Attention and Performance XVII.
Daniel Gopher and Asher Koriat, editors.

13 The Strategic Regulation of Memory Reporting: Mechanisms and Performance Consequences

Morris Goldsmith and Asher Koriat

ABSTRACT In most real-life memory situations, as opposed to traditional laboratory settings, people have a great deal of freedom to control their memory reporting in accordance with personal and situational goals. For instance, they may choose to report only information they feel sure about, or they may choose to answer at a level of generality where they are unlikely to be wrong. In the present chapter, we focus on these two types of metamnemonic control, examining both their underlying mechanisms and their consequences for memory performance. We first describe a theoretical framework developed in our previous work (Koriat and Goldsmith 1996c) that addresses the monitoring and control processes underlying the decision to volunteer or withhold particular items of information in free-report situations. Simulation and empirical results demonstrate the critical role of these processes in allowing rememberers to strategically regulate the amount and accuracy of reported information. We then show how this framework can be extended to address how people control the level of generality or "grain size" of the information they report. Preliminary results indicate that here, too, people utilize their monitoring and control processes in a strategic manner, taking into account competing demands for both accuracy and informativeness. Finally, we discuss the more general implications of our work, and how it may be applied to the study, assessment, and enhancement of memory performance.

Q: Could you please tell us what you saw as you were getting out of your car?

A: I had just opened the door when I heard someone scream on the other side of the street. As I looked up, a man in a dark sweatsuit burst through the gate of the yard and ran at full speed down the alley. He seemed to be carrying a bag or something over his shoulder. I lost sight of him when he reached the end of the alley.

Q: Did you see which way he turned at the end of the alley?

A: I'm not sure.

Q: Do you remember what time it was?

A: Around 6 o'clock, maybe 6:30.

Q: Could you be more specific?

A: Between 6:15 and 6:30.

Although fictional, this short transcript illustrates some of the vast flexibility that people generally have in recounting past events from memory. There is no official "list" of input items that must be reproduced, as there is in traditional laboratory experiments. Instead, the person is free to choose which aspects of the event to relate and which to ignore, what perspective to adopt, how much detail to volunteer, what degree of confidence to impart, and so forth. Such decisions will naturally depend on a variety of personal and situational goals, whether these involve aiding a criminal investigation, succeeding on an exam, or impressing an experimenter or one's friends. Of course, in many situations one may find oneself pressed to report a particular piece of information or to give a more specific answer, but even then it is the rememberer who ultimately decides whether to provide the solicited information or reply "I don't know."

How has this type of personal control typically been handled in memory research? In general, experimental psychologists have shied away from tackling the implications of person-controlled regulatory processes in memory reporting, presumably because of the perceived conflict between the operation of these processes and the desire to maintain strict experimental control (see Nelson and Narens 1994). Thus one approach has been to take control away from rememberers, for instance by using forced-report testing techniques (Erdelyi and Becker 1974). Another alternative is to allow a small degree of personal control, but then to apply some sort of correction technique, such as those provided by the signal detection methodology (Banks 1970) or standard correction-for-guessing formulas (Budescu and Bar-Hillel 1993). A third approach is simply to ignore personal control, assuming that it has little effect on performance (see later discussion).

Our work, in contrast, is founded on the assumption that rather than constituting a mere methodological nuisance, personal control over memory reporting is in fact an intrinsic aspect of memory functioning. Hence it must be allowed to operate so that both its underlying dynamics and its performance consequences can be systematically investigated. In what follows, we present an integrative review of work we have done and new work in progress that demonstrates the utility of this approach. We focus on two different but related types of personal control: The first is report option, which involves the decision to volunteer or withhold specific pieces of information. The second is control over "grain size," that is, the choice of the level of generality at which remembered information is reported.

13.1 ROLE OF REPORT OPTION IN ACCURACY-ORIENTED MEMORY ASSESSMENT

To appreciate the role played by personal control over memory reporting, it is critical to distinguish between two different properties of memory—its quantity and its accuracy (Klatzky and Erdelyi 1985; Stern 1904). These two properties have received rather different emphases in contemporary approaches to memory and associated research practices (Koriat and Goldsmith 1996a,b): On the one hand, traditional memory research has been guided by a *storehouse* conception (Roediger 1980), evaluating memory primarily in terms of the number of stored items that can be recovered. On the other hand, the more recent wave of naturalistic, "everyday memory" research has inclined more toward a *correspondence* conception (Bartlett 1932), exhibiting a greater concern for the accuracy or "faithfulness" of memory in representing past events. Here the focus is on the extent to which memory reports can be *relied on* to provide accurate information.

Not coincidentally, the focus on quantity and accuracy has generally been associated with two different attitudes toward the issue of personal control over memory reporting. As mentioned earlier, personal control has not figured prominently in traditional quantity-oriented memory research, perhaps because of its incompatibility with the desire for strict experimental control (Banaji and Crowder 1989). It turns out, however, that in any case, personal control over memory reporting has little effect when memory performance is evaluated in terms of its quantity. Thus, for instance, offering participants monetary incentives to provide as many correct answers as possible does not increase their quantity performance relative to control participants who are not given any special incentive (e.g., Nilsson 1987; Weiner 1966a,b; but see Loftus and Wickens 1970). Also, studies investigating the effects of recall criterion (e.g., Bousfield and Rosner 1970; Britton et al. 1980; Cofer 1967; Erdelyi 1970; Erdelyi, Finks, and Feigin-Pfau 1989; Keppel and Mallory 1969; Roediger and Payne 1985; Roediger, Srinivas, and Waddil 1989) generally indicate that encouraging or forcing participants to recall more items does not improve their memory quantity performance much or at all, relative to standard free-recall instructions (e.g., Roediger and Payne 1985). The implication is that in quantity-oriented research, personal control over memory reporting might simply be ignored (Roediger, Srinivas, and Waddil 1989; but see Erdelyi, Finks, and Feigin-Pfau 1989).

