Jacoby, L.L., Bjork, R.A. & Kelley, C.M. (1994) In D. Druckman & R.A. Bjork (Eds.), Learning, remembering, believing: Enhancing human performance. Washington, DC: National Academy Press.

4

Illusions of Comprehension, Competence, and Remembering

People's performances as individuals, and their contributions as members of groups, depend not only on their actual competence, but also on their assessment of that competence. The reading people take of their own state of knowledge or level of skill determines how they allocate their time and energy and the influence they have on others. Many things depend on people's subjective assessment of what they know and do not know: whether they volunteer for certain roles or tasks, whether they seek further practice or instruction, and whether they instill confidence in others, as well as the answers they give to questions from superiors and subordinates and the affect they induce in others by facial expressions and body language (see Chapter 10). Recent evidence suggests, however, that under certain conditions, people's assessments of what they know or remember can be seriously flawed, particularly when they use one index, such as familiarity, recognition, or fluency, to predict something else, such as unaided recall or production.

There are a variety of ways in which people can be fooled. Consider some of the assessments college students typically make in a course. They must decide if they are prepared for an upcoming test. In order to allocate study time, they must monitor the state of their own learning and comprehension across the topics for which they are responsible. If they end up doing poorly on the test, they may take a reading of their memories for how much time they spent studying. At the end of the course, they may also monitor their learning and comprehension in the course—when asked to judge the effectiveness of the instructor, for example, or deciding whether to recommend the course to a friend.

Cognitive Illusions

Newton (1990) provides a powerful demonstration of how imagination can color subjective experience. She asked half of her student subjects to be "tappers" and half to be "listeners." The tappers were asked to select songs from a list of 25 common titles and tap out the rhythm to listeners sitting near them. Listeners did not know the list of potential songs. Then the tappers were asked to estimate the probability that their listeners could correctly identify the song from the taps and to estimate the numbers of students in general who could identify the song from the taps. To appreciate the study, you might want to pause and tap out a song and reflect on your subjective experience. Tappers do not hear merely a series of taps on the table, but a whole imagined rendition of the tune, with orchestration and voices. The imaginative additions apparently also transform one's experience of the taps: tappers estimated the likelihood that listeners could correctly identify the songs at 50 percent; the actual rate of correct identification was a mere 2.5 percent. Griffin and Ross (1991) make the point that to correctly estimate the likelihood that listeners would identify the song, tappers needed to recognize the extent to which their subjective experience of the taps differed from the subjective experience of the listeners, and they needed to make allowance for the difference in experience when making their estimates. Tappers may have failed to grasp the degree of difference their privileged information made in their experience of the taps, or they may have been unable to adequately judge how much they needed to adjust their estimates.

It is not difficult to imagine how such processes probably color many interactions in the workplace. Inefficient communication from superiors to subordinates, and vice versa, and among coworkers in general, is no doubt frequently the product of people's inability to make allowances for the difference between their subjective experience and the likely subjective experience of their listener.

Griffin and Ross (1991) argue that a variety of anomalies in social judgment and prediction stems from variations in how particular people construe a situation, and their failure to appreciate that their construal is only one of many possibilities. The "false consensus effect" in social judgment refers to people's tendency to overestimate the commonness of their own beliefs and choices (Ross et al., 1977). For example, subjects might be asked the question: "Which kind of music do you like better, music from the 1960s or music from the 1980s?" and then asked to estimate the proportion of college students who would make the same choices. The subject has to make a particular interpretation of what is meant by "1960s music" and what is meant by "1980s music" and then choose between the two. But as was the case for the tappers in Newton's study, subjects who interpret such

questions may not realize the degree to which their particular construal of the elements of the choice is idiosyncratic. That is, they might think of five or six really good bands from the 1960s and several bad bands from the 1980s, choose "1960s music," and assume that most people would make a similar choice. In contrast, someone else might quickly think of a number of good bands from the 1980s, and several lousy bands from the 1960s, choose "1980s music" and assume that their preference would be widely shared. Gilovich (1990) argued that the false consensus effect occurs because the phrasing of the options only partially specifies their meaning. Dawes (1990) questioned the false consensus effect, showing that errors are often people's overestimating the extent to which they are unique, rather than the opposite.

The important point for present purposes is that reliance on subjective experience as a basis for judgments is prone to error because of failure to take individual differences into account. Griffin and Ross (1991) point out that much human misunderstanding may stem from people's failure to appreciate the degree to which their construal of a situation differs from that of other people. Disagreements often arise because of what Asch (1969) called different objects of judgments (different construals), rather than different judgments of the same object. Conflicts are exacerbated by people's resistance to accepting the possibility of multiple construals. It is common for partisans in a disagreement to feel that their own beliefs follow from an unbiased consideration of facts and arguments, while the other side's beliefs are affected by biased interpretation of facts and arguments. That is, others are open to the effects of beliefs on construal, but not oneself.

One source of variation in people's interpretation or construal of a situation is past experience. In one study, people who walked down a hallway that contained a poster of a missing child were more likely to interpret an ambiguous interaction between a child and an adult as a possible kidnapping (James, 1986). That effect could be an instance of memory without remembering: recent experience with one event made particular construals of a later event more available. Impressions of others are also influenced by "priming" effects of this sort. People's judgments about another person are influenced by what categories are made most easily accessible by recent prior experience (see e.g., Higgins, 1989; Srull and Wyer, 1979). The same behavior, for example, can be judged as "assertive" or as "aggressive," depending on which category has been primed. These effects occur even though subjects cannot "remember" or are unaware of the relevance of the priming stimulus.

