14

Structuring the Conditions of Training to Achieve Elite Performance: Reflections on Elite Training Programs and Related Themes in Chapters 10–13

ROBERT A. BJORK

There can be arguments as to what training revolution we are in the midst of currently (Chatham, Chapters 2 and 10; Hunt, Chapter 5), but what does not seem debatable is that our society now confronts social, technological, and cultural conditions that require a revolution in training. Personnel in industry, health, military, and the public sector have more to learn in less time. Increasingly sophisticated and rapidly changing electronic and other technologies impose additional training demands at the same time that they offer the potential to make training more effective and more portable. The chapters in this section focus on these problems from the perspective of the military, which has always tried to be at the cutting edge of research and knowledge of learning and mastery. The United States Army, for example, requested that the National Academy of Sciences commission a National Research Council Committee study on the techniques for the enhancement of human performance (Druckman & Bjork, 1991, 1994; Druckman & Swets, 1988).

From the perspective of the military, changes in technology, types of wars, and the social/cultural contexts of warfare require that training must continue after military personnel are deployed, including in combat zones. Such changes also require that an elite performer possess a high level of interpersonal and cultural skills and sensitivities, as well as a high level of traditional warfighter skills. To optimize training, given these changes and challenges, requires new perspectives and new technologies, as well as capitalizing on progress in the science of learning.

BASIC CONSIDERATIONS IN STRUCTURING TRAINING TO ACHIEVE ELITE PERFORMANCE

My charge is to comment on the four chapters that comprise the section of this volume on the training of elite performers in the armed forces. The following considerations provide a framework for my comments.

The Need to Distinguish Between Learning and Performance

One of the most time-honored distinctions in research on learning is the distinction between learning and performance. Basically, what we can observe during training is *performance*, whereas *learning* – conceived of as relatively permanent changes in cognitive, perceptual, and motor representations that will support long-term retention and transfer of knowledge and skills to the post-training contexts that are the target of training – must be inferred. The distinction is important because performance during training is often an unreliable guide to whether the desired learning has actually happened. Considerable learning can happen across periods when performance is not improving and, conversely, little or no learning can happen across periods when performance is improving markedly.

Learning Without Performance

That learning can happen without improvements in performance was demonstrated decades ago by both animal- and human-learning researchers. Latent-learning experiments with animals (e.g., Tolman & Honzik, 1930), for example, demonstrated that a period of free exploration, without systematic changes in performance, could nonetheless result in considerable learning as evidenced by an animal's performance once reinforcement was introduced. With humans, research on the acquisition of motor skills demonstrated that training trials structured so that a buildup of fatigue prevented any further improvement in performance nonetheless resulted in further learning, as evidenced by subsequent performance after fatigue had dissipated (see Adams & Reynolds, 1954).

Such findings led learning theorists – whatever their other arguments and differences – not only to distinguish, empirically, between learning and performance, but also to distinguish, theoretically, between *habit strength* (learning) and *momentary action potential* (performance), to use Hull's (1943) terms, for example, or between *habit strength* and *response strength*, to use Estes's (1955) terms. More recently, Elizabeth Bjork and I have resurrected such a distinction in the context of a "new theory of disuse" (Bjork & Bjork, 1992), which distinguishes between the *storage strength* of the memory representation corresponding to a skill or knowledge, versus the *retrieval strength* of that representation. Storage strength is assumed to reflect how entrenched or inter-associated the memory representation is with respect to related knowledge and skills that exist in memory, whereas retrieval strength is assumed to reflect the current activation or accessibility of that representation, which is a product of factors such as current situational cues and recency. Importantly, current performance is presumed to be solely a product of current retrieval strength, but storage strength acts to retard the loss of retrieval strength and enhance the gain of retrieval strength, given practice or further study.

Performance Without Learning

That the converse is true as well – namely, that little or no learning, as measured by long-term post-training retention or transfer, can result from training during which there are pronounced improvements in performance – is demonstrated by a wide range of more recent studies (for reviews, see Bjork, 1994, 1999 ; Christina & Bjork, 1991; Schmidt & Bjork, 1992). Manipulations such as blocking or massing practice on a given subtask of what is to be learned can result in rapid apparent learning, but not support post-training retention or transfer, whereas conditions that introduce difficulties and challenges for learners, slowing the rate of apparent learning, can optimize retention and transfer (see Shea & Morgan, 1979 ; Simon & Bjork, 2001). Such "desirable difficulties" (Bjork, 1994a, 1994b) include spacing rather than massing study sessions; interleaving rather than blocking practice on separate topics or tasks; varying how instructional materials and tasks are presented or illustrated; reducing feedback; and using tests rather than presentations as learning events.