In contrast, in the context of naturalistic, accuracy-oriented research, not only is personal control over memory reporting taken more seriously, but there is also an increased willingness to allow participants control over their memory reporting. One reason for this (beyond the methodological limitations sometimes imposed by naturalistic research contexts) is the common belief that personal control over memory reporting plays a critical role in eliciting accurate accounts from memory. Indeed, it is established wisdom is eyewitness research that witnesses should first be allowed to tell their story in their own words (i.e., in a free-narrative format) before being subjected to more directed questioning, and that even then, greater faith should be placed in the accuracy of the free-narrative type of testimony (e.g., Flanagan 1981; Hilgard and Loftus 1979; Timm 1983). This wisdom has been incorporated, for instance, into the "Cognitive Interview" technique (Fisher and Geiselman 1992), and also into various government documents concerning the proper way to interview witnesses (see Memon and Stevenage 1996).

Such attention notwithstanding, even in accuracy-oriented research the actual effects of personal control on memory performance are far from clear. One obstacle that has perhaps blocked progress is the general failure to distinguish between two distinct yet intertwined factors, *test format* and *report option*. Test format refers to whether rememberers produce their own answers (production or recall format), or instead must choose responses from a limited set provided by the questioner (selection or recognition format). It is this variable that has been emphasized, for instance, by the work of Loftus and her colleagues on the potentially harmful effects of postevent information (e.g., Loftus 1979; Loftus and Hoffman 1989; Loftus, Miller, and Burns 1978; Wagenaar and Boer 1987). That work is in part responsible for the belief that directed questioning or recognition testing can have contaminating effects on memory (see also Brown, Deffenbacher, and Sturgill 1977; Gorenstein and Ellsworth 1980; Lipton 1977). Report option (free versus forced), on the other hand, has to do with whether or not people are allowed to abstain from giving an answer (i.e., to respond "I don't know") or forced to answer each and every question.

Unfortunately, report option and test format are generally confounded in the reality of both naturalistic and traditional laboratory research. For instance, in free-narrative and recall testing, people produce their own answers (production format) and report only what they feel they actually remember (free report), whereas in directed questioning and recognition testing, people are not only confined to choosing between the alternatives presented by the interrogator (selection format), they are generally also induced, by implicit or explicit demands, to answer each and every question (forced report). Hence, it is not clear whether the commonly found accuracy advantage of recall (free-narrative) over recognition (directed questioning) is due to test format or report option. Also, it is not clear why, recall is generally superior to recognition in terms of accuracy performance, but recognition is generally superior to recall in terms of quantity performance, as testified to by a wealth of traditional laboratory research (see, for example, Brown 1976). We have called this pattern, the "recall recognition paradox" (Koriat and Goldsmith 1994).

In order to unravel this paradox and expose the contributing effects of report option, we conducted several experiments in which report option (free versus forced), test format (recall versus recognition), and memory property (accuracy versus quantity) were orthogonally manipulated. In one experiment (Koriat and Goldsmith 1994, experiment 1), we had participants answer 60 general-knowledge questions in a recall or a five-alternative recognition format (all items required a one-word answer in order to equate the "grain" of the answers across the two test formats). In addition to the standard tests of free recall and forced-choice recognition, however, two relatively uncommon procedures were added: forced recall (requiring participants to respond to all questions) and free recognition (permitting participants to skip items). A payoff schedule provided all participants with a common performance incentive, essentially rewarding them for each correct answer, but penalizing them by an equal amount for each incorrect answer.

Table 13.1 Mean Quantity and Accuracy Performance (Percentage Correct) as Function of Report Option (Free versus Forced) and Test Format (Recall versus Recognition) in Koriat and Goldsmith 1994, Experiment 1

Memory Property	Quantity		Accuracy	
Report option	Free	Forced	Free	Forced
Test format				
Recall	47.8	47.6	76.6	47.6
Recognition	61.5	67.0	76.9	67.0

Performance in all conditions was scored for both quantity and accuracy. Quantity was scored as the (input-bound) percentage of questions correctly answered, whereas accuracy was scored as the (output-bound) percentage of provided or selected answers that were correct. Note that under forced-report conditions, the two measures are operationally equivalent; both reflect the likelihood that an input question will be answered correctly. Under free-report conditions, however, the amount of volunteered information is generally less than the amount solicited, and hence the conditional output-bound accuracy measure uniquely reflects the correctness or *dependability* of the information that is *reported.*

The results are presented in table 13.1. When comparing the standard memory measures, free recall and forced recognition, the results produce the "paradoxical" pattern: recall is superior to recognition on the accuracy measure, but recognition is superior to recall on the quantity measure. Examination of the remaining means, however, shows that although memory quantity performance does vary with test format, recognition superior to recall, it is *report option* that is critical for memory accuracy. Free report increased accuracy performance substantially relative to forced report for both recall and recognition testing. In fact, under free-report conditions (in which memory accuracy can be distinguished from quantity), test format had no effect at all on memory accuracy: Given equal opportunity to screen their answers, the recall and recognition participants achieved virtually identical accuracy scores!

