Shimamura, A. P., & Squire, L. R. (1986). Memory and metamemory: A study of the feeling-of-knowing phenomenon in amnesic patients. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 452–460.
Uttal, W. R. (2001). *The new phrenology: The limits of localizing cognitive processes in the brain.* Cambridge, MA: MIT Press.
Van Overschelde, J. P., & Nelson, T. (2006). Delayed judgments of learning cause both a decrease in absolute accuracy (calibration) and increase in relative accuracy (resolution). *Memory and Cognition, 34,* 1527–1538.
Wade, C. (2006). Some cautions about jumping on the brain-scan bandwagon. *The Observer, 19,* 23.
Wager, T. D. (2006). Do we need to study the brain to understand the mind? *The Observer, 19,* 25–27.
Wager, T. D., Hernandez, L., Jonides, J., & Lindquist, M. (2007). Elements of functional neuroimaging. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (4th ed., pp. 19–55). Cambridge: Cambridge University Press.
Widner, R. L., Otani, H., & Winkelman, S. E. (2005). Tip-of-the-tongue experiences are not merely strong feeling-of-knowing experiences. *Journal of General Psychology, 132,* 392–407.
Widner, R. L., Smith, S. M., & Graziano, W. G. (1996). The effects of demand characteristics on the reporting of tip-of-the-tongue and feeling-of-knowing states. *American Journal of Psychology, 109,* 525–538.
Wilson, T. D. (1994). The proper protocol: Validity and completeness of verbal reports. *Psychological Science, 5,* 249–254.
Yarmey, A. D. (1973). I recognize your face but I can't remember your name: Further evidence on the tip-of-the-tongue phenomenon. *Memory and Cognition, 1,* 287–290.

17

LEARNING FROM THE CONSEQUENCES OF RETRIEVAL

Another Test Effect

Elizabeth Ligon Bjork, Benjamin C. Storm, and Patricia A. deWinstanley

OVERVIEW

In the present chapter, we describe research that integrates two lines of investigation in which Robert A. Bjork, in whose honor this volume is dedicated, has played a significant leadership role: research regarding the benefits of testing for long-term retention and research regarding the metacognitive processes underlying the development of illusions of competence during study. In our conjoining of these two areas of research, we have explored whether the metacognitive experience gained from taking a test can lead learners to adopt improved strategies for encoding future to-be-learned material. Or, put slightly differently, we have explored whether learning to learn can be another beneficial effect of test taking.

In the sections to follow, we first describe some of the more relevant findings regarding the beneficial effects of testing on memory, with this section followed by one in which we describe findings relevant to the question of how individuals monitor their learning during study. Then, we turn to a description of the research in which we have explored whether learning to learn can occur as a consequence of the testing experience.

THE POWER OF TESTS AS LEARNING EVENTS

Although to many educators and learners alike, tests are simply measures of what is already known; tests can, in fact, also serve as powerful learning events. Indeed, much laboratory research (e.g., Landauer & Bjork, 1978; Carrier & Pashler, 1992) has demonstrated the power of tests as learning events and, moreover, has shown that a test, even when no corrective feedback is given, can be considerably more effective for the long-term retention of material than additional study of it. This power of tests as learning events comes about because, as pointed out by R. A. Bjork (1975), retrieval processes do not simply assess the contents of memory and then leave the representations of items in memory in the same state as they were before being retrieved. Rather, the act of retrieving information modifies its representation in memory such that it becomes more recallable in the future—a phenomenon that Bjork argued represents a kind of Heisenberg principle for retrieval processes: "an item can seldom, if ever, be retrieved from memory without modifying the representation of that item in significant ways" (p. 123). Or, as similarly observed by Roediger and Karpicke (2006a) in their excellent review of testing effects, "just as measuring the position of an electron changes that position, so the act of retrieving information from memory changes the mnemonic representation underlying retrieval—and enhances later retention of the tested information" (p. 182).