Construal of a current situation is often determined by small details that make the current situation analogous to a particular past experience. Gilovich (1981) has shown that subjects' recommendations of how to resolve a hypothetical international crisis were influenced by irrelevant similarities be-

tween the hypothetical crisis and historical analogies. Politicians know that by describing a conflict as analogous to Vietnam they can very effectively color the interpretation of that conflict. Such effects on interpretation of an ambiguous stimulus can depend on specific details that recruit memory for one particular prior event rather than another. For example, Smith and Zarate (1992) showed that the likelihood of interpreting someone in terms of a stereotype depends on the similarity of the current situation to a prior situation in which the stereotype was invoked. Ross (1977) demonstrated similar specificity in the likelihood of solving a problem that is analogous to a prior problem: even irrelevant details in the current situation can change the probability of recruiting and using memory for a prior problem. This work shows that past experience influences people's interpretation or what comes to mind in a current situation and, thereby, their judgments.

Past experience can also alter social judgments by making cognition more fluent. In a series of experiments, subjects were asked to solve anagrams (e.g., fscar) and then rate the difficulty of the anagrams for others (Jacoby and Kelley, 1987). Subjects tended to use their own experiences in solving the anagrams as a basis for judging the difficulty of the anagrams for others—the correlations between speed of solving the anagrams and difficulty ratings were quite high. Before attempting to solve and rate the difficulty of the anagrams, subjects in some conditions read a list of words that were the solutions to half of the anagrams they were later asked to solve. Prior reading of those words led to faster solution of those anagrams. However, subjects continued to use their subjective experience of difficulty as a basis for judging difficulty for others, even though their experience had been affected by the effects of prior reading of the solution words. They rated anagrams of the studied words as easier for others to solve than anagrams of equal difficulty constructed from words not presented earlier.

People in the anagram experiments may have failed to recognize that half of the solution words they produced had been on the list they studied earlier. Or, even if they recognized the solution words from the list, they may have failed to appreciate that reading the solution words led them to solve those anagrams more quickly. In follow-up experiments, subjects were informed of the effect of prior reading of solution words on later anagram difficulty or informed of the effect and asked to attempt to recognize each solution word before judging the anagram difficulty. In the first case, subjects who were informed of the effect nonetheless rated old anagrams as easier for others than new anagrams. However, subjects who were informed of the effect and asked whether they recognized the solution words as from the list they had studied did not rate the old anagrams as easier to solve than new anagrams: those subjects avoided the biasing effects of memory by switching from subjective experience to a more objective basis for judgments (see below). That switch was necessary because there is no

sure way to correct subjective experience. That is, the question "How easy would this anagram seem had I not read the solution?" is impossible to answer.

Subjects in the anagram experiments were extremely reluctant to give up subjective experience as a basis for predicting the performance of others. As described in the next section, there is good reason for that reluctance: even spoiled subjective experience can provide a more accurate basis for predicting the performance of others than do alternative, more objective bases for judgment.

Subjective experience has a bad reputation as a basis for judgments. Treating one's own experience as a mirror of objective reality can lead to egocentric errors as extreme as those made by children (Piaget and Inhelder, 1956). Olson (1986) argued that children fail to distinguish between what they see and what they know and therefore mistakenly extend their privileged interpretation to others. The examples given in the preceding sections show that such egocentrism is not restricted to children; adults make similar errors. Apparently, people do not develop a general ability to distinguish what is given in a stimulus and what is an interpretation. Instead, they continue to rely on their subjective experience—even though that experience may be spoiled by factors such as memory.

People do come to realize that in certain situations they must give up subjective experience and rely on a more analytic, objective basis for judgments. Good teachers must be mindful that their understanding of material is not an appropriate basis for judging their students' understanding. As Piaget (1962:5) pointed out:

> every beginning instructor discovers sooner or later that his first lectures were incomprehensible because he was talking to himself, so to say, mindful only of his own point of view. He realizes only gradually and with difficulty that it is not easy to place oneself in the shoes of students who do not yet know what he knows about the subject matter of his course.

Goranson (1976) created a series of demonstrations of the problems of hindsight in judging the difficulty of problems for others. When people were given the answers to puzzles and problems and asked to "go through the motions of solving the puzzle just as if you didn't already know the solution," he found that they vastly underestimated the time it would take to solve the problem in comparison with a group that actually attempted to solve it. This hindsight effect could lead teachers who rely on their own subjective experience of understanding as a basis for judging the difficulty of problems to judge their students as dull and slow. Goranson suggested a number of strategies for teachers to use to more accurately assess students' understanding: frequent tests, checking students' lecture notes, and individual tutorials.

The ability to know that subjective experience is a bad basis for predicting for others may be situation specific. To avoid egocentrism, one must have an objective basis for judgment that can replace subjective experience. One objective basis for judgments is a theory. In the case of the anagram experiments described above, subjects who had read the solutions to anagrams in the first phase and then realized that their subjective experience of difficulty was a poor basis for judging difficulty for others, could have switched over to analytic judgments based on some theory of what makes an anagram easy or difficult. Judgments could be based on rules such as "low frequency words are difficult to solve," or "anagrams with many vowels are more difficult to solve." To see how effective those rule or theory-based judgments were, subjects in one condition in the anagram experiment were shown the solution word to the anagram immediately before seeing the anagram, for example, "scarf" then "fscar." Reading the solution word prevented subjects from having any experience of the difficulty of the anagram and forced them to rely on rules as a basis for judgments. Those rules turned out to be inadequate: the correlation between judged difficulty for anagrams presented with solutions and the actual time to solve those anagrams in the control condition was quite low. Furthermore, the rule-based judgments were even worse than judgments based on subjective experience that had been altered by prior reading of the solution words. Unless people have a well-specified theory, they may be better off relying on their subjective experience as a basis for judgments.