These results indicate that people can improve their memory accuracy performance considerably when they are allowed to control their own memory reporting, irrespective of the test format (and irrespective of the research context—naturalistic or laboratory). Similar results were obtained using a standard list-learning paradigm (Koriat and Goldsmith 1994, experiment 2).

Moreover, memory accuracy was found to be under strategic control. We compared performance under two levels of accuracy incentive. In a high-incentive condition (Koriat and Goldsmith 1994, experiment 3), participants received the same monetary bonus for each correct answer as in the first experiment, but forfeited all winnings if even a single incorrect answer was volunteered. These participants achieved substantially better accuracy for both recall (90.1%) and recognition (92.7%) compared to participants performing under the more moderate incentive (see table 13.1). Indeed, fully one-fourth of the high-incentive participants succeeded in achieving 100% accuracy! Thus the participants were able to adjust their memory accuracy in accordance with the operative level of accuracy incentive. The improved accuracy was accompanied by a corresponding reduction in quantity performance.

In sum, a comparison of the accuracy-oriented (naturalistic) and quantity-oriented (traditional) approaches to memory suggests that personal control over memory reporting may play very different roles in these two approaches. Whereas in quantity-oriented research, the effects of personal control have generally been regarded as negligible, in accuracy-oriented research, the effects of personal control on both memory accuracy and memory quantity performance can be substantial and hence must be taken more seriously. We now turn to the mechanisms of such control, and to a more systematic account of their performance consequences.

13.2 FRAMEWORK FOR FREE-REPORT MEMORY REGULATION

How can the strategic control of memory performance in free-report situations be conceptualized and investigated? In searching for a viable research approach, we first turned to signal detection theory (Green and Swets 1966; Swets, Tanner, and Birdsall 1961). Of course, signal detection theory has been very influential in bringing to the fore the role of person-controlled regulatory processes in memory responding (see, for example, Banks 1970; Lockhart and Murdock 1970; Norman and Wickelgren 1969). That framework has been used extensively to investigate the decision processes underlying forced-report recognition memory. Participants in the standard "old/new" recognition paradigm are assumed to set a response criterion on a continuum of memory strength in order to decide whether to respond "old" (studied) or "new" (foil) to any given test item. Depending on various further assumptions, two indices are typically derived: a measure of retention, d', and a measure of criterion level, β.

Unfortunately, however, the signal detection approach is not very helpful in dealing with the decision process underlying *free*-report memory performance, that is, with the decision whether to report an answer or to abstain. Indeed, under such conditions, the signal detection methodology cannot be properly applied (Lockhart and Murdock 1970). Our approach to the problem was therefore to extend the basic logic underlying signal detection theory to free-report situations (as others have done; see Klatzky and Erdelyi 1985), and to augment that logic with concepts and methods borrowed from the study of metamemory.

Figure 13.1 presents a simple model of how metamemory processes are used to regulate memory accuracy and quantity performance under free-report conditions (Koriat and Goldsmith 1996c). The model is deliberately

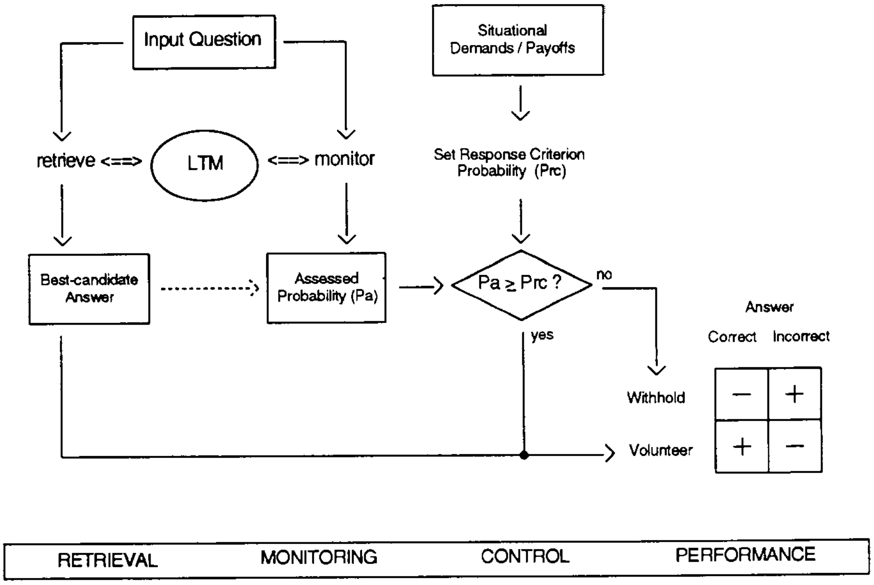


Figure 13.1 Schematic model of the strategic regulation of memory accuracy and memory quantity performance. The plus and minus signs signify positive and negative performance outcomes. Adapted from Koriat and Goldsmith 1996c.

schematic, focusing on the manner in which metamemory processes at the *reporting* stage affect the ultimate memory performance (cf. the distinction between "ecphory" and "conversion" in Tulving 1983). Thus, in addition to an unspecified retrieval (ecphory, reconstruction, etc.) mechanism, we posit a *monitoring* mechanism that is used to subjectively assess the correctness of potential memory responses, and a *control* mechanism that determines whether to volunteer the best available candidate answer (for a similar model, see also Barnes et al., chap. 10, this volume). The control mechanism operates as a threshold on the monitoring output: The answer is volunteered if it passes the threshold, but withheld otherwise. The threshold is set on the basis of implicit or explicit payoffs, that is, the gain for providing correct information relative to the cost of providing incorrect information.