It is important to emphasize, first, that many of the difficulties one can create for learners are undesirable during learning, at the end of learning, and forever after, and second, that desirable difficulties can become undesirable if such difficulties cannot be overcome. Such difficulties are desirable because responding to them engages processes that foster the understanding, linkages, and elaboration that support long-term retention and transfer, but they become undesirable if learners, by virtue of their prior learning and current motivation, are unable to overcome those difficulties. Having learners generate to-be-learned skills and knowledge, for example, versus presenting such knowledge and skills for observation and study, is a desirable difficulty, but if learners are not in a position to succeed at such generation, it can become an undesirable difficulty (see McNamara, Kintsch, Songer, & Kintsch, 1996). For that reason and perhaps others, it becomes a matter for research to determine whether and when such difficulties should be introduced. (See Wulf & Shea, 2002 , for evidence that the procedures for optimizing simple skills and complex skills may differ.)

The Potential for Instructors and Learners to Be Misled

That performance during training is an unreliable index of learning, as measured by post-training retention and transfer, has important implications. To the extent that instructors and trainees interpret current performance as learning, instructors become susceptible to choosing less effective conditions of training over more effective conditions, and trainees become susceptible to preferring – and evaluating more highly – poorer instruction over better instruction. Said differently, to the extent that current retrieval strength is assumed to reflect storage strength – that is, learning – instructors and learners are susceptible to being misled. As several of the chapters in this section stress, especially the chapters by Chatham (Chapter 10) and by Schreiber, Bennett, Colegrove, Portrey, Greschke, and Bell (Chapter 11), good assessment and measurement of performance is critical to the evaluation of training, but good measures of performance in the post-training contexts that are the target of training are usually lacking. Additionally, when such measures do exist, they may not get back to training personnel at all or not in a way that they can be related to a given trainee's conditions of training.

Without such post-training assessments and measures, it is natural to assume that performance *during* training indexes learning and provides a basis for distinguishing between and choosing among instructional practices. In fact, that assumption is so compelling that it is unlikely to even be questioned. As a consequence, however, the potential exists for training personnel to be shaped into choosing conditions of training that make performance improve rapidly and not choosing the kind of desirable difficulties that might enhance performance by military personnel when it matters most – that is, after trainees are deployed. The motivation to choose conditions that enhance trainees' performance – and apparent learning – is fueled by another consideration: Individuals responsible for the conditions of training are likely themselves to be evaluated by the performance of their trainees during training.

The problem can be exacerbated when trainees' evaluations of their ongoing training are used as a basis for choosing among alternative conditions of training. There is evidence from laboratory studies (see Kornell & Bjork, 2008; Simon & Bjork, 2001) that learners, too – in predicting their future performance on a post-training criterion test – are fooled by their own current performance, even when that performance is a product of local conditions, such as blocking and massing of practice, that result in poor long-term retention and transfer. Aside from the potential for trainees to evaluate less-than-optimal training very highly, training conditions that are made artificially easy – by keeping conditions of practice constant and predictable, for example, or by massing practice on a given subtask of a more complex task – can create illusions of comprehension and competence. As my colleagues and I have emphasized elsewhere (e.g., Bjork, 1999; Jacoby, Bjork, & Kelley, 1994), there is something far worse than individuals not possessing critical skills and knowledge – namely, not possessing such skills and knowledge, but thinking they do. Illusions of comprehension and competence become especially hazardous in settings, such as military settings, where mistakes and miscommunication can be so costly not only to one's self, but also to so many others.

The Potential to Misunderstand the Meaning and Role of Errors

As I emphasized in a final epilogue (Bjork, 1994b) to the third of three reports by a National Research Council Committee on Techniques for the Enhancement of Human Performance (Druckman & Bjork, 1994), a committee commissioned by the National Academy of Sciences on behalf of the United States

Army, one of the "institutional impediments to effective training" in military settings is a misunderstanding of the meaning and role of errors. To the extent that optimizing training requires introducing desirable difficulties and simulating, during training, the complexity, variety, and unpredictability of combat conditions, errors and mistakes by trainees during training will be frequent. Within military cultures, however, there remains a tendency to view errors and mistakes as evidence of a less-than-optimal training, rather than as opportunities for learning and a necessary component of maximally effective training. Expressions such as "we do it right the first time" and "we don't practice mistakes" reflect and contribute to that tendency.

The meaning of errors can be misunderstood, too. Within both the military and our society more generally, in my view, differences in innate ability between individuals tends to be over-appreciated, whereas the power of training, effort, and experience tends to be under-appreciated. As a consequence, differences in individual performance tend to be over-attributed to differences in innate ability – meaning that errors, rather than being interpreted as an opportunity for learning, are seen as evidence of the inadequacies of trainees, trainers, or both.