That tests can be more powerful as learning events for the long-term retention of material than additional opportunities for studying it, even when the tests do not provide any corrective feedback, was recently and impressively demonstrated by Roediger and Karpicke (2006b) using both educationally realistic materials and ecologically significant retention intervals. To illustrate, in one of their reported experiments, participants in one condition were given four 5-minute opportunities to study the same passage (resulting in the passage being read an average of 14.2 times by these participants), while participants in another condition were given only one 5-minute opportunity to study the passage and then were tested on it three times in a row (resulting in the passage being read an average of only 3.4 times for these participants). Additionally, participants in these two conditions were given a final test on the passage at either a five-minute or a one-week delay. Although at the five-minute delay, participants who had received four opportunities to study the passage recalled significantly more idea units than did those who had only studied the passage once and been tested on it three times, at the one-week delay, this difference in recall was dramatically reversed. Now, the participants who had studied the passage only once but who had been tested on it three times recalled significantly more idea units than those who had studied the passage four times. Additionally, an analysis of the proportion of information forgotten across the one-week delay revealed that the participants who repeatedly studied the material forgot far more (52%) than did the participants who were repeatedly tested on it (only 14%), demonstrating the power of tests to retard the forgetting that would otherwise occur.

In addition to such direct effects of testing for the improvement of retention, tests can also enhance learning in several indirect ways. First, when given frequently across a course, tests can lead students to space their studying across the course rather than massing it right before a final exam, thus inducing the oft-demonstrated power of spaced or distributed practice for long-term retention (e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Dempster, 1996). Second, tests can serve to optimize future study activities. From a metacognitive standpoint, tests allow learners to make more accurate assessments than do additional study events concerning whether information is likely to be recallable in the future (e.g., Koriat & Bjork, 2005, 2006; Nelson & Dunlosky, 1991). Furthermore, when feedback is provided after a test, students are informed as to what they do and do not already know and can thus more efficiently allocate their future study efforts. Third, as suggested by the work of Izawa (1970), tests can enhance the effectiveness of subsequent study relative to the effectiveness of such study when not preceded by a test. Finally, as illustrated in several recent studies by Kornell, Hays, and Bjork (2009) using materials that ensured unsuccessful initial retrieval attempts, even failed tests can potentiate the effectiveness of subsequent study opportunities.

MONITORING ONE'S LEARNING DURING STUDY

Much recent research in the area of memory and learning has been concerned with the study of the metacognitive processes by which individuals monitor their level of knowledge during study (e.g., Benjamin, Bjork, & Schwartz, 1997; R. A. Bjork, 1999; Dunlosky & Nelson, 1994; Koriat, 1997, 1998). Undoubtedly, some of the interest in this area of research stems from the possibility, as put forth by R. A. Bjork and others (e.g., R. A. Bjork, 1999; Jacoby, Bjork, & Kelley, 1994), that the readings learners take on their level of comprehension during study, or their judgments of how likely they are to be able to recall the material being studied in the future, are as important as their actual comprehension and degree of learning because such readings play a powerful role in determining how they will decide to allocate their future study activities and learning resources. On the basis of such readings, for example, students may well decide to read one chapter versus another or to study one set of materials versus another in preparation for an upcoming examination.

A primary method that has been used to study such metacognitive processes is to ask learners to make judgments of learning (JOLs) during acquisition. A typically used procedure, for example, is to present learners with a list of cue-target pairs to learn and, following the presentation of each pair for study, to ask learners to judge the likelihood of their remembering the target in response to presentation of the cue alone on a later retention test. In a number of experiments using such a procedure, the JOLs made by participants have been found to be moderately accurate (e.g., Dunlosky & Nelson, 1994; Lovelace, 1984; Mazzoni & Nelson, 1995), and considerable research in this area has thus been focused on the question of what accounts for the accuracy of JOLs in predicting future memory performance.

Learners, however, can also be far from accurate in taking such readings of comprehension or in making JOLs, and thus other research in this area has been directed to the question of what accounts for such illusions of comprehension (e.g., R. A. Bjork, 1999; Jacoby, Bjork, & Kelley, 1994). Learners, for example, can be led to think that their level of comprehension or skill is greater than it actually is owing to conditions of learning (such as massed practice) that enhance or support performance during study or training, but actually impair long-term retention or transfer (e.g., Simon & Bjork, 2001). Similarly, learners can be led to make JOLs that perfectly mismatch their later performance on a test by basing them on the fluency with which they can retrieve answers from long-term memory in the presence of cues available at the time of study, but that will not be present at the time of test (Benjamin et al., 1998).

One account for the occurrence of such dissociations between the JOLs made by learners during study and their actual performance on a later test is offered by the new theory of disuse (NTD), a theory proposed by R. A. Bjork and E. L. Bjork in 1992 to account for a number of unique characteristics or peculiarities of human memory. According to the NTD, such dissociations occur when learners base the JOLs they make during study on retrieval strength (i.e., the current activation or accessibility of an item's representation in memory) rather than on storage strength (i.e., how entrenched or interassociated a memory representation is with related knowledge and skills), and because the former is a poor indicator of actual learning, it is also a poor indicator of long-term performance. (For more information concerning the new theory of disuse as a model of learning and memory as well as its applications for training and instruction, the reader is also referred to R. A. Bjork & E. L. Bjork, 2006; E. L. Bjork & R. A. Bjork, 2011).

From a slightly different perspective, Koriat (1997) has argued that learners can suffer from illusions of competence and be led to make inaccurate JOLs because, during original study, they tend to be relatively insensitive to the presence of extrinsic factors, which entail both conditions of learning (such as number of repetitions, presentation durations, and massed vs. distributed repetitions of items) and encoding operations used by the learner (such as level of processing and interactive imagery) that do enhance learning and later performance on a test, while being overly sensitive to intrinsic factors (such as the perceived association between cues and targets when both are present during study, or the relative difficulty or imagery values of individual words in a list) that do not necessarily enhance performance on later retention tests.

Indeed, in several recent studies, Koriat and Bjork (2005) have demonstrated the oversensitivity of learners to the intrinsic factor of the perceived relationship between cues and targets when both are present during study by manipulating the types of associations existing between cue-target pairs to which learners must make JOLs during acquisition. In one of their studies, for example, learners were asked to make JOLs during study of a list of cue-target pairs in which some of the pairs were high in a priori relatedness (i.e., the preexperimental likelihood that presentation of the cue alone would bring the target to mind rather than other possible associates of the cue) and some were what the authors referred to as purely a posteriori pairs (i.e., pairs for which the perceived relationship between the cue and target is high when presented together, but for which the a priori degree of relatedness is low—that is, when presented alone, the cue would be very unlikely to bring that particular target to mind vs. other possible associates of the cue). The JOLs learners made to the a priori cue-target pairs corresponded to their later performance; in contrast, the JOLs they made to the purely a posteriori pairs were highly inflated.

Koriat and Bjork (2005, 2006) have interpreted this overconfidence, or illusion of knowing, produced by the a posteriori pairs as being a type of foresight bias akin to, but different from, hindsight bias (Fischhoff, 1975; Hawkins & Hastie, 1990). More specifically, they proposed this foresight bias to be a mirror image of hindsight bias. That is, whereas hindsight bias refers to our tendency to distort our memory of a previously made judgment once the answer is known to us, the foresight bias proposed by Koriat and Bjork occurs when we predict our future success in recalling a correct answer in the presence of that answer. Thus, the authors argue, both represent biases that reflect our inability to "escape" the influence of the correct answer.

LEARNING FROM THE EXPERIENCE OF TEST TAKING

We turn now to a discussion of research conducted by the present authors that we believe integrates some of the findings and concepts from the research on testing effects and metacognitive processes described in the previous two sections. Because our research has focused on the encoding strategy of generation as one of the conditions of learning to which learners do not seem sensitive, we first define and illustrate the generation effect or advantage, including brief descriptions of two successful accounts of it. We then discuss a series of our studies, some complete and some ongoing, that address the general issue of the sensitivity of learners to the memorial benefits of generation and whether—if made sensitive to this benefit in the context of a test—they would then adopt more effective encoding strategies in the processing of new information.

GENERATION AS A CONDITION OF LEARNING

When learners take an active part in generating the information they are to learn, as opposed to having that information provided to them intact and simply reading it, they tend to remember the to-be-learned information better. To illustrate, if learners are required to generate the word *memory* from a word fragment (e.g., *m _ m _ ry*) versus being given the intact word to read, they will recall the word *memory* better on a later test. Or, if required to generate an exemplar, say *lemon*, to a category-plus-letter-stem cue (e.g., fruit–le___) versus being given the intact pair, *fruit–lemon*, to study, they will recall *lemon* better in response to the cue *fruit* on a later test.

Considerable research has shown this memorial benefit of generation (e.g., Jacoby, 1978; Slamecka & Graf, 1978) to be both a robust phenomenon and one that extends to a variety of materials, including lists of words, word pairs, and trivia questions (e.g., deWinstanley, 1995; Hirshman & Bjork, 1988) as well as mathematical problems (e.g., McNamara & Healy, 1995a, 1995b; Pesta, Sanders, & Murphy, 1999). The generation advantage, however, can also be diminished or even eliminated under certain conditions. For example, McNamara and Healy (1995a, 1995b) found that generation advantages do not occur on a later test for arithmetic problems unless retrieval strategies that were employed by learners during study are evoked again during the test. Similarly, deWinstanley, Bjork, and Bjork (1996) demonstrated that even when participants are learning the same materials, generation advantages may or may not occur depending on the match between the information strengthened as a result of the learner performing the generation task during study and the type of information required for optimal performance on a later test.

More specifically, deWinstanley et al. (1996) manipulated the conditions of learning during study so as to force the processing of different types of information in order to generate targets for the same list of cue-target pairs. To illustrate, in one condition, the cue-target pairs to be learned were blocked on the basis of the categorical membership of the targets, leading participants to focus on target-target-relational information (to which free-recall tests are assumed to be most sensitive)—as opposed to cue-target-relational information (to which cued-recall tests are assumed to be most sensitive)—in order to perform the generation task. On subsequent tests, these participants showed a generation advantage when given a free-recall test, but not when given a cued-recall test. In a second condition, the pairs to be learned were not blocked by category membership of the targets, essentially eliminating target-target-relational processing as a basis for generating targets and forcing participants to rely on cue-target-relational information instead. On subsequent tests, these participants showed a generation advantage on a cued-recall test, but not on a free-recall test. In short, a striking reversal was observed in the relative levels of free and cued recall for targets that had been generated versus read, depending on the type of information participants had been forced to use in order to generate those targets during learning.

The occurrence of generation effects can also be influenced by the encoding instruction given to learners. Begg, Vinski, Frankovich, and Holgate (1991), for example, showed that the advantage of generation over reading could be eliminated when participants were given other effective strategies, such as imagery, to use when encoding intact to-be-read items. Similarly, deWinstanley and Bjork (1997) eliminated a previously observed generation advantage for identical materials by giving participants explicit instructions concerning the type of retention test to expect and how to process information optimally in anticipation of such a test. In fact, changes in a variety of factors—such as the type of test learners expect, whether to-be-read or to-be-generated items are mixed together (i.e., between- or within-subject manipulations of generation vs. read), and the specific requirements of the generation task—have led to a continuum of outcomes ranging from large to small to no generation advantages.

These types of findings, delineating conditions under which generation advantages do and do not occur, are largely consistent with two explanations of the generation effect: the procedural account and the transfer-appropriate multifactor account, both of which emphasize the critical nature of the relationship between encoding and retrieval processes in the production of generation effects. Briefly characterized, the procedural account (Crutcher & Healy, 1989; McNamara & Healy, 1995a, 1995b) assumes that when learners are required to generate information at study, as opposed to reading it, they are more likely to utilize encoding procedures that can then be reinstated during a later retention test. When a later test does invoke such procedures, a generation advantage should occur; if not, a generation advantage should not occur. The transfer-appropriate multifactor account (deWinstanley et al., 1996)—built upon the two-factor account of Hirshman and Bjork (1988) and the multifactor account of McDaniel, Waddill, and Einstein (1988)—assumes that the act of generation strengthens whatever type of information is used by the learner to complete the generation task, and thus the consequence of the generation task for later memory performance depends on whether the information so enhanced is information to which a later test is sensitive. Thus, when there is a good match between these types of information, generation advantages should occur; when there is not, generation advantages should not occur.

MAKING LEARNERS SENSITIVE TO GENERATION AS AN EFFECTIVE CONDITION OF LEARNING

As indicated in our discussion of the research by Koriat, R. A. Bjork, and others on metacognitive processes, investigations of how individuals monitor their level of learning during study paint a picture of learners as being insensitive to many of the conditions of learning that can enhance long-term retention, reflected in the relatively small influence of such variables on the JOLs they make during study. Such conditions, referred to as extrinsic factors by Koriat (1997), include not only aspects of presentation, such as number and duration of study opportunities, but also encoding operations applied by the learner during study, such as generation and levels of processing. In our research, we have asked whether learners could be made sensitive to the encoding effectiveness of one such extrinsic factor—that of generation—if they were to experience its memorial consequences in their own recall performance in the context of a test. We did this not to see if the accuracy of their

JOLs might thereby be enhanced during future study (although that is a topic of other research under way), but to see if their encoding strategies might thereby be enhanced during future study.

In our research addressing this general issue, we have typically adopted the following general experimental strategy, initially used in the studies conducted by deWinstanley and Bjork (2004). Participants were first presented with a short passage to study of the type that would appear in an undergraduate introductory textbook, but in which we had embedded both to-be-generated and to-be-read critical items. Next, participants' recall for these critical items was assessed in a fill-in-the-blank test. Then, after the experience of this test, a new text passage, also containing both to-be-generated and to-be-read critical items, was presented for study and then also followed by the same type of test for the critical items. Thus, before presentation of the second text passage for study, participants would have the opportunity to engage in both generating and reading of critical items in a previous passage as well as the opportunity to experience a generation advantage in their own performance on the test of those items. Hence, if as hypothesized, such an experience could be sufficient to induce participants to adopt a more effective way of encoding future to-be-read information, a generation advantage should be attenuated, or possibly eliminated, in the test of the second passage.

In the first two studies employing this procedure, deWinstanley and Bjork (2004) obtained results consistent with this hypothesis. Specifically, while a generation advantage was observed in the test of the first passage, no generation advantage was observed in the test of the second passage. Importantly, however, the absence of a generation advantage on the second test did not occur at the expense of the generated items. Instead, recall of the to-be-read items presented in the second passage improved to the level of that for the to-be-generated items, which did not differ from the level obtained in the test of the first passage.

Given the results of these studies, deWinstanley and Bjork (2004) next attempted to determine whether it was, in fact, the testing experience that was critical in leading participants to develop more effective encoding strategies in two follow-up studies. In their first follow-up study, they used the same basic procedure, but rather than presenting both to-be-generated and to-be-read items within the same passage, they manipulated the requirement to generate versus read between passages. Accordingly, in the first passage for a given participant, the encoding task for all critical items was the same—either generating or reading—and then, in the second passage, the encoding task for critical items was switched. Consequently, participants did not have the opportunity to experience the memorial consequences of generating versus reading within the context of the same test before they were presented with a new passage for study. Thus, if such an experience is critical for leading learners to adopt more effective encoding strategies, then the generation advantage should not be eliminated on the test of the second passage, and indeed, this was the result observed: A generation advantage was obtained on both tests, and furthermore, its size did not differ across tests. Apparently, then, simply having the experience of encoding critical items via generation in the first passage was not sufficient to make participants aware of the need to develop a better processing strategy for encoding to-be-read critical items in the second passage, pointing to the critical role of experiencing the relative memorial consequences of the two types of encoding within the same testing episode.

In their second follow-up study addressing the critical nature of the testing experience, deWinstanley and Bjork (2004) examined whether something less specific—like a general dissatisfaction with the number of critical to-be-read items they were able to recall in the first test—had led participants to process to-be-read items in the second passage more effectively. This possibility could not be ruled out by the first follow-up study because the switch of encoding tasks between passages made it impossible for participants presented with only to-be-read items in the first passage to reveal any such improved encoding strategies for subsequent to-be-read items, as they only received to-be-generated critical items in their second passage. Thus, to address this possibility, they next manipulated the requirement to generate versus read between participants rather than between passages.

Given this way of manipulating the encoding variable, a generation advantage would be expected on the test of the first passage whichever hypothesis was correct, whereas different outcomes would be expected on the test of the second passage. If the general feeling of dissatisfaction explanation is correct, then the generation advantage should be reduced or eliminated in the test of the second passage. If, however, the opportunity to experience the memorial benefits of generating relative to reading is critical for inducing a processing change, then participants only reading critical items in the first passage should not change their processing strategy for the second passage, and a generation advantage should be seen on the second test as well. Consistent with the testing experience explanation, a generation advantage was obtained in the tests for both passages and, as with the first follow-up study, the size of this advantage did not differ across tests. Thus, when participants were denied the opportunity to experience the memorial advantage of generation in their own test performance—because the read versus generate encoding variable was manipulated either between passages or between participants—their ability to recall to-be-read items remained significantly poorer than their ability to recall to-be-generated items.

ADDITIONAL QUESTIONS AND POTENTIAL IMPORTANT APPLICATIONS

As indicated in our description of the studies of deWinstanley and Bjork (deWinstanley & Bjork, 2004; E. L. Bjork, deWinstanley, & Storm, 2007), this research explored whether learners could discover for themselves how to become more effective processors or encoders of to-be-learned information if given an informative test experience. The pattern of results observed across the four studies described above indicated that experiencing the advantages of encoding by generation versus only reading could induce learners to develop more effective encoding strategies. Or, in the terms of Koriat (1997), making learners sensitive to the power of generation as a condition of learning led them, in turn, to adopt enhanced strategies for the encoding of new information via reading—that is, even for information that they were not required to generate. These findings raise many interesting questions: some regarding the underlying cause of the effect observed by deWinstanley and Bjork (2004) and some regarding how these findings could best be applied to educational practices. It is to a discussion of some of these questions—topics of our recent and ongoing research—that we now turn.

One of our first investigated questions concerned the effect's durability. In the series of studies by deWinstanley and Bjork (2004) just reviewed, the second passage was always presented with little or no delay after the test of the first passage, raising the question of whether the testing experience only leads to enhanced encoding of new information when the new information is presented immediately after such a test. Perhaps, for example, the insertion of a delay between the testing experience and the presentation of the next passage would prevent participants from adopting a more effective processing strategy for subsequently presented to-be-read information. Should a delay produce such an effect, the applicability of these findings for educational purposes would be lessened. A related issue, in terms of the educational applications of these findings, would be whether the test experience must occur immediately after presentation of the passage in which participants both generated and read critical items. Or, might it be possible—as might frequently be necessary in educational settings—to delay the test without eliminating the learners' ability to benefit from the test experience?

In two recently conducted studies, we have addressed both of these questions as follows. In a first study, we inserted a delay filled with a number of other activities between the presentation and testing of a first passage and the presentation and testing of a second passage, and in a second study, we inserted the same type of delay between the presentation of the first passage and its subsequent test. Furthermore, as in the original deWinstanley and Bjork (2004) studies, both the first and second passages always contained both to-be-read and to-be-generated critical items, and the subsequent tests for the passages were in the form of a fill-in-the-blank type of test for the critical items. Importantly, for the applicability of the present effect for educational purposes, the results obtained replicated those of the original studies. Thus, it appears that the observed effect of the testing experience—that is, its ability to lead learners to develop more effective encoding strategies for processing future information—can persist across a delay filled with other activities and, furthermore, does not require that the test be administered immediately after presentation of the first passage.

Also being addressed in our current research is the question of the necessity for learners actually to experience the differential effectiveness of encoding via generation versus reading in the context of a memory test. From the deWinstanley and Bjork (2004) studies, we know that this experience is critical in that it was only the participants given this experience who then went on to adopt more effective processing strategies for future to-be-read items. Additionally, that such an experience would be necessary is consistent with previous research indicating that learners are typically unable to judge the efficacy of a given processing strategy during its execution and do not switch from a less to a more effective strategy without an opportunity to experience their relative effectiveness (see, e.g., Dunlosky & Hertzog, 2000). What remains unclear, however, is whether the relative effectiveness must be experienced in the context of an actual testing episode. Perhaps, for example, simply instructing or informing learners regarding the differential effectiveness of the two types of encoding might be sufficient, and we are currently addressing this possibility in ongoing research by varying across participants the type of experience they have following study of the first passage. For example, for some participants, they are given the opportunity to experience the relative effectiveness of generating versus reading for later performance via a testing episode, whereas for others, we are instructing them in various ways regarding the relative memorial effectiveness of generation versus reading as encoding strategies. Although still ongoing, results so far are strongly indicating the critical

nature of the actual test experience for producing the desired enhancement of future encoding strategies.

We are also currently addressing the theoretical question of how participants are improving their encoding of information in the second passage. One possibility that we are considering is that during original study, participants used contextual information provided by other words in the passage to help them complete or encode the to-be-generated critical items and, then, used this information again in the subsequent fill-in-the-blank test to aid their recall. Indeed, the use of such a strategy—that is, to use contextual information first to help complete and then to help recall the generated items—could underlie the generation advantages observed on the tests of the first passages and, additionally, would be an explanation consistent with both the procedural and the transfer-appropriate multifactor accounts described earlier. Specifically, it would have been the match between the information strengthened while completing the generation task and the information needed to perform well on the later test, or the ability to reinstate during test the cognitive procedures used during study, that had resulted in the observed generation advantages.