Analysis and reason may actually lead to worse judgments than unanalyzed feelings or intuitions when one is predicting preferences or other personal judgments. Wilson and Schooler (1991) found that translating affect into reasons may be disruptive. They asked subjects to taste and rate the quality of five different brands of strawberry jam. Their criterion measure of "goodness of ratings" was the rank-ordering of the jams by a panel of experts who had rated the jams on 16 dimensions for *Consumer Reports*. When subjects were asked simply to taste the jams and rank them, their rankings correlated fairly well with the experts' rankings (average correlation, .55). However, when another group of subjects was asked to provide reasons for their judgments, their rankings of the quality of the jams was less like that of the experts (average correlation, .11). Wilson and Schooler argued that forcing their subjects to think about why they liked or disliked each jam turned an affective response into a more cognitive one, and in this case the cognitive judgments appeared to be less correct.

In another study by Wilson et al. (1984), couples in steady dating relationships were asked how well adjusted they thought their relationship was. In the "control condition," that's all they were asked. In the "reasons condition," they were also asked to list all the reasons they felt the way they did about their dating partners. Several months later, it appeared that asking subjects for reasons undermined the accuracy of their assessment of their relationships: the correlation between ratings of the relationship and whether they were still dating their partner several months later was only .10 in the reasons condition; it was .62 in the control condition. (It is safe to assume, the researchers hope, that there was no causal relationship between having subjects list their reasons and their reduced likelihood of continuing to date the rated partner.) People's analysis of what makes a jam good or what makes a relationship good may simply be bad theories. Perhaps an unanalyzed affective response captures more important dimensions or weights the dimensions more appropriately, about a relationship than any theory of what makes a couple compatible.

Self-Monitoring of Learning: Illusions of Comprehending and Knowing

An important determinant of learning is a learner's ability to self-monitor knowledge and learning. Among the consequences of self-monitoring are effects on judgments of one's self-efficacy, which is important for motivation and performance (see Chapter 8). An attribution process clearly underlies people's evaluation of their own abilities, as shown in a classic series of experiments done by Jones et al. (1968), who investigated the effects of varying distributions of success and failure on attributions of intellectual ability.

Jones et al. (1968) arranged conditions such that success rate in solving problems either increased or decreased across a series of problems; the total number of successful solutions was the same for the two conditions. Subjects were asked to estimate their future performance. Subjects in the "ascending success" condition were more confident about their own ability than were those in the "descending success" condition. However, results were the opposite when subjects were given the same patterns of successes and failures of others and asked to evaluate their abilities. In that case, a performer with a descending success rate was judged as more intelligent and expected to outperform a performer with an ascending success rate. Jones et al. explain their results in a way that is reminiscent of Heider's (1958:157) proposition that perceivers are prone to attribute the reactions of another person to characteristics of the person (e.g., intelligence), while one's own reactions are attributed to the objective world (e.g., the difficulty of problems). The early failures of others in the ascending success condition give rise to a first impression of low intelligence and that first impression is not reversed even in the face of others' future successes. For oneself, in contrast, the early failures in the same ascending series is attributed to the greater difficulty of early problems, settling into the task, and so forth.

Jones et. al. (1968) considered the practical implications of their research. They ask one to imagine a situation in which A is observed by B to

show systematic improvement in performance, for example, an improving player (A) observed by a coach (B). Results of their experiments suggest that A will feel very confident about his or her future performance, while B will predict that A's future performance will be relatively low. The improving player is outraged and does not understand why he or she was kicked off the team just as his or her true ability was beginning to show. Examples of this sort underline the importance of considering attribution processes in educational settings.

As is true when predicting the performance of others, monitoring one's own performance can rely on subjective experience and, consequently, is subject to errors that reflect cognitive illusions. For example, the ease or fluency of identifying a briefly presented word is often used as a basis for judging the visual duration of the word. Similarly, the ease of generating an answer to a question may be one basis for confidence that the answer is correct. To illustrate, what was Buffalo Bill's last name? If an answer comes quickly to mind, how do you know it was correct?

Kelley and Lindsay (1993) found negative correlations between the latency to think of an answer and confidence in that answer (see also Nelson and Narens, 1990). They then manipulated the speed with which correct and incorrect answers came to subjects' minds by having subjects first read a list of answers to general knowledge questions and then attempt to answer those questions. Prior exposure to correct answers (e.g., Cody) increased the speed and frequency with which correct answers were given on the general knowledge test and also increased subjects' confidence in those answers. Even stronger evidence for the claim that the speed with which an answer comes to mind is a basis for confidence was gained by changing what answers came to mind. Illusions of knowing occurred when subjects studied closely related but incorrect answers (e.g., Hickock). The studied but incorrect answers were more likely to be confidently given as answers to the general knowledge questions. Kelley and Lindsay (1993) argued that those effects could occur without conscious memory for the list of words.

Nelson and his colleagues (for a review, see Nelson and Narens, 1990) point out the importance of monitoring one's own learning for the control of study time in self-paced learning. Unfortunately, judgments of learning are surprisingly inaccurate under certain conditions. For example, subjects might study a list of unrelated paired associates, such as *ocean-tree*, and then be asked to predict the probability of their being able to recall *tree* to the cue *ocean* on a memory test a short time later. On some judgment-of-learning trials, the pair studied earlier was shown intact; other trials consisted of the cue *ocean* being shown alone. Judgments of learning have little predictive value if the judgments are made immediately after one has studied the item. However, the predictive accuracy soars to almost perfect if the judgments of learning are delayed for several minutes after studying

the item and if the cue word is shown without the target word (Dunlosky and Nelson, 1992; Nelson and Dunlosky, 1991). Nelson and Dunlosky argued that people make judgments of learning by using the cue *ocean* to retrieve information from memory. If the judgment is made immediately after studying an item, such as *ocean-tree*, the response term *tree* is still available from short-term memory, and the subject may be unable to assess how readily available *tree* would be from long-term memory. Delaying the judgment for more than 30 seconds means that it is then based on information in long-term memory, which appears to be a much better source of information for the later memory test. When the to-be-recalled target word was shown to subjects, they, like the subjects who were presented the answer to an anagram along with the anagram, could apparently no longer judge the likelihood that they would have been able to come up with the answer on their own. Giving the target word paired with the cue word at a delay appears to ruin subjects' subjective experience in a fashion analogous to that experienced by subjects in the anagram experiments. Spellman and Bjork (1992) have argued that with the target word present subjects are denied the opportunity to exercise and benefit from the retrieval processes that will be later required on the final cued test.

Evaluation of Instruction

During training and education, learners monitor changes in their performance or understanding and use those changes to evaluate the instruction. However, changes in ease of performance or increased understanding can be attributed to a variety of sources and so give rise to different judgments. The new understanding can be mistakenly interpreted as something "obvious" that the student knew all along.

One "knew it all along" effect of this sort has been documented in laboratory investigations (e.g., Fischhoff, 1975; Hasher et al., 1981). People were unable to remember their original knowledge state—that Aladdin, of the magic lamp story, originated in Persia—after the experimenter has provided them with outcome information—it was really China. The effect is an example of hindsight bias: people are prone, with hindsight, to treat an observed outcome as inevitable even though the outcome may have objectively been an unlikely one (Fischhoff, 1975). Errors of this sort are pervasive. However, Hasher et al. (1981) showed that the "knew it all along" effect can be avoided. Under the proper test conditions, people can recover their original state of knowledge and realize the difference between that original state and outcome information. That is, the later outcome information does not erase or "overwrite" original knowledge. These findings are important for questions about recall processes, especially those involved in eyewitness testimony (considered below).

The "knew it all along" effect can be understood as another example of fluency being misattributed to produce a cognitive illusion. The effect is quite similar to the effect of the ease of generating an answer to a question on confidence in the correctness of the answer (Kelley and Lindsay, 1993). When a question is answered easily, the subjective experience may be that the answer came from one's general knowledge held prior to the experiment—one "knew it all along" even when the "it" was misleading new information gained in the experimental setting (e.g., Buffalo Bill's last name was Hickock). Similar misattributions can produce errors in self-monitoring of knowledge and learning.

The opposite effect can also occur in instruction: students can misattribute their improved performance or understanding to a quality of the task or material. For example, across a series of lectures, students may think their instructor's style is improving—that the instructor is speaking more slowly, perhaps, or making more organized presentations. The students' subjective experience of greater clarity may be a misattribution of effects due to their improved understanding of the subject; the speech habits of the lecturer may be completely unchanged. The subjective experience of greater clarity of speech may depend on increased ease of comprehension, analogous to the influence of memory on noise judgments, as discussed below.

Speculatively, it might be possible to intentionally arrange conditions to create an illusion of effective instruction. This devious possibility was suggested by an acquaintance of the authors who became suspicious about a book he bought to help him prepare for the Law School Admissions Test (LSAT). The book provided sample questions and discussion of strategies to use, followed by repeated tests that were scored in ways that supposedly paralleled the actual LSAT. The person worked through the first few tests and was horrified when he obtained low scores. He had scored well on standardized tests in the past, and after calming down, he wondered if there was something peculiar about the book. He turned to the end of the book and took the last test, and scored very well. After closer inspection of the questions, he became convinced that the sample questions on the earlier tests were objectively more difficult than those on the later tests. He argues that the book was written in a way to intentionally create an illusion of improvement so as to increase sales of the book. This case history is not cited to suggest applications to how one should structure training to give an illusion of improvement, but to illustrate the ambiguity in interpreting changes in one's performance.

Knowing and Learning

Although people may be unable to remember the answer to a knowledge question, they sometimes have a feeling of knowing the answer (a

"feeling of knowing"), and in that state they can predict with considerable accuracy the likelihood that they will be able to pick out the correct alternative on a subsequent recognition test (e.g., Blake, 1973; Hart, 1967). The "calibration" of judgments refers to the correlation of people's self-predicted performance and their actual performance. People are usually overconfident: their predicted level of performance is higher than their actual performance. Such overconfidence resembles that of people's confidence in their responses to inquiries concerning matters of fact (e.g., Fischhoff et al., 1977). Despite such overconfidence, feeling-of-knowing judgments are reasonably well calibrated (see Nelson, 1988). People might rely on feeling of knowing as an indication of what question answering strategy should be used: given a positive feeling of knowing, people attempt to answer a question rather than simply to give up (e.g., Reder, 1987). This use of feeling-of-knowing judgments requires that they be made rapidly. In fact, Reder (1988) found faster reaction times for making feeling-of-knowing judgments than for answering questions.

What mechanisms underlie feeling-of-knowing judgments? Studies show that prior exposure of the target item, which would aid its later retrieval, is less important for feeling-of-knowing judgments than is prior exposure to the question or cue. For example, Reder and Ritter (1992) found that when exposure to parts of problems requiring arithmetical calculation were manipulated along with prior exposure to answers for the problems, it was frequency of exposure to parts that increased the feeling of knowing (see also Schwartz and Metcalfe, 1992, who obtained analogous results in subjects' memory for cue-target word pairs). Reder and Ritter (1992) concluded that assessing familiarity of questions allows one to make fast decisions without retrieving the target information. To the extent that feeling-of-knowing judgments reflect familiarity of the cue rather than partial retrieval of the target, feeling-of-knowing judgments will not accurately predict performance on subsequent tests of memory for the target. That is, illusory feeling of knowing can be produced by misattributing the familiarity of a question to knowledge of the answer to that question.

A similar misattribution underlies illusions of comprehension. Self-assessment of comprehension of expository text is often a poor predictor of objectively assessed comprehension, as shown in a series of studies by Epstein et al. (1984) and by Glenberg et al. (1982). High confidence in understanding of text was often coupled with failure to detect contradictions within the text. That is, people claimed to fully comprehend a text even though it could be shown that they failed to note that sentences embedded in the text were contradictory. Methodological problems in those early experiments led to the underestimation of the correlation between self-assessment and objectively assessed comprehension (Weaver, 1990); however, even when those methodological problems were corrected, the correla-

tion is still low. More recent investigations have used measures that are akin to those used in feeling-of-knowing experiments to reveal illusions of comprehension. For example, Glenberg and Epstein (1985) required subjects to rate their confidence in their ability to verify inferences derived from a text they had read. The calibration of subjects' judgments was assessed by computing the correlation between their predicted and actual performance: the correlation was typically near zero!

One cause of illusions of comprehension is confusion between knowledge of the domain that a text represents and knowledge gained from the text—misattribution of familiarity. Glenberg and Epstein (1987) had students with expertise in physics or music theory read expository texts from both domains and predict their ability to verify inferences based on the central premises of the texts. Results showed that subjects were well calibrated across domains; physics students, for example, were more likely to correctly predict that they could verify inferences from a physics text than were music students. More interesting, within a domain, expertise was inversely related to calibration, that is, experts were more confident than they should have been. Confidence judgments were apparently based on self-classification as expert or nonexpert in the domain, rather than on comprehension of the particular text that was read. Familiarity with the domain was misattributed to comprehension of the text. Because of reliance on prior knowledge when judging comprehension, an expert is even more subject to illusions of comprehension than is a novice.

Similarly, Begg et al. (1989) point to the importance of the form of questions posed to subjects when they are asked to predict the likelihood that they will remember an item on a future memory test. They argue that memory predictions are based on implicit judgments of how easily the item is processed while answering the predictive question. If items are processed easily for reasons that also contribute to successful memory performance, predictions are accurate; otherwise, predictions are less accurate. An interesting implication of their view is that "well-polished" lectures contribute to an illusion of comprehension. The danger is that students mistakenly attribute the fluency of their understanding to actual comprehension of the substantive aspects of the lecture rather than to the lecturer's easily accessible language and sentence construction. (The "Feynman effect" is the supposed evidence that the students of the famous lecturer and physicist, Richard Feynman, did poorer on standard physics exams than did students in other sections—owing presumably to Feynman's making difficult content seem easy to understand, which in turn gave students a false reading of their own knowledge—and their need to study.) The problem is that the criterion of "easy to listen to" that the student uses to judge comprehension is radically different from the later test used by the lecturer as a measure of comprehension and memory. Good speakers may pay the price of being accused of giving unfair tests.

Ability to judge one's reading comprehension is important for budgeting study time, as well as many other purposes, and there is no easy substitute for subjective experience as a basis for monitoring comprehension. This being the case, are there means by which illusions of comprehension can be avoided? The most promising way of helping people to avoid illusions of comprehension is by giving them experience with the type of test that will be used to measure comprehension. Glenberg et al. (1987) provide evidence that calibration is enhanced when the processes and knowledge tapped by a pretest are closely related to those required by the criterion test. Familiarity with the form of comprehension test is critical. Drum et al. (1981) looked at predictors of performance on standardized, multiple-choice tests of reading comprehension. These predictors included passage difficulty, characteristics of the question, characteristics of the correct choice, and characteristics of the distractors incorrect choices. The major explanation of the variance was characteristics of the distractors: distractor plausibility alone accounted for about 25 percent of the variance. Given that the distractors are not known until a person takes a test, it is virtually impossible for a person to accurately predict test performance without some good foreknowledge of the test.

In light of the importance of foreknowledge of a test, illusions of comprehension can be understood as reflecting a problem that is similar to that of construal problems (described above). The problem is that of misconstruing the form of test. But as in the examples of construal, such as the "tappers" judgment of the likelihood that a listener will identify the tune, the illusion of comprehension does not feel like a particular "construal" of what the test will be, but, rather, like some sort of absolute comprehension of the text. The problem is a general one: it is often the case that present performance is misleading.

Druckman and Bjork (1991:Chapter 3) and Schmidt and Bjork (1992) have argued that individuals responsible for the design of training programs can be seriously misled by assuming, explicitly or implicitly, that the conditions that enhance performance during training are also the conditions that enhance the posttraining or transfer performance. What one sees during training is current performance, which is an unreliable indicator of the learning that can support the longer-term performance that is the goal of training. There are many manipulations that speed the rate of improvement in performance during training—such as massing practice, providing very frequent feedback, and keeping the conditions of practice constant—that are among the very worst training conditions in terms of long-term retention and the ability to transfer or generalize training to altered conditions and tasks. Conversely, certain manipulations that appear to introduce difficulties for the learner, slowing the apparent acquisition of the skills and knowledge to be learned— such as spacing practice sessions on a given subtask, provid-

ing only intermittent summary feedback, and inducing variability in the conditions of practice—can enhance long-term performance and the ability to transfer or generalize. (Examples of the many studies demonstrating the differing short-term and long-term consequences of the spacing of practice, the scheduling of feedback, and the variation of conditions of practice are, respectively, Bahrick, 1979; Schmidt et al., 1989; and Catalano and Kleiner, 1984.)

The important point for present purposes, as argued by Bjork (1994), is that not only teachers, trainers, and instructors can be fooled by performance during training, but also learners themselves can be fooled. As discussed throughout this chapter, is very difficult to assess one's true state of knowledge and skill. In the absence of being able to take a "pure" subjective reading of the current level of learning achieved during training, a learner is susceptible to making the same inferential error to which trainers are susceptible: if I am performing well and appearing to improve rapidly, I am learning; if I am making errors and struggling, I am not. Such inferential errors can result in learners' being enthusiastic about training and instruction that are far from optimal, while being displeased with an excellent training regimen.

In short, learners do not necessarily know what is best for them. An experiment by Baddeley and Longman (1978) nicely illustrates this point. They had postal workers in Britain learn keyboard skills under varying conditions of the spacing of practice sessions. Consistent with the results of laboratory studies going back decades, they found that the more distributed in time were the practice sessions, the more efficient the learning per session. When, however, subjects were asked at the end of training to rate how pleased or displeased they were with their own particular training schedule, their reactions were negatively correlated with the efficiency of their training. Those with the most massed (and inefficient) schedule of practice were the most pleased; those with the most distributed (and efficient) schedule were the least pleased.

In general, people learn by making and correcting mistakes. In that sense, making errors during training can be viewed as an important part of subjective experience. Conditions during training that serve as crutches for performance—massing practice by subtask, for example, or keeping multiple aspects of the task environment fixed and predictable—are conditions that, in effect, deny learners the opportunity to learn what they don't know. Rather than learning from errors in the training context, those errors are deferred to some posttraining real-world environment in which performance matters. This is a particularly critical consideration for individuals in occupations for which society, and the individuals themselves, cannot afford on-the-job learning: police officers, air-traffic controllers, nuclear-plant operators, and so forth.

It is important, then, that training routines simulate the types of demands and variability of task conditions that can be expected in the transfer or real-world environment (see Chapter 3). As Bjork (1994) has argued, however, attempting to simulate the exact posttraining environment during training may not be optimal. There are experimental results that suggest that training programs should introduce variability and other demands that go beyond that expected in the real-world environment. A good example is an experiment by Kerr and Booth (1978), in which 8-year-old children were given practice in a task that required throwing small beanbags underhanded at a 4-inch by 4-inch target on the floor. During a series of training trials, each child was permitted to view the target, then asked to throw with his or her vision of the target blocked, and then shown the outcome to that throw. For half the children, the target distance was fixed at 3 feet; for the other children, the distance trial by trial was either 2 feet or 4 feet, but never 3 feet. On a final criterion test, during which all children were tested at 3 feet, the children who had practiced at 2 and 4 feet, but never at 3 feet, were significantly more accurate than the children who had been trained exclusively at the criterion distance. The same pattern was obtained with 12-year-old children, where the criterion distance was 4 feet and the training trials were either at that distance or were alternated between 3 feet and 5 feet.

Using adult subjects, Shea and Morgan (1979) found a strikingly analogous result in the transfer of blocked and random conditions of training. Subjects were given training on three different movement patterns, each of which involved knocking over several of a set of hinged barriers in a prescribed sequence as rapidly as possible. All subjects received 51 training trials, 17 on each movement pattern. For half the subjects, the trials were in blocks by movement pattern: they received all 17 trials on a given pattern before moving on to practice the next pattern. For the other half of the subjects, the patterns to be practiced on successive trials were determined by a random schedule, so the trials on the differing patterns were interleaved and unpredictable from the subject's standpoint. Not surprisingly, during training, the subjects in the blocked practice trials had better performance results across the trials on a given pattern than did the subjects in the intermixed trials. On a final retention test, however, carried out under either blocked or random conditions, the picture was dramatically different. Not only did the subjects who received random practice perform much better than blocked-practice subjects when tested under random conditions, they also performed better when the final criterion test was carried out under blocked conditions! Recently, Hall et al. (1992) have found the same pattern of results when players on the baseball team at California Polytechnic State University, San Luis Obispo, were given two sessions of extra batting practice a week for 6 weeks. Each such sessions consisted of 45 pitches (15 curve balls, 15 fast balls, and 15 change-ups): some players

received those pitches blocked by type; for other players the pitch on a given trial was determined by a random schedule. The percentage of solid hits during practice, and during the two posttraining criterion sessions—the first carried out under random conditions, the second under blocked conditions—were judged by assistant coaches who were blind to the purpose of the experiment. As in the Shea and Morgan (1979) study, random practice not only yielded better performance on a retention test carried out under random conditions, it also yielded better performance on the blocked criterion test.

ILLUSIONS OF REMEMBERING

Memory Without Remembering

It is logically possible that the feeling of remembering reflects faithfully the process of having access to a memory trace, and that such access is necessary and sufficient to produce a subjective experience of remembering. It is easy to demonstrate, however, that the link between subjective experience and the actual changes that have or have not taken place in one's memory is weak—in both directions. In a subsequent section we illustrate that people can experience illusory remembering. In this section we discuss the opposite case, that is, unconscious effects of past experience.

Cases of unconscious plagiarism illustrate how the past can unconsciously affect people. One famous case of plagiarism involved Helen Keller (Bowers and Hilgard, 1988). When Helen was 11 years old, she wrote a short story that was published as part of a newsletter by Helen's school. Readers of the story identified it as remarkably similar to a story published years earlier by an author named Margaret Canby. Helen was accused of plagiarism, but vigorously denied ever having read Canby's story. Eventually, however, a family friend came forward and said that she had signed the story to Helen during a visit 3 years earlier. Helen apparently used memory for the story when writing her own, without consciously remembering having previously heard it. That is, the story came to mind without an accompanying feeling of familiarity and was mistakenly accepted as her own invention. (For further discussion of unconscious plagarism, see Brown and Murphy, 1989; Reed, 1974.)

The most dramatic examples of such memory without remembering come from amnesic patients. Amnesic patients, who totally lack the subjective experience of remembering, still show specific effects of prior experience on performance as measured by indirect tests of memory. On an indirect test, people are not asked to report directly on memory for an event, as would be the case on a direct test of memory, such as recall or recognition. Rather, a person performs some task that can indirectly reflect the

effects of prior experience. For example, an amnesic patient might first be asked to read a list of words and then later asked to complete words based on only some of the constituent letters in those words. Amnesiacs, who perform very poorly on direct tests of memory, such as recall and recognition, nonetheless perform better on the identification of fragmented words that have been presented earlier than they do on other words not presented earlier (see, e.g., Warrington and Weiskrantz, 1974). People with normally functioning memories also show dissociations between performance on indirect and direct tests. They, too, for example, show effects of memory on fragment-completion tasks that are not accompanied by the subjective experience of remembering (for reviews, see Richardson-Klavehn and Bjork, 1988; Roediger and McDermott, 1993).

Prior experience may increase the ease or efficiency of later perception and thought: reading a list of words makes them easier to identify later from a fragmented or degraded form, and Helen Keller's experience of Margaret Canby's story may have made "creation" of a similar story easier 3 years later. One may either correctly interpret that ease or fluency of perception or thought as due to past experience or may misattribute those effects to qualities of the current situation.

Jacoby and Dallas (1981) examined the effects of prior study of words on people's ability to identify the words when they were presented briefly on a computer screen. In comparison with words not studied earlier, words that had been previously read in the experimental setting were more likely to be correctly identified when flashed for 35 milliseconds. The subjects in the experiment commented that some words were easier to read than others, but rather than correctly attributing that difference to the effects of having read the words in the first phase of the experiment, they said they thought some words were presented for a longer duration than other words. That is, the ease of perception due to memory was misattributed to a difference in presentation duration. Witherspoon and Allan (1985) then changed the subjects' task in the perceptual-identification part of the experiment by asking them to identify the duration of the words: they found that a single prior presentation was sufficient to lengthen the apparent duration of a word's presentation. Similarly, Jacoby et al. (1988) had subjects read a list of sentences and then listen to those sentences and analogous new sentences presented one at a time against a background of noise. Subjects misattributed their ease of hearing the old sentences to a lowering of the intensity of the background noise.

Past experiences may be a pervasive source of unconscious influences on subjective experience, and, in turn, on judgments based on subjective experience. For example, people judge a problem as easy if they can solve it easily, a paper as well written if they can read it fluently, and an argument as well reasoned if they can follow it easily. In each case, however,

prior experience may have affected the subjective experience. The apparent difficulty of a problem might decrease if one had solved it previously, the apparent quality of a paper might increase if one had read it (or written it!) previously, and the apparent force of an argument might increase if one had heard it before. As noted above, subjective experiences are important because people tend to assume their own experiences will be shared by others and so base judgments for others on the basis of their own experiences.

False Memories

Just as a past experience can change subjective experience of the present, without conscious awareness that one has been influenced by the past, the opposite attributional error can occur and produce illusions of remembering. The subjective experience of remembering partially depends on an inference or an interpretation of various cues. One cue that may lead one to experience remembering is the ease or fluency of current perception and cognition. Because past experience commonly enhances later perception and cognition, the fluency of perception and thinking is a valid indicator that one has seen a face before or read a text before. In particular, fluent perception may be one basis for the feeling of familiarity in recognition. We noted above that perceptual identification of briefly presented words is enhanced if the words have been studied previously (Jacoby and Dallas, 1981), and that if subjects are directed to make visual-duration judgments, they will misattribute the easy perception of old words to a change in duration of presentation. Can that same perceptual fluency be correctly interpreted as familiarity due to having read the words previously? Johnston et al. (1985) found a correlation between ease of perception and recognition memory judgments for pseudo words. Pseudo words that were easily identified perceptually were more likely to be recognized as "old."

If ease of perceptual processing is a cue that can serve as the basis for the feeling of remembering, then it should be possible in subtle ways to manipulate ease of perceptual processing and to create illusions of remembering. In an attempt to create memory illusions, Jacoby and Whitehouse (1989) varied the ease with which words on a recognition memory test were perceived. Subjects studied a list of words and then took a recognition test. Immediately prior to the presentation of each word on the recognition test, the same word (match) or a different word (mismatch) was briefly flashed on the screen, so briefly that subjects were unaware that any character string had been presented. For both old and new words on the recognition memory test, a matching context word increased the probability of judging an item old, while a mismatching word decreased the probability of judging the item old. The matching word facilitated perceptual processing of the following test word and so increased subjects' feeling of familiarity. In the

case of new words, the brief presentation of a matching context word created an illusion of memory. An important control condition in the Jacoby and Whitehouse experiment illustrates how these illusions of memory depend on an inference or an attribution about the source of easy perceptual processing. In a second condition, the matching or mismatching context words were presented for much longer so that subjects were fully aware of them. When subjects were aware of the context words, they were actually less likely to call either an old or a new recognition test word old when the context word matched the test word than when no context word or a mismatch context word was presented. When they were unaware of the context word, people mistakenly attributed their enhanced processing of the test word to having studied it and so judged it old. In contrast, when aware of the context word, people correctly attributed their enhanced processing of the test word to having just read it as a matching context word. In fact, subjects in the aware condition tended to overcorrect for the effect of the matching word and so were less likely to judge the test word old than if no context word had been presented.

Whittlesea et al. (1990) manipulated the fluency of perceiving the test words shown on a recognition test by masking those words with greater or lesser amounts of visual noise. Words made relatively easy to perceive because of less visual noise were more likely to be judged as having appeared in a studied list of words, whether or not the target words were actually old or new. When subjects were made aware that fluency of perception was being manipulated by the experimenter, they no longer interpreted variations in perceptual fluency as familiarity, but instead correctly interpreted those variations as due to changes in visual clarity.

These experiments illustrate how ease of perception can be manipulated to produce illusions of familiarity. Similarly, variations in the ease of more conceptual analyses may be interpreted as familiarity. Whittlesea (1993) tested this possibility by manipulating the conceptual analysis of recognition test words. Each trial in the experiment began by having subjects read a rapidly presented list of seven words (66 milliseconds per word). Following the last word, there was a brief pause, and then a recognition test word was presented as the last word in a sentence. The sentences were constructed to make the last word more or less predictable: "The stormy seas tossed the BOAT" or "He saved up his money and bought a BOAT." (Previous experiments in the series confirmed that subjects could generate the last word of the sentence more rapidly for the more predictable sentence contexts.) Subjects were more likely to judge the target word as from the list of seven briefly presented words when it was the last word in a more predictable sentence context, whether or not the target word actually had been presented in the list of seven words. That is, subjects mistakenly attributed the ease of reading the word due to the predictable sentence

context as familiarity due to having read the word in the seven-item study list: they experienced illusions of familiarity.