A final important consideration is that efforts to avoid errors and mistakes being made during training – by providing artificial supports, such as creating constant and predictable conditions; providing cues that would not be present in actual combat; encouraging imitation; and so forth – will *not* tend to eliminate such errors being made when such supports are absent. Rather, the effect can be to delay such errors being made until after training, possibly in circumstances where such errors really matter.

Top Gun, the National Training Center (NTC), and Similar Programs as Counterexamples

Having said that there is a "tendency" in the military to misunderstand the meaning and role of errors, I must now hasten to add that there are striking exceptions to that generalization. The fact that people learn by making and correcting mistakes is not only well understood by many individuals in the military, training programs such as the Navy's Top Gun program and Army's NTC have provided among the most compelling, real-world examples of that principle. General Paul Gorman, now retired, who served as a consultant to the National Research Council Committee on Techniques for the Enhancement of Human Performance (see Druckman & Bjork, 1994), and who was a principal architect of the NTC, provides an especially good example of a key individual who understands, rather than misunderstands, the meaning and role of errors during training. (For more on General Gorman, see Chatham, Chapter 2.)

After-Action Reviews

If people are to learn by making and correcting errors during training, then it is crucial that performance is measured in a rigorous and accurate way, that there be feedback – bottom up as well as top down – to individuals and units that is informative, and that there is an opportunity to discuss and analyze why errors were made. To the military's credit, after-action reviews (AARs) typically have all of those properties.

The Prevalence of Forgetting and Its Role in Relearning

Finally, before turning to comments on the individual chapters in this section, I cannot resist chiming in on the discussion (Chatham, Chapter 2; Hunt, Chapter 5) on the importance of understanding the nature of forgetting and the interaction of forgetting and learning. Chatham rightly emphasizes the importance of acknowledging that skills and knowledge, once apparently acquired, will often not remain accessible over time – that is, they will be forgotten – and often at a rapid rate. It is, in fact, a fundamental – and even adaptive – property of human memory that skills and knowledge, with disuse, become inaccessible.

Forgetting, of course, is often frustrating to us, but it is arguably as important to our day-to-day functioning as is remembering. To be efficient requires that we continually update procedures, facts, and episodes in our memories: We need to remember, for example, how the current version of a given software program works, or how it works on a new computer, not how the old program worked on our old computer; we need to remember current names and numbers, not the names and numbers they replaced; we need – as military personnel – to remember the details of a current military engagement or how a new piece of military hardware works, not the details of an earlier and similar engagement, or how the old hardware worked; we need to remember where we parked the car today, not yesterday or a week ago; and on and on. Any efficient information-processing system, whether living or non-living, needs some means to set aside or erase information that is no longer relevant and a possible source of confusion. In human memory, forgetting plays that role.

Research on forgetting, tracing back more than 120 years, has demonstrated that forgetting, rather than being a process of decay, analogous to footprints fading in the sand, is a consequence of interference and competition among memories (for a brief history of research on forgetting, see Bjork, 2003). Learning new information and using that information renders inaccessible information that is no longer being used. The acceleration of forgetting that accompanies disuse is an especially pertinent consideration in military contexts. Much of military training focuses on procedures, skills, and knowledge that – under typical conditions – are accessed only rarely and intermittently, such as in actual combat situations. Without systematic relearning and rehearsal procedures, much of what has been "learned" during military training will be inaccessible when needed.

The Importance of Overlearning and Relearning

Hunt (Chapter 5) is also right when he mentions that the efficiency of learning must be measured not simply by whether a learning curve has appeared to plateau, but also – or even especially – by the forgetting curve that follows various degrees of practice and training. As he mentions, overlearning procedures that appear to be inefficient, as measured by the small or nonexistent improvements in performance across overlearning trials, can be efficient as measured by the subsequent forgetting rate (e.g., Krueger, 1929; for a review, see Christina & Bjork, 1991). In terms of the distinction between storage strength and retrieval strength mentioned earlier, overlearning can continue to build storage strength – and, hence, slow the rate of subsequent forgetting – even though retrieval strength is no longer increasing, or increasing only slightly.

Forgetting as a Necessary Condition for Optimal Relearning

One of the misconceptions laypeople have about how they learn and remember, or fail to learn and remember, is that learning is a matter of building up something in memory, and forgetting is then a matter of losing some or all of what has been built up. In fact, a variety of conditions that increase forgetting (two examples being changing contextual cues or increasing a retention interval) also enhance learning. Thus, after instruction, individuals who are later tested in a context that differs from the context of instruction will tend to exhibit more forgetting than do individuals who are tested back in the context of instruction (see, e.g., Smith, 1988), but if both groups are re-instructed, rather than tested, the relationship tends to reverse (see Smith, Glenberg, & Bjork, 1978). Similarly, as exemplified by the well-known "spacing effect" (see Dempster, 1996; Glenberg, 1992), increasing the retention interval from an initial study episode to a test or second study episode decreases performance in the test condition, but increases the effectiveness of the repeated-study condition, as measured by long-term retention. Increasing the difficulty (Bjork & Allen, 1970) or similarity (Cuddy & Jacoby, 1982) can produce similar pattern.

Basically, conditions that induce forgetting, rather than undoing learning, create the opportunity for additional learning, over and beyond the level achieved earlier, when skills and knowledge are rehearsed or relearned. Forgetting, in effect, can enable learning. A discussion of *why* that is the case would take this commentary too far afield, but within the context of Bjork and Bjork's (1992) new theory of disuse it comes about because increments in storage strength (learning) are assumed to be smaller the higher the current retrieval strength of a given memory representation. Said differently, when some skill or knowledge is maximally accessible from memory, given local conditions, little or no learning results from additional instruction or practice.

Implications for Sustaining Elite Performance

From the perspective of traditional research on learning and skill acquisition, it is important to distinguish between the phase of training and its distinct environment from the subsequent phase of performance in a target environment. The problem for a training organization is to maximize performance when it matters, that is, *after* training and, especially, when individuals are deployed. To achieve that goal requires not only that initial training be designed in ways that foster durability and flexibility of to-be-learned skills and knowledge, but also that – following initial training – there be well-designed schedules of relearning and rehearsal of critical skills and knowledge.

More considerations go into such relearning being "well designed" than appropriate for me to discuss here, but two considerations merit emphasis. First, it is crucial for there to be a kind of task analysis that sorts needed skills and knowledge into those that are simpler versus those that are more complex and those that are, in the normal course of an individual's job, accessed more frequently and less frequently. The more complex the skill or knowledge in question and the less frequently it is accessed on a day-by-day, week-by-week, and month-by-month basis, the more it needs to be the target of relearning/refresher procedures. Second, it needs to be recognized that the potency of relearning procedures will be enhanced by such factors as the delay from initial learning and changes in situational and interpersonal contexts – that is, by factors that produce forgetting and enhance learning.

It must also be recognized, though, that in many – perhaps most – training situations in professional and military contexts, the line between the training phase and the after-training phase is blurred. Graduates from professional schools or military training facilities end up spending a long period in on-the-job-training before they become proficient (see Lajoie, Chapter 3) and some of them eventually reach expert levels of performance (see Ericsson, Chapter 18). During the on-the-job-training, the trainees engage in a majority of the representative daily activities and, thus, the period of delay (and the associated forgetting) between successive executions is often short.

SPECIFIC COMMENTS ON THE PRECEDING CHAPTERS

Quantifying Warfighter Performance in Distributed SimulationEnvironments

In this chapter, Schreiber, Bennett, Colegrove, Portrey, Greschke, and Bell (Chapter 11) provide a brief history of the use of simulators in flight training, which they point out has a history nearly as long as manned flight itself. Their primary focus is on the current state of distributed mission operations (DMO), technologically demanding simulations that provide learners an opportunity "to operate in a team-oriented, multiplayer environment, and thereby train higher-order and more varied skills than would normally be trained in a single-simulator environment."

The authors provide a compelling argument for the potential of distributed/team simulations, but they couple that argument with an even more compelling argument that quantitative measurements of the improvements in performance owing to such simulations is sorely lacking. Without measuring and quantifying performance and improvements in performance, they argue, we cannot test the actual effectiveness of such simulations, which are complex and expensive, nor do we have a basis for refining such simulations over time. They assert that "for networked simulation, subjective evaluations have been the foundation for large-scale aggregate analyses" and that "published works using objective, quantifiable metrics for empirical investigations of performance transfer from multiplayer networked environments to the actual environment appear, for DMO, to be nonexistent."

With respect to the "basic considerations" I provided as a framework for my comments, there are several points that merit emphasis. First, the authors' emphasis is on measuring *performance* – and mostly on measuring performance during simulation exercises. Given the unreliability of performance *during* training as an index of learning, as measured by post-training retention and transfer of skills and knowledge, there is a risk of developing simulations that maximize performance during training, but are not optimal for the post-training performance that is the target of training. To be fair to the authors, however, they emphasize that improvements in performance must be measured in objective, quantifiable ways both during DMO and "resulting from" DMO training. The following are other points that warrant emphasis.

Introducing Desirable Difficulties

A considerable virtue of distributed simulations, in my view, is that they have the potential to incorporate, in a natural way, a number of the conditions of learning and practice that I have labeled desirable difficulties. By the very fact that they are interactive, for example, they incorporate generation activities in a natural way – unless, of course, they are structured so that most of what trainees do is to observe and/or imitate. In addition, because they involve multiple players, the conditions of practice depend on the responses of other individuals, meaning that those conditions will incorporate variation, versus being constant and predictable.