Should this explanation for the generation advantage initially observed in the deWinstanley and Bjork (2004) studies be correct, perhaps participants—becoming aware of both their superior recall of generated items and their use of such contextual information in recalling them on the test—then attended to such contextual information during the study of the second passage for both types of critical items, consequently eliminating a generation advantage in their subsequent recall. Such an account would also be consistent with the finding that the generation advantage was not eliminated on tests of the second passage when participants had only received to-be-generated critical items during study of the first passage. It may have been more difficult for such participants—even if using contextual information in the same way during study of the first passage—to notice the role of this strategy in aiding their recall during the test because they were only recalling items they had generated and thus were not able to experience a contrast between their ability to recall words encoded via generation and reading. Consequently, they would have been less likely to transfer the use of this strategy when encoding to-be-read items presented in the second passage.

In research recently completed, we have tested this potential explanation by using different types of retention tests following study of the first passage: in particular, ones that provide contextual information during the testing process and ones that do not. Our reasoning in so doing was as follows: If this explanation is correct, then when the test following study of the first passage does not provide contextual information, the testing experience should not lead participants to the discovery of this encoding strategy, and thus the generation advantage should not be eliminated in the testing of subsequently presented material. Consistent with our hypothesis, when participants were given the same type of fill-in-the-blank test as initially used by deWinstanley and Bjork (2004), a generation advantage was then eliminated in the test of the second passage, but in contrast, when a test that did not provide such information (e.g., a free-recall test) was administered following study of the first passage, a generation advantage continued to occur in the test of the second passage (Sin, Storm, Bjork, & deWinstanley, 2006).

Finally, in an even more direct test of this potential strategy in leading to enhanced encoding of the second passage, we have recently conducted a study in which we varied both the nature of the test given to participants following their study of the first passage (i.e., tests that did or did not provide contextual information) and the type of information for which we tested following the second passage (either critical items or contextual items). Our reasoning in so doing was that it would only be participants who could discover this strategy during the test experience (i.e., those given a test involving contextual information) that would then go on to process such information more effectively in the second passage. Consequently, these participants should reveal a superior ability to recall contextual items in the test of the second passage, and the results obtained were consistent with this reasoning (Little, Storm, & Bjork, 2008).

CONCLUDING COMMENTS

As clearly documented by research on testing effects, it is not only during study that learning takes place. Learning also occurs during tests: Successful retrieval modifies the representation of the material so retrieved, making it more retrievable in the future. In addition to such specific effects on learning that can occur as a consequence of retrieval during tests—that is, the modification of the representations in memory of the retrieved information—we believe our research demonstrates that another type of learning can also take place during tests—in particular, that a higher-order type of learning can occur as well, such as the learning of an improved strategy for encoding future information.

Additionally, while our research has focused on only one such strategy—that engendered by the generation of to-be-learned information—it seems possible that learners could be made sensitive to other extrinsic factors or conditions of learning that enhance long-term performance through similar testing experiences. Thus, the line of research that we have outlined in this chapter seems to us to paint a promising picture from an applied perspective: namely, that providing students with opportunities to experience the consequences of differentially effective encoding processes in their own performance—either in the context of tests, as was done in our research, or potentially in other ways as well—can lead students to discover and then to adopt on their own more effective ways of processing future to-be-learned information. That is, beyond the more effective learning of the information in question, they may also, in general, be learning how to learn more effectively.

REFERENCES

Begg, I., Vinski, L., Frankovich, L., & Holgate, B. (1991). Generating makes words memorable, but so does effective reading. *Memory and Cognition, 19*, 487–497.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1997). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*, 55–68.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). New York: Worth Publishers.

Bjork, E. L., deWinstanley, P. A., & Storm, B. C. (2007). Learning how to learn: Can experiencing the outcome of different encoding strategies enhance subsequent encoding? *Psychonomic Bulletin and Review, 14*, 207–211.

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123–144). Hillsdale, NJ: Erlbaum.

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

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: Erlbaum.

Bjork, R. A., & Bjork, E. L. (2006). Optimizing treatment and instruction: Implications of a new theory of disuse. In L.-G. Nilsson & N. Ohta (Eds.), *Memory and society: Psychological perspectives* (pp. 109–133). New York: Psychology Press.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20,* 633–642.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132,* 354–380.

Crutcher, R. J., & Healy, A. F. (1989). Cognitive operations and the generation effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 669–675.

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork & R. A. Bjork (Eds.), *Human memory* (pp. 197–236). San Diego, CA: Academic Press.

deWinstanley, P. A. (1995). A generation effect can be found during naturalistic learning. *Psychonomic Bulletin and Review, 2,* 538–541.

deWinstanley, P. A., & Bjork, E. L. (1997). Processing instructions and the generation effect: A test of the multifactor transfer-appropriate processing theory. *Memory, 5,* 401–421.

deWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory and Cognition, 32,* 945–955.

deWinstanley, P. A., Bjork, E. L., & Bjork, R. A. (1996). Generation effects and the lack thereof: The role of transfer appropriate processing. *Memory, 4,* 31–48.

Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging, 15,* 462–474.

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language, 33*(4), 545–565.

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

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107,* 311–327.

Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 484–494.

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

Jacoby, L. L., Bjork, R. A., & Kelley, C. M. (1994). Illusions of comprehension, competence, and remembering. In D. Druckman & R. A. Bjork (Eds.), *Learning, remembering, believing: Enhancing human performance* (pp. 57–81). Washington, DC: National Academy Press.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 127,* 349–370.

Koriat, A. (1998). Illusions of knowing: A window to the link between knowledge and metaknowledge. In V. Y. Yzerbyt, G. Lories, & B. Dardenne (Eds.), *Metacognition: Cognitive and social dimensions* (pp. 16–34). London: Sage.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 187–194.

Koriat, A., & Bjork, R. A. (2006). Illusions of competence during study can be remedied by manipulations that enhance learners' sensitivity to retrieval conditions at test. *Memory and Cognition, 34,* 959–972.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 989–998.

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.

Little, J. L., Storm, B. C., & Bjork, E. L. (2008, November). *How does experiencing the generation advantage lead to improved reading of new information?* Poster presented at the annual meeting of the Psychonomic Society, Chicago.

Lovelace, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 756–766.

McDaniel, M. A., Waddill, P. J., & Einstein, P. J. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language, 27,* 521–536.

McNamara, D. S., & Healy, A. F. (1995a). A generation advantage for multiplication skill training and nonword vocabulary acquisition. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Learning and memory of knowledge and skills: Durability and specificity* (pp. 1–29). Thousand Oaks, CA: Sage.

McNamara, D. S., & Healy, A. F. (1995b). A procedural explanation of the generation effect: The use of an operand retrieval strategy for multiplication and addition problems. *Journal of Memory and Language, 34,* 399–416.

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

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.
Pesta, B. J., Sanders, R. E., & Murphy, M. D., (1999). A beautiful day in the neighborhood: What factors determine the generation effect for simple multiplication problems? *Memory and Cognition, 27*(1), 106–115.
Roediger, H. L., III, & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–210.
Roediger, H. L., III, & Karpicke, J. D. (2006b). Test enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255.
Simon, D., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 907–912.
Sin, N. L., Storm, B. C., Bjork, E. L., & deWinstanley, P. A. (2006, April). *Learning how to learn: How does experience enhance subsequent encoding?* Poster presented at the annual meeting of the Western Psychological Association, Palm Springs, CA.
Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 592–604.

18

FAILING TO PREDICT FUTURE CHANGES IN MEMORY

A Stability Bias Yields Long-Term Overconfidence

Nate Kornell

If you have ever experienced that panicky moment when you realize, while taking a shower in Los Angeles, that you are supposed to be in Seattle telling 75 business executives about their memories, you may have learned something about the instability of memory (Robert A. Bjork, personal communication, November 15, 2009). It is easy to forget travel plans that once seemed memorable, and it is easy to add embarrassing experiences to one's memory.

Human memory is anything but stable. We constantly forget old information and form new memories. Yet recent research has demonstrated a *stability bias* in human memory: People act as though their memories will remain stable in the future. They fail to predict future forgetting (Koriat, Bjork, Sheffer, & Bar, 2004) and future learning (Kornell & Bjork, 2009). In this chapter, I discuss the importance of assessing one's memory in everyday life, draw a distinction between predicting future remembering versus predicting future changes in remembering, and review evidence substantiating the stability bias. I then describe an experiment examining the cause of the stability bias. I asked participants ($n = 430$) to predict their ability to remember word pairs they would study once or four times and would be tested on in five minutes or one week. Participants predicted significant learning and forgetting but vastly underpredicted both effects, demonstrating