Although the model is simple, its implications for memory performance are not; indeed, within this framework free-report memory performance depends on four contributing factors:

1. *Overall retention*—the amount of correct information (i.e., the number of correct candidate answers) that can be retrieved;
2. *Monitoring effectiveness*—the extent to which the assessed probabilities successfully differentiate correct from incorrect candidate answers;
3. *Control sensitivity*—the extent to which the volunteering or withholding of answers is in fact based on the monitoring output; and
4. *Response criterion setting*—the probability threshold that is set in accordance with the incentive to be accurate (e.g., payoff schedule).

Most previous treatments of the effects of the recall criterion, borrowing from signal detection theory, have focused on the first and fourth factors only (see, for example, Klatzky and Erdelyi 1985). The general assumption is that although people cannot increase the quantity of correct information they retrieve (e.g., Nilsson 1987), they can enhance the accuracy of the information they report, by withholding answers that are likely to be incorrect. Hence the widely acknowledged prediction is for a *quantity-accuracy trade-off*: raising the response criterion should result in fewer volunteered answers, a higher percentage of which are correct (*increased* output-bound *accuracy*), but a lower number of which are correct (*decreased* input-bound *quantity*). Because raising the response criterion is assumed to increase accuracy at the expense of quantity, the strategic control of memory performance requires the rememberer to weigh the relative payoffs for accuracy and quantity in reaching an optimal criterion setting.

It has not generally been noticed, however, that underlying this expected dynamic are two further implicit assumptions. First, of course, is the assumption that people do in fact volunteer and withhold information on the basis of subjective confidence. Although this assumption accords well with introspection, it nevertheless requires empirical verification (see below).

The second assumption is that people's probability assessments are reasonably, but not perfectly, diagnostic of the correctness of their candidate answers. The importance of this assumption has gone largely unnoticed. Indeed, although monitoring effectiveness has attracted much attention among students of metacognition (see Metcalfe and Shimamura 1994; Schwartz 1994), its performance consequences have only recently begun to be investigated (see, for example, Barnes et al., chap. 10, this volume; Bjork 1994; Metcalfe 1993; Nelson and Narens 1994).

The critical contribution of monitoring effectiveness to both memory accuracy and memory quantity performance emerged in several simulation analyses based on our model (Koriat and Goldsmith 1996c). These analyses assume a testing situation in which 50% of a hypothetical rememberer's candidate answers are correct (varying this percentage does not change the basic pattern of results), but both monitoring effectiveness and the response criterion are manipulated. Figure 13.2 depicts the accuracy and quantity performance that should ensue under the model from the use of various response criteria, assuming three different levels of monitoring effectiveness.

Consider first the plot for the "prototypical" monitoring condition (plot *B*). In this condition the rememberer's assessed probability judgments are assumed to be uniformly distributed across 11 confidence levels, ranging from 0 (certainly wrong) to 1.0 (certainly right). In addition, these judgments are assumed to be perfectly calibrated, that is, 20% of the answers with an assessed probability of .20 are correct, 30% of the answers with an assessed probability of .30 are correct, and so forth. People have generally been found to be well calibrated, though a tendency for overconfidence is often observed (see Lichtenstein, Fischhoff, and Phillips 1982). Under such conditions, rais-