Those intrinsic benefits of distributed simulations notwithstanding, however, I believe that certain types of desirable difficulties should be designed into such simulations. The emphasis by Schreiber et al. on "repetition" is important, for example, but it is important that such repetitions be spaced or interleaved, versus massed or blocked. The authors' emphasis on providing feedback seems well advised, too, but their emphasis on providing "immediate warfighter feedback" after each complex trial may not be optimal. In research on the learning of relatively simple motor skills (e.g., Schmidt & Bjork, 1992; Schmidt, Young, Swinnen, & Shapiro, 1989), providing intermittent or summarized feedback rather than constant/immediate feedback has been found to enhance long-term retention. Whether those findings would extend to more complex simulations remains to be seen, but I think it is non-controversial to say, given those findings, that research is needed to explore how the acquisition of different types of skills can be optimized by design about the amount, content, and timing of the feedback.

Engineering Fidelity Versus Psychological Fidelity

Designing distributed simulations in a way that even comes close to simulating the real-world dynamics they are designed to simulate is an impressive technological achievement. It also tends to be the case that each additional increment in such engineering fidelity tends to multiply the development costs of such simulations. From that perspective, it is important to remember that in research on simulators it has frequently been the case that low-tech simulators have often produced training outcomes that are as good or better than high-tech simulators. What is crucial in simulations is not engineering fidelity, per se, but the psychological fidelity of a simulation – that is, does it exercise the processes that will be needed later, when it matters? Early in training, for example, a simpler representation of the task and environment might be more effective than representing fully the complexities that characterize the actual real-world task. (For a review of the considerations in the context of optimizing transfer of training, see Reder & Klatzky, 1994.)

Learning Versus Relearning

Finally, it is important to emphasize that the optimal way to design such distributed simulations – or single-simulator environments, for that matter – is likely to depend on whether the goal is to optimize the original learning of skills and knowledge, or to refresh existing skills and knowledge. Stated in terms of the new-theory-of-disuse framework, when the problem is to build storage strength (learning), as in the case of original learning, the considerations are different than when storage strength already exists but retrieval strength has been lost owing to disuse. In the latter case, relearning will tend to be very rapid, practice can be more massed in time, and a high level of engineering fidelity may be very desirable.

Contrasting Submarine Specialty Training: Sonar and Fire Control

In this chapter, Kirschenbaum, McInnnis, and Correll (Chapter 12) focus on the training of the two largest groups of enlisted submarine specialists, sonar technicians (STs) and fire-control technicians (FTs). The authors assert and demonstrate that the skills necessary to excel in each specialty differ and, hence, that training and "performance metrics" must differ, as well, if training is to be optimized in each case. More specifically, they argue that STs must develop a high level of perceptual skills, whereas FTs must develop skills that are more cognitive in nature, including mastering a "set of tools, procedures, and algorithms." They describe and illustrate the tasks confronting each type of specialist; they characterize the unique environment in which submariners must work; they summarize the problems that confront current military training, such as inadequate time for training; and they contrast what they refer to as the "old model" of training with the new "training continuum model."

The focus of the authors' analysis is on two specialties in submarine training, but the implications of their analysis are quite general, in my opinion. The following are some observations.

The Old Training Model Versus the New Training Model

Among the problems the authors identify with the old training model are that inadequate time for training resulted in an emphasis on maintaining and using equipment, rather than on understanding concepts and processes; that when trainees went to sea the first time, they were not able to perform basic functions independently, meaning that extensive on-the-job training was required; and that such on-the-job training consumed the time of shipboard experts, meaning their time to focus on their own tasks was reduced.

The new training model reflects what the authors refer to as a revolution in training within the Navy. The components of the revolution include a careful and validated specification of the "skill objects" associated with the tasks that characterize a job within the Navy; a re-engineering of the training curriculum to focus on those tasks; and incorporating the curriculum within a "career path that takes a sailor from raw recruit to expert." The new "training continuum model" results in a shortening of the training before a sailor goes out to sea and includes individual computer-based instruction that can be completed both on shore and on ship. Another "radical change" is that sailors periodically leave the ship and return to the classroom for refresher and technology training stints of about two weeks' duration.

Learning Versus Performance

The basic goals of the new training program include "providing as good or better training in a shorter timeframe." Analyzed in terms of basic principles of learning and memory, the new program would seem to have many desirable characteristics. An early exposure to the realities of submarine life and tasks has to be good in terms of motivating and framing subsequent training; going back and forth between the classroom and the ship must introduce variation, spacing of practice, and an integration of higher-order and lower-order learning and concepts; and the interspersing of individualized computer-based training seems both efficient and potentially effective.

But is the new program an improvement? According to the authors, an assessment is ongoing, but documenting that the new model is better in any rigorous way "cannot be evaluated" because performance measurements are not, for some reason, comparable between the two models, a fact that again supports the arguments of Schreiber et al. that agreed-to and quantifiable measures of performance are essential if programs of training are to be evaluated and optimized. From the description given by Kirschenbaum et al., it also seems to be the case that the ongoing assessment rests heavily on measures of performance *during* training and on sailors' "perception of the course," both of which are, as I emphasized earlier, likely to be unreliable indicators of whether the long-term learning that is the target of training has been achieved. One of the measures included, however, is a "follow-up evaluation by the student's on–the-job supervisor." Although perhaps not as valuable as rigorous measures of actual on-the-job performance, this measure should at least reflect the longer-term consequences of training.

Career-Long Learning

The description and analysis by Kirschenbaum et al. provide a nice example of a trend in military training that is likely to become pervasive. It may have been true many years ago – though perhaps it was never entirely true – that military training largely preceded and was isolated from individuals actually going to war. If that were ever true, it clearly is no longer true. Because training capabilities are more transportable than ever before, because equipment and technologies are changing more rapidly than ever before, and because new types and contexts of combat pose new challenges, the training of military personnel needs to be viewed as a career-long process. Continuing medical education is an interesting example of how the domain of medicine has developed methods for maintaining and advancing the skills of its practitioners (Davis, Chapter 8).

Training Complex Cognitive Skills: A Theme-Based Approach

In their chapter, Shadrick and Lussier (Chapter 13) provide a report on a program designed to train adaptive thinking skills. They argue that being able to think adaptively, which has always been important in military conflicts, has become more important than ever given the complexities and non-traditional nature of recent conflicts, such as the current war in Iraq. Dimensions of adaptive thinking include being able to respond effectively when confronted with unanticipated situations and responding effectively to rapidly changing conditions when multiple events and inputs compete for one's attention.

The authors refer to their approach to training thinking as "theme based." Broadly, the idea is to identify the cognitive behaviors and habits – or themes – that contribute to flexible thinking and problem solving. Examples might be learning to adopt a different perspective or modeling how an enemy is thinking and how that thinking might be changing as a function of battlefield events. The approach is embodied in a Think Like a Commander (TLAC) training program that, among other things, incorporates deliberate-practice (Ericsson, Krampe, & Tesch-Römer, 1993) techniques in an effort to develop expert mental models and understanding of presented tactical scenarios. In an analogy with how chess masters analyze chess positions from games and then obtain feedback on the correctness of their move selection and their analyses, Shadrick and Lussier assume that military officers can be trained to think by being presented with tactical situations that require them to analyze those situations before taking action and then, later, to obtain feedback on their analyses and thoughts.

To the authors' great credit, they have evaluated the TLAC training program not via testimonials or some kind of face validity, but by well-designed research. The results of the four experiments they report not only provide strong support for the effectiveness of the TLAC program, but are also interesting and provocative with respect to some broader considerations in training. The following are some observations.

The Generality of Training Principles

The authors argue that traditional training methods, such as those applied to the learning of motor skills and verbal and spatial information, can also be applied to train thinking skills. In contrast to other opinions in the literature, the authors do not think that training adaptability and adaptive thinking require training paradigms that are uniquely suited to higher-order cognitions and to people being able to perform appropriately in novel or ambiguous situations.

I, personally, share the authors' opinion. I believe the principles and methods that optimize learning – such as that generating information and skills leads to better, later production of that information or those skills than does studying and observing such information or skills – are very general indeed. My opinion in that regard has been shaped by a range of experimental findings, but also by an experience co-teaching a graduate seminar with my colleague, Richard Schmidt. The goal of the seminar was to identify how the techniques that optimize the learning of motor skills (his area) differ from the techniques that optimize the learning of verbal and conceptual information (my area). Somewhat to our surprise, we were able to identify far more principles that were common to motor and verbal learning than we could unique principles, which led us to write an article entitled "New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training" (Schmidt & Bjork, 1992).

The Interaction of Experience and Training

Among the most striking aspects of the findings the authors report is that the effectiveness of TLAC training interacts with whether participants in their studies did or did not have deployment experience in Iraq or Afghanistan.

The training program was effective for both groups, but having been deployed increased the effectiveness of the training for all ranks – from lieutenants, to captains, to majors, to lieutenant colonels. The authors' findings in that respect may illustrate a broad and important generalization with respect to military training: Whenever it proves feasible to have personnel go back and forth between deployment settings and training settings, the effectiveness of training is likely to be enhanced. One can think of motivational reasons why such a generalization may hold, but there may be other reasons as well having to do with deployment experiences yielding an interpretive framework for the content of training. The new program for training submarine specialists, described by Kirschenbaum et al., may profit from that standpoint by interleaving classroom training ashore with training aboard ship.

Deliberate Practice and Learning Versus Performance Revisited

The TLAC program includes deliberate practice as a component of training (see Ericsson , Chapter 18). On the surface, aspects of deliberate practice, especially practicing one tactical scenario after another with analysis and feedback after each scenario, appear to be at odds with my earlier arguments that one should introduce desirable difficulties, such as spacing versus massing, and intermittent or summarized feedback versus immediate/continuous feedback, to enhance learning. Ericsson, however, would – and has – argued that sampling scenarios from a large collection of challenging tactical situations introduces variation and spacing of key concepts in an intrinsic, natural way.

Aside from any arguments as to whether deliberate-practice routines could be made even more effective by a planned introduction of spacing, variation, and summarized feedback, it is important to emphasize that if the focus is on performing at one's best when it matters (as would often be the goal, say, in a high-level athletic competition), then my earlier assertions need to be modified. That is, if the goal is not learning, but preparing one's self to perform some learned skill or procedure at the highest possible level, the considerations are different. There may, for example, be an important role for massed repetition, partly to keep a skill or routine maximally activated, but also to build confidence. Continuous feedback from a coach or expert may be important, too, once a skill has been acquired, if the fine-grain adjustments necessary for maximal performance are to be made.

From the standpoint of manipulating confidence and motivation, it can also be important, in some circumstances, to start with practice conditions that are *not* optimal with respect to long-term learning, such as massed practice and constant conditions, and then gradually withdraw such crutches and introduce desirable difficulties. Such a combination has the virtue of encouraging the learner via inducing early rapid improvement while still incorporating conditions of instruction that can be expected to foster learning (storage strength) versus performance (retrieval strength).

Toward a Second Training Revolution: Promise and Pitfalls of Digital Experiential Training

In his chapter on digital experiential training, Chatham (Chapter 10) argues strongly, if with occasional caveats, for the training potential of "light-weight digital simulations" – that is, training on personal computers (PCs). He sees the potential of the lessons learned from what he refers to as the "first training revolution," which led to training programs and settings such as the NTC, to be incorporated into PC training. As a kind of background for his comments, he observes that one-on-one tutoring can be remarkably effective, but is impossibly expensive and impractical for other reasons as well; that the state of assessment and measurement of training results is "abysmal" (he adds the cynical comment, "what you can't measure, you can ignore"); and that proximity in time between training and experience in the context in which that training is needed is one key to making training maximally effective.

Chatham argues that "grand-scale, high-fidelity simulations" can, at their best, be superb trainers, but they are very expensive, have fixed sites, and tend to serve specialized needs of some subset of military personnel. By contrast, he sees lighter weight, PC-based simulations as being accessible to "most of the military" who are in need of training "most of the time." In the remainder of his chapter, Chatham explores the potential of existing and future computer games as training platforms. In that analysis, he draws on examples such as *DARWARS Ambush!, Operation Flashpoint,* and *America's Army.*

Games as Tools for Training

The success of PC games, whether measured by their global market, their increasing sophistication, or by the hours that people, especially young people who otherwise have short attention spans, spend immersed in such games, has suggested to many individuals – including individuals at military training sites, such as Fort Monroe, and at non-military sites, such as the Institute for Education Sciences – that such games might be harnessed for science learning and other types of learning. Chatham argues that games may or may not be effective and that any game that is "unexamined" – that is, not tested for the training gains it produces – is an "empty trainer."

In that assessment, I wholeheartedly concur. Games as possible tools for learning often seem to be evaluated – and positively – in terms of their novelty, their whiz-bang technological features, and their attractiveness to users. The truth, though, is that – as tools – they can be used poorly or well. If a game is not structured to exercise the processes that support comprehension, understanding, and the production of to-be-learned procedures or knowledge, it will not be effective as a learning platform, however glitzy and attractive it may seem.

Performance Gains Versus Learning Gains
I am inclined to give the author the benefit of the doubt, because he refers to the "results" of training in one or two places, but his emphasis on "performance gains" as the criterion by which simulations should be evaluated, suggests that he is referring to performance across and during simulation sessions. If so, then my earlier cautions with respect to performance during training being misleading as an index of learning are relevant to his analysis. If versions of the simulation are evaluated by rate of improvement across experience with the simulation, as opposed to long-term post-training retention and transfer, then designers of the simulation will be susceptible to including massing, predictability, and other conditions that act to prop up current performance but not support learning.