In addition to perceptual and conceptual ease of processing as a basis for the feeling of familiarity, other qualities of current experience can lead to the inference that one is remembering rather than imagining. For example, events are embedded in a train of events, so the ability to remember what happened before and after a particular memory bolsters a feeling that it is a real memory (Johnson, 1988). The vividness of an image can also be a cue that one is remembering rather than imagining. Brewer (1988) studied students' ability to remember events up to 55 days later. Subjects were most confident that they were remembering when they had vivid visual imagery for the event. However, just as perceptual and conceptual fluency on a memory test can be manipulated to produce illusions of memory, the vividness of images and the ease of elaborating on a "memory" can be manipulated and so create illusions of remembering.

Hypnosis appears to create the experience of enhanced imagery (e.g., Crawford and Allen, 1983) and a feeling of effortless imagining. Thus people attempting to remember while hypnotized may be prone to interpret the images that come to mind as memories because they have the qualities of memories: vivid, detailed, and effortlessly generated. Dywan and Bowers (1983) found that people who had been hypnotized and attempted to recall pictures they had studied a week earlier were more confident in their memories, and that they did remember more pictures than their nonhypnotized counterparts. However, they also confidently recalled a number of items they had never studied (see also Laurence and Perry, 1983).

If memory is indeed an interpretation of current perception or imagination, then the social context of the rememberer can also contribute to the creation of illusions of memory. When one is attempting to remember something and other people, perhaps with some authority, treat the ideas produced as memories, then one might be more likely to also treat those ideas as memories. Loftus (1992) described a series of demonstrations that memories can be created through suggestion. Subjects were younger relatives of the experimenters. The experimenters essentially reminisced with the subject about shared childhood events over several days. However, one of the reminiscences that was introduced was a false memory of the subject getting lost on a shopping trip. Chris, the younger brother of one experimenter was told the following story:

> It was 1981 or 1982. I remember that Chris was 5. We had gone shopping at the University City shopping mall in Spokane. After some panic, we found Chris being led down the mall by a tall, oldish man (I think he was wearing a flannel shirt). Chris was crying and holding the man's hand. The man explained that he had found Chris walking around crying his eyes out just a few moments before and was trying to help him find his parents.

Initially, Chris said he didn't remember getting lost at the mall. But after several days of working to recall the real memories and this false memory, he began to add in details to the "lost in the mall" event and to feel that he was remembering.

Loftus notes that children and their parents share a fear that the child will get lost—there are even children's books on the topic, such as one in which the Sesame Street character Ernie gets lost in the mall. If people are attempting to recall details of an event that someone else seems to remember clearly, they may have a ready source of details from books, movies, or other people's lives that allow them to generate details. In a social context that biases one to interpret those details as remembered, they may take on a feeling of familiarity.

Loftus points out how similar social dynamics could occur in therapy sessions when a therapist believes that a client has been sexually or physically abused as a child and sets out to assist that client in uncovering repressed memories of that abuse. Some therapists assess their clients' symptoms and inform the clients that the symptoms are suggestive of childhood abuse or trauma of some sort. Popular books on recovering repressed memory of abuse also suggest that people give up their normal criteria for judging whether a memory really happened, suggesting that the memories recovered will not feel like normal memories. The combination of certainty from an expert that "you have the kind of symptoms I've seen in people who have been abused as children" and the suggestion that any image of abuse that comes to mind should be accepted as a memory may produce a potential for the creation of false memories.

There is increasing evidence against a "naive realist" view of memory as an objective record of the past. Instead, the feeling of remembering depends on a particular interpretation of evidence, such as the ease of generating images, the vividness of those images, and the social support for the "memory." People can mistake memories of imagined events or fictional events for events that really happened to them, and they can mistake current imagining for remembering.

CONCLUSIONS

In our previous book, we concluded (Druckman and Bjork, 1991:47):

> The effectiveness of a training program should be measured not by the speed of acquisition of a task during training or by the level of performance reached at the end of training, but, rather, by learner's performance in the posttraining tasks and real-world settings that are the target of training.

We amend that conclusion by noting that it is as important to educate learners about subjective experience as about performance. Subjective ex-

perience is of obvious relevance to motivation, and is also important as a basis for self-monitoring learning and performance. It is important that the conditions of practice be structured to reveal when understanding has not yet been achieved or when a procedure cannot yet be executed in conditions that differ from those experienced thus far in training. Experience in the setting that is the target of training serves to educate subjective experience of the learner as well as providing a measure of the effectiveness of training. In fact, it may be optimal to make the conditions of practice be more demanding, and, hence, more revealing to the learner, than are the anticipated real-world conditions.

Subjective experience does not faithfully reflect "situation-free" knowledge. Rather, subjective experience relies on an unconscious attribution or inference process and, consequently, can be misleading. The effects of memory can be misattributed to the characteristics of a present situation and give rise to unconscious plagiarism, mistakes in evaluating the level of comprehension of one's students, and so forth. The effects of factors that influence fluency, in turn, can be misattributed and experienced as remembering, such as the false recovery of a "memory." Much like the person who taps a song, everyone lives in a world that is, in part, of his or her own making, not realizing that subjective experience might rest on a misconstrual of the current situation.

That subjective experience is error prone might be taken as a good reason for banishing it as a concern, and for concentrating only on objective performance. However, subjective experience does not appear to its holder as being error prone. Although the instructor might attempt to banish the student's subjective experience, the student does not do so nor does the instructor banish his or her own subjective experience, and, rightfully, they should not. More "objective" bases for evaluation may not exist and some, such as a theory, are sometimes inferior to subjective experience as a basis for judgments.

More research is needed to better specify the cues that are used and the processes that are involved in the creation of subjective experience. Doing so is important to allow people to avoid the deleterious effects of misleading subjective experience. Of more immediate utility, people can be aware that subjective experience is largely unavoidable and serves important functions, but it cannot be fully trusted.