healy, S.M. Kosslyn, & R.M. Shiffrin (Eds.), From learning processes to cognitive processes: Essays in honor of William K. Estes, Vol. 2 (pp. 35-67). Hillsdale, NJ: Erlbaum.

Bransford, J.D., Franks, J.J., Morris, C.D., & Stein, B.S. (1979). Some general constraints on learning and memory research. In L.S. Cermack & F.I.M. Craik (Eds.), Levels of processing in human memory (pp. 331-354). Hillsdale, NJ: Erlbaum.

Carson, L. M., & Wiegand, R. L. (1979). Motor schema formation and retention in young children: A test of Schmidt's schema theory. Journal of Motor Behavior, 11, 247-251.

Catalano, J.F., & Kleiner, B.M. (1984). Distant transfer in coincident timing as a function of variability of practice. Perceptual and Motor Skills, 58, 851-856.

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.

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.

Cuddy, L.J., & Jacoby, L.L. (1982). When forgetting helps memory: Analysis of repetition effects. Journal of Verbal Learning and Verbal Behavior, 21, 451-467.

Dempster, F. N. (1990). The spacing effect: A case study in the failure to apply the results of psychological research. American Psychologist, 43, 627-634.

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.

Epstein, W., Glenberg, A.M. & Bradley, M.M. (1984). Coactivation and comprehension: Contribution of text variables to the illusion of knowing. Memory and Cognition, 12, 355-360.

Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. Psychological Review, 62, 369-377.

Farr, M. J. (1987). The Long-term Retention of Knowledge and Skills: A Cognitive and Instructional Perspective. New York: Springer-Verlag.

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effects of outcome knowledge on judgment under uncertainty. Journal of Experimental Psychology: Human Perception and Performance, 1, 288-299.

Gates, A. I. (1917). Archives of Psychology #40

Gick, M. L. & Holyoak, K. J. (1980). Analogical problem solving. Cognitive Psychology, 12, 306-355.

Glenberg, A. M. (1992). Distributed practice effects. In L. R. Squire (Ed.), Encyclopedia of Learning and Memory. (pp. 138-142) New York, NY: Macmillan.

Glenberg, A.M. & Epstein, W. (1985). Calibration of comprehension. Journal of Experimental Psychology: Learning, Memory & Cognition, 11, 702-718.

Glenberg, A.M. & Epstein, W. (1987). Inexpert calibration of comprehension. Memory & Cognition, 15(1), 84-93.

Glenberg, A.M., Sanocki, T., Epstein, W. & Morris, C. (1987). Enhancing calibration of comprehension. Journal of Experimental Psychology: General, 116, 119-136.

Glenberg, A.M., Wilkinson, A.C. & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. Memory and Cognition, 10, 597-602.

Greeno, J.G., Smith, D.R., & Moore, J.L. (1993). Transfer of situated learning. In D.K. Detterman and R.J. Sternberg (Eds.), Transfer on Trial: Intelligence, Cognition, and Instruction. (pp. 99-167) Norwoood, NJ: Ablex.

Hall, K. G., Domingues, E. & Cavazos, R. (1992). The effects of contextual interference on extra batting practice. Unpublished paper.

Hart, J.T. (1967). Memory and the memory-monitoring process. Journal of Verbal Learning and Verbal Behavior, 6, 685-691.

Hogan, R.M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. Journal of Verbal Learning and Verbal Behavior, 10, 562-567.

Homa, D., & Cultice, J. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. Journal of Experimental Psychology: Learning, Memory & Cognition, 10, 83-94.

Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. Journal of Experimental Psychology, 83, 340-344.

Jacoby, L. L., Bjork, R. A., & Kelley, C. M. (In press). Illusions of comprehensions and competence. In D. Druckman and R. A. Bjork (Eds.), Learning, Remembering, Believing: Enhancing Individual and Team Performance. Washington, DC: National Academy Press.

Jacoby, L. L., & Kelley, C. M. (1987). Unconscious influences of memoryfor 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.

Kerr, R., & Booth, B. (1978). Specific and varied practice of a motor skill. Perceptual and Motor Skills, 46, 395-401.

Koriat, A. & Lieblich, I. (1974). What does a person in a "TOT" state know that a person in a "don't know" state doesn't know? Memory & Cognition, 2, 647-655.

Landauer, T.K., & Bjork, R.A. (1978). Optimum rehearsal patterns and name learning. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.), Practical aspects of memory (pp. 625-632). London: Academic Press.

Langley, D. J., & Zelaznik, H. N. (1984). The acquisition of time properties associated with a sequential motor skill. Journal of Motor Behavior, 16, 275-301.

Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge, England: Cambridge University Press.

Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. Research Quarterly for Exercise and Sport, 59, 277-287.

Lee, T.D., & Magill, R.A. (1983). The locus of contextual interference in motor-skill acquisition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 730-746.

Mannes S.M., & Kintsch, W. (1987). Knowledge organization and text organization. Cognition and Instruction, 4, 91-115.

Metcalfe, J., Schwartz, B.L., & Joaquim, S. G. (in press). The cue familiarity heuristic in metacognition. Journal of Experimental Psychology: Learning, Memory, and Cognition.

Morris, C.D., Bransford, J.D., & Franks, J.J. (1977). Levels of processing versus transfer appropriate processing. Journal of Verbal Learning and Verbal Behavior, 16, 519-533.

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. Bower (Ed.), The psychology of learning and motivation (Vol.26, pp.125-141). San Diego, CA: Academic Press.

Patrick, J. (1992). Training: Research and Practice. San Diego, CA: Academic Press.

Rea, C.P., & Modigliani, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. Human Learning, 4, 11-18.

Reder, L. M. (1987). Selection strategies in question answering. Cognitive Psychology, 19, 90-138.

Reder, L. M. (1988). Strategic control of retrieval strategies. In G. Bower (Ed.), The Psychology of Learning and Motivation, (Vol. 22, pp. 227-259). San Diego, CA: Academic Press.

Reder, L. M., Charney, D. H. & Morgan, K. I. (1986). The role of elaborations in learning a skill from an instructional text. Memory and Cognition, 14, 64-78.

Reder, L. M., & Klatzky, R. L. (in press). The effect of context on training: Is learning: Is learning situated? In D. Druckman and R. A. Bjork (Eds.), Learning, Remembering, Believing: Enhancing Individual and Team Performance. Washington, DC: National Academy Press.

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

Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. Annual Review of Psychology, 39, 475-543.

Roediger, H. L., & Blaxton, T. A. (1987). Retrieval modes produce dissociations in memory for surface information. In D. Gorfein an R. R. Hoffman (Eds.), Memory and Cognitive Processes: The Ebbinghaus Centennial Conference. (pp. 349-379). Hillsdale, NJ: Erlbaum.

Roediger, H. L., & McDermott, K. B. (in press). Implicit memory in normal human subjects. In H. Spinnler and F. Boller (Eds.), Handbook of Neuropsychology, Vol. 8. Amsterdam: Elsevier.

Schacter, D. L. (1987). Implicit memory: History and current status. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 84-99.

Schmidt, R.A. (1991). Frequent augmented feedback can degrade learning: Evidence and interpretations. In G.E. Stelmach & J. Requin (Eds.), Tutorials in motor neuroscienc (pp. 59-75). Dordrecht: Kluwer Academic Publishers.

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., Young, D.E., Swinnen, S., & Shapiro, D.C. (1989). Summary knowledge of results for skill acquisition: Support for the guidance hypothesis. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 352-359.

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

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.

Shimamura, A. P. (1986). Priming effects in amnesia: Evidence for a dissociable memory function. Quarterly Journal of Experimental Psychology, 38A, 619-644.

Smith, S. M., Glenberg, A. x., & Bjork, R. A. (1978). Environmental context and human memory. Memory & Cognition, 6, 342-353.

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. Cognition and Instruction, 1(3), 341-358.

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.

Winstein, C.J., & Schmidt, R.A. (1990). Reduced frequency of knowledge of results enhances motor skill learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 677-691.