User Authoring, Generation Effects, and Cooperative Learning
The author cites *user authoring* as a very important factor in the success of the *DARWARS Ambush!* program. Because the program could be altered, students' experience was not limited to the initial 24 scenarios, and the author cites this property as a reason that students did not lose interest in the simulation. Students could both change scenarios and create new scenarios. I am sure user authoring was a very important feature in exactly that way, but it was probably also important because doing such authoring constitutes a kind of generation that no doubt had a powerful learning effect. That is, the kind of learning that results from modifying or creating a scenario is going to be powerful and long lasting, far more so than simply playing through an existing scenario.

Finally, the author mentions – in the context of the Tactical Language and Culture trainer – that the simulation, "while usable by the highly motivated individual, delivered much more useful training when done in a group setting." More specifically, the author mentions that, to his surprise, a pair of novice students working at the same computer seemed to gain more from the simulation than did students who had a computer to themselves. Such a finding, which is consistent with the voluminous evidence that cooperative learning, structured appropriately, can enhance individual learning (e.g., Aronson, 1978; Dansereau, 1987; for a review, see Druckman & Bjork, 1994), suggests it could well be profitable, in simulations designed to train individual skills, to design the games as a two-person interaction.

CONCLUSION

The dominant reaction I have after working through the chapters in this section is optimism. The training challenges faced by our society – in educational, medical, and industrial contexts, as well as military contexts – are daunting in their complexity, their novelty, and their sheer scope, but the chapters in this section illustrate that new technologies, informed by the science of learning, have unprecedented potential to create and sustain elite performers.

REFERENCES

Adams, J. A., & Reynolds, B. (1954). Effect of shift in distribution of practice conditions following interpolated rest. *Journal of Experimental Psychology*, *47*, 32–36.
Aronson, E. (1978). *The jigsaw classroom*. Beverly Hills, CA: Sage.
Bjork, R. A. (1994a). 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.
(1994b). Institutional impediments to effective training. In D. Druckman & R. A. Bjork (Eds.), *Learning, remembering, believing: Enhancing human performance* (pp. 295–306) Washington, DC: National Academy Press.
(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.
(2003). Interference and forgetting. In J. H. Byrne (Ed.), *Encyclopedia of learning and memory* (2nd ed., pp. 268–273). New York: Macmillan Reference USA.
Bjork, R. A., & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior*, *9*, 567–572.
Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35–67). Hillsdale, NJ: Lawrence Erlbaum Associates.
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.
Cuddy, L. J., & Jacoby, L. L. (1982). When forgetting helps memory: Analysis of repetition effects. *Verbal Learning and Verbal Behavior*, *21*, 451–467.
Dansereau, D. F. (1987). Transfer from cooperative to individual studying. *Journal of Reading*, *30*(7), 614–619.
Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork & R. A. Bjork (Eds.), *Memory:* Vol. 10. *Handbook of Perception and Cognition* (pp. 317–344). New York: Academic Press.
Druckman, D., & Bjork, R. A. (Eds.). (1991). *In the mind's eye: Enhancing human performance*. Washington, DC: National Academy Press.
(1994). *Learning, remembering, believing: Enhancing human performance*. Washington, DC: National Academy Press.
Druckman, D., & Swets, J. A. (Eds.). (1988). *Enhancing human performance: Issues, theories, and techniques*. Washington, DC: National Academy Press.
Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*, 363–406.
Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review*, *62*, 369–377.
Glenberg, A. M. (1992). Distributed practice effects. In L. R. Squire (Ed.), *Encyclopedia of learning and memory* (pp. 138–142). New York: Macmillan.
Hull, C. L. (1943). *The principles of behavior.* New York: Appleton-Century-Crofts.
Jacoby, L. L., Bjork, R. A., & Kelley, C. M. (1994). Illusions of comprehensions and competence. In D. Druckman & R. A. Bjork (Eds.), *Learning, remembering, believing: Enhancing human performance* (pp. 57–80). Washington, DC: National Academy Press.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science*, *19*, 585–592.
Krueger, W. C. F. (1929). The effect of overlearning on retention. *Journal of Experimental Psychology*, *12*, 71–78.
McNamara, D. S., Kintsch, E., Songer, N., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*(1), 1–43.
Reder, L. M., & Klatzky, R. L. (1994). Transfer: Training for performance. In D. Druckman & R. A. Bjork (Eds.), *Learning, remembering, believing: Enhancing human performance* (pp. 25–56). Washington, DC: National Academy Press.
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.
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.
Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 907–912.
Smith, S. M. (1988). Environmental context-dependent memory. In D. M. Thomson & G. M. Davies (Eds.), *Memory in context: Context in memory* (pp. 13–33). New York: Wiley.
Smith, S. M., Glenberg, A. M., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*, 342–353.
Tolman, E. C., & Honzik, C. H. (1930). Introduction and removal of reward and maze performance of rats. *University of California Publications in Psychology*, *4*, 257–275.
Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic Bulletin & Review*, *9*(2), 185–211.