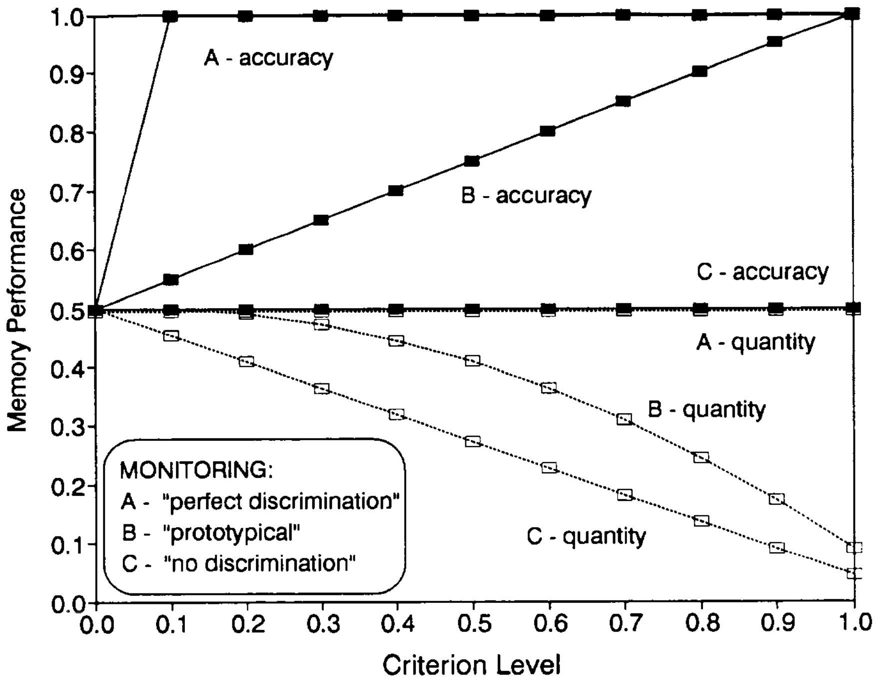


Figure 13.2 Simulated memory quantity and memory accuracy performance (proportion correct) plotted as a function of response criterion level, assuming three different levels of monitoring effectiveness (see text for explanation). Adapted from Koriat and Goldsmith 1996c.

ing the response criterion from 0 (forced report) to 1.0 yields the prototypical quantity-accuracy trade-off: accuracy increases but quantity decreases as the criterion becomes more strict.

Now, however, consider the plot for the "no discrimination" monitoring condition (plot *C*). In this condition, we assume that the rememberer's confidence judgments bear no relationship to the actual correctness of the answers. The person may believe that his or her judgments are diagnostic, but in fact the probability that an answer is correct is .50, regardless of the confidence attached to it. In this extreme case, the rememberer is unable to enhance his or her memory performance at all by exercising the option of free report: raising the response criterion does not increase accuracy performance, but simply decreases quantity performance.

Finally, consider the more positive extreme. In the "perfect discrimination" condition (plot *A*), we assume that the person can discriminate perfectly between correct and incorrect candidate answers.[1] Here all correct answers are assigned a subjective probability of 1.0, and all incorrect answers are assigned a probability of 0. Thus the ideal situation is reached, in which the option of free report allows the rememberer to achieve 100% accuracy with no quantity-accuracy trade-off: for any criterion level greater than zero (forced report), he or she will volunteer only correct answers and withhold only incorrect answers.

These simulations help illustrate the critical role assigned to two factors within the proposed framework: monitoring effectiveness and accuracy motivation. With regard to monitoring effectiveness, clearly some ability to distinguish between correct and incorrect candidate answers is necessary for the control of memory reporting to yield any benefits at all. Moreover, as this ability improves, greater increases in accuracy can be achieved at lower costs in quantity, so that at the extreme, when monitoring effectiveness is perfect, there is no quantity-accuracy trade-off at all.

As far as accuracy motivation is concerned, one can generally increase the accuracy of a memory report by employing a more conservative response criterion. However, under most monitoring conditions, enhancing one's accuracy becomes relatively costly in terms of quantity performance as the criterion level is raised (note the accelerated drop in quantity on the prototypical plot in figure 13.2). Thus simply giving a person the option of free report may allow a fairly large accuracy improvement to be achieved without much loss of quantity, but placing a larger premium on accuracy should lead to a more serious quantity reduction.[2]

More generally, when considering free-report memory performance, it is both necessary and useful to distinguish between the independent contributions of retention, monitoring, and control. Overall retention (50%, as indexed by forced-report performance at criterion = 0) was the same for all three conditions in figure 13.2. Yet the observed levels of free-report performance could vary drastically, depending on both the person's control policy (criterion level) and degree of monitoring effectiveness. We will return to these points again shortly.

Some Empirical Evidence

Do the monitoring and control processes in fact operate in the postulated manner? To test the basic assumptions of the model, we used a special procedure that combines both free and forced reporting (Koriat and Goldsmith 1996c, experiment 1). A general-knowledge test was administered in either a recall or a recognition format. (In order to control the grain level of the answers, all questions required either a single-word term or a proper name to be provided or selected.) The participants first took the test under forced-report instructions (phase 1) and provided confidence judgments regarding the correctness of each answer. Immediately afterward, they took the same test again under free-report instructions (phase 2), with either a moderate accuracy incentive (receiving a monetary bonus for correct answers, but paying an equal penalty for wrong answers) or a high accuracy incentive (in which the penalty was ten times greater than the bonus).

This design enabled us to trace the links postulated by the model (see figure 13.1) between retrieval, monitoring, control, and memory performance (for a more detailed treatment of this operationalizaton, see Koriat and

Goldsmith 1996c). Retrieval (recall or recognition) was tapped by treating the forced-report answers provided in phase 1 as representing the participants' best candidate responses for each item. Monitoring was tapped by treating the confidence ratings as representing the assessed probability associated with each best candidate answer. This allowed monitoring effectiveness to be evaluated. Control was tapped by examining which answers were volunteered or withheld on phase 2. This allowed us to determine the sensitivity of the control policy to the monitoring output, and to derive an estimate of the response criterion probability set by each participant. In addition, a comparison of the estimated criterion levels for the two incentive conditions allowed an examination of the predicted effects of accuracy incentive on the participants' control policy. Finally, the design allowed us to evaluate the contribution of monitoring and control processes to the ultimate free-report memory accuracy and memory quantity performance.

The results accorded well with the model. First, participants were found to be fairly effective in monitoring the correctness of their answers. Within-subject Kruskal-Goodman gamma correlations between confidence and correctness averaged .87 for recall and .68 for recognition. Second, the tendency to report an answer increased greatly with increased confidence in the answer. In fact, the gamma correlations between confidence and volunteering averaged .97 for recall and .93 for recognition! Third, participants who were given the high accuracy incentive were more selective in their reporting, adopting a stricter criterion than those given the more moderate incentive: they volunteered fewer answers on the average (26.9) than did the moderate-incentive participants (30.9), and mean confidence for those answers (.93) was higher than those volunteered by the moderate-incentive participants (.84). In addition, the "best-fit" estimate of the criterion used by each participant averaged .84 for the high-incentive condition versus .61 for the moderate-incentive condition.

Finally, by employing these monitoring and control processes, participants in both incentive conditions were able to enhance their free-report accuracy performance relative to forced-report. However, a quantity-accuracy tradeoff was observed both in comparing free- and forced-report performance, and in comparing performance under the two incentive conditions (see table 13.2).[3] Consistent with the simulation analyses, the quantity cost of the improved accuracy increased in relative terms when a higher criterion was employed. Whereas, under a moderate accuracy incentive, the option of free report enabled participants to enhance their accuracy substantially at a relatively low cost in quantity performance (a 64% accuracy improvement achieved at a 19% quantity cost for recall; a 33% accuracy improvement achieved at a 26% quantity cost for recognition), the introduction of a stronger accuracy incentive resulted in a further increase in accuracy, but now at a relatively high quantity cost (a further 12% accuracy improvement achieved at a 10% quantity cost for recall; a 6% accuracy improvement achieved at a 15% quantity cost for recognition, based on adjusted means).

Table 13.2 Mean Free-Report Quantity and Accuracy Performance (Percentage Correct) as Function of Test Format (Recall versus Recognition) and Accuracy Incentive (High versus Moderate); Forced-Report Performance (quantity or accuracy) as Function of Test Format in Koriat and Goldsmith 1996c, Experiment 1

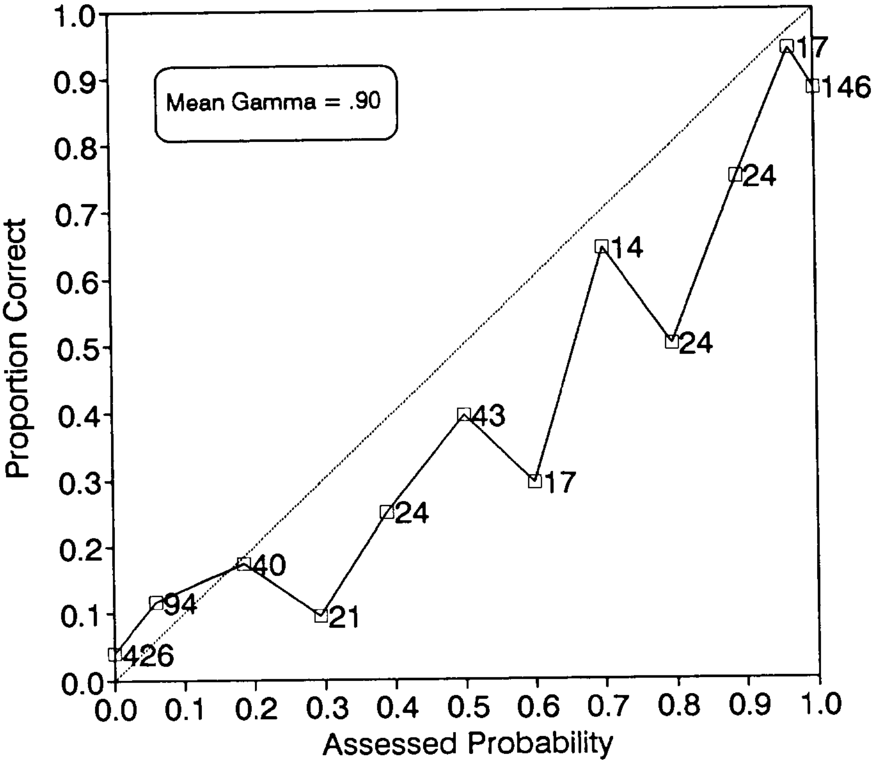
	Free Report (Phase 2)				Forced Report (Phase 1)
Property	Quantity		Accuracy		Quantity or Accuracy
Incentive	Moderate	High	Moderate	High	
Test format					
Recall	38.0	38.5	76.4	88.1	46.7
Recognition	47.3	43.6	84.4	91.1	63.5

A second experiment evaluated the role of monitoring effectiveness (Koriat and Goldsmith 1996c, experiment 2). That experiment used basically the same procedure as did the first (recall and moderate incentive only), but in addition, monitoring effectiveness was manipulated within participant by using two different sets of general-knowledge items. One set (the "poor" monitoring condition) consisted of items for which the participants' confidence judgments were expected to be generally uncorrelated with the correctness of their answers (see Fischhoff, Slovic, and Lichtenstein 1977; Gigerenzer, Hoffrage, and Kleinbolting 1991; Koriat 1995), whereas the other set (the "good" monitoring condition) consisted of more typical items, for which the participants' monitoring was expected to be more effective. The success of the manipulation can be verified by examining figure 13.3.

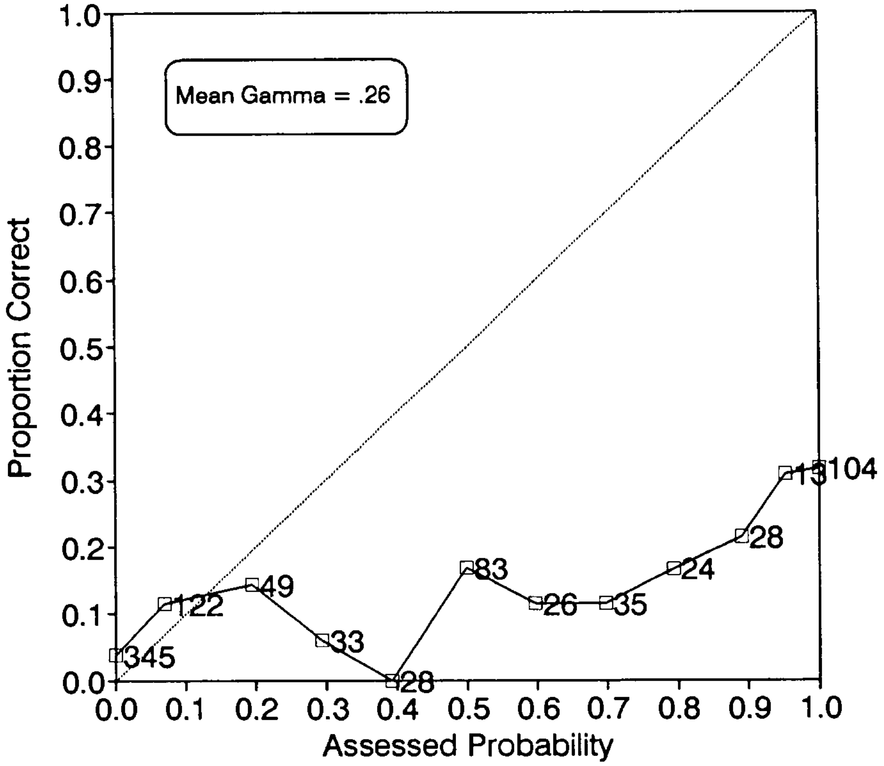
We expected that because they lack any better predictor, participants in both monitoring conditions would base their volunteering decisions on their monitoring output, and indeed, the gamma correlations between confidence and volunteering averaged .95 and .88 for the good and poor monitoring conditions, respectively. More importantly, even when the two sets were matched on retention (by adding some very difficult items to the good monitoring set) so that *forced*-report quantity performance was equivalent, the good monitoring condition allowed participants to attain a far superior joint level of *free*-report accuracy and quantity performance: much better accuracy performance was achieved while maintaining equivalent quantity performance, compared to the poor monitoring condition (see table 13.3).

These results, then, reinforce the earlier simulation results in highlighting the criticality of monitoring effectiveness for free-report memory performance. When people's confidence judgments are reasonably diagnostic of the correctness of their answers, the option of free report can allow them

Figure 13.3 Calibration plots for the good monitoring and poor monitoring conditions in Koriat and Goldsmith 1996c, experiment 2. The frequency of judgments in each category appears beside each data point, and the mean within-participant gamma correlations between assessed probability and actual correctness are also presented.



(a) Good Monitoring Condition



(b) Poor Monitoring Condition

Table 13.3 Mean Quantity and Accuracy Performance (Percentage Correct) for the Good Monitoring Condition (after Matching on Forced Report) and the Poor Monitoring Condition in Koriat and Goldsmith 1996c, Experiment 2

	Free Report		Forced Report
Property	Quantity	Accuracy	Quantity or Accuracy
Monitoring condition			
Good	8.6	63.0	11.2
Poor	7.6	21.0	11.8

to achieve high levels of accuracy. In other situations, however, people's monitoring may be undiagnostic to the point of being useless. People still control their memory reporting according to their monitoring output, but the attained level of free-report accuracy may be little better than when they are denied the option of deciding which answers to volunteer.[4]

Of particular importance is the demonstration that monitoring effectiveness can affect memory performance independent of memory "retention" (cf. figure 13.2). Even when retention, as indexed by forced-report quantity performance, was equated across the good and poor monitoring subtests in experiment 2, the joint levels of free-report accuracy and quantity performance were far superior for the good monitoring subtest than for the poor monitoring subtests (and see figure 13.4, later). Clearly, then, free-report memory performance depends on the effective operation of metamemory processes that are simply not tapped by forced-report performance.

Results from several other studies also suggest a dissociation between monitoring and retention. For instance, Kelley and Lindsay (1993) observed that advance priming of potential answers to general-information questions increased the ease of access to these answers, raising subjective confidence regardless of whether those answers were right or wrong. Similarly, research investigating the cue familiarity account of the feeling of knowing indicates that feeling-of-knowing judgments can be enhanced by advance priming of the cue, again even when such priming has no effect on actual memory quantity performance (e.g., Reder and Ritter 1992; Schwartz and Metcalfe 1992). Finally, Chandler (1994) found that exposing participants to an additional set of pictures similar to the studied set increased their confidence ratings on a subsequent forced-choice recognition test, while in fact their actual performance was impaired.

Such dissociations help to emphasize a basic difference between our proposed framework for conceptualizing the strategic regulation of memory reporting and the signal detection approach to memory. The signal detection framework does not address the separate contributions of memory retention (or memory strength) and monitoring effectiveness to memory performance. Subjective confidence and memory strength are generally treated as synonymous (Chandler 1994); indeed, confidence is often used to *index* memory

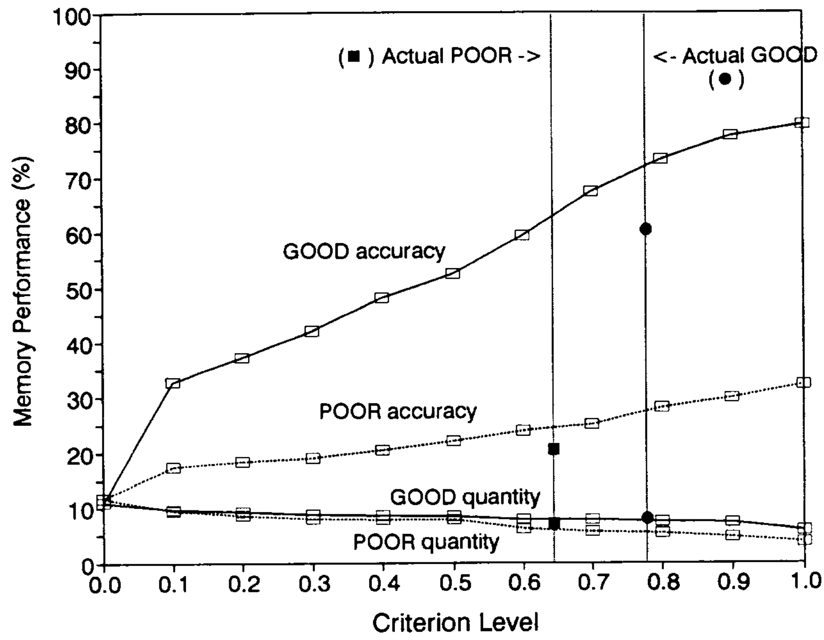


Figure 13.4 Illustrative quantity-accuracy profiles comparing performance under two levels of monitoring effectiveness. Potential free-report memory quantity and memory accuracy performance (mean percentage correct) is plotted as a function of criterion level for the participants in Koriat and Goldsmith 1996c, experiment 2, for both the good monitoring condition (after equating forced-report performance) and the poor monitoring condition. Actual free-report quantity and accuracy scores for each monitoring condition are also plotted as bullets above the criterion estimate for that condition.

strength (see, for example, Lockhart and Murdock 1970; Parks 1966). Thus, in the forced-report old/new paradigm to which signal detection methods are typically applied, "control" is isolated in terms of the parameter β. Yet "retention" (overall memory strength) and "monitoring effectiveness" (the extent to which the participant's confidence distinguishes "old" from "new" items) cannot be operationally or conceptually separated: both are equally valid interpretations of d' (see, for example, Banks 1970; Lockhart and Murdock 1970).

By contrast, in our proposed framework for conceptualizing free-report performance, retention and monitoring effectiveness (as well as control) are given a separate standing: one may have effective monitoring, yet very poor retention, or vice versa. Furthermore, poor free-report memory performance, for instance, could derive from poor retention, poor monitoring, an inappropriate control policy, or all three. The separation of these components of free-report performance has important theoretical and practical implications. Before these are considered, however, let us first examine how the proposed framework can be extended to encompass other types of strategic memory regulation.

Control over "Grain Size"

We now turn to a second means by which people can regulate their memory performance, control over the "grain" or level of generality of the reported information. Like the option of free report, this type of control is also quite prevalent both in real-life remembering and in accuracy-oriented memory research. It, too, may underlie some current empirical puzzles. To illustrate, consider a study reported by Neisser (1988), who tested students' memory for events related to a seminar that he taught, using either an open-ended recall format or a forced-choice recognition format. He found the recall format to yield more accurate remembering than the recognition format, and noted that this might come as a surprise to memory researchers who are accustomed to the general superiority of recognition testing over recall testing. As discussed earlier, such a finding can perhaps be explained by the effects of report option. Neisser, however, also pointed out a further consideration: whereas in the recognition format participants had to make relatively fine discriminations between correct and incorrect response alternatives, in the recall format they seemed to choose "a level of generality at which they were not mistaken" (Neisser 1988, 553).

Along similar lines, Fisher (1996a,b), in assessing participants' freely reported recollections of a filmed robbery, was recently surprised to find that both quantity performance (number of correct statements) and accuracy performance (proportion of correct statements) remained constant over two retention intervals covering a 40-day span. The anomaly was resolved by considering the grain size of the reported information: statements made after 40 days were as likely to be correct as those made sooner after the event (about 90% accuracy in both cases), but this seeming equivalence at the second testing was achieved by providing information of coarser grain (as rated by two independent judges) than that contained in the earlier statements.

Clearly, then, when rememberers control their own memory reporting, differences in the grain size of the reported information can pose a troubling methodological problem. Here, too, the traditional remedy has been to take control away from rememberers, for instance, by using recognition testing, or by using stimulus materials, such as word lists, that greatly limit the scope of the problem. (This, in fact, is what we tried to do in our studies focusing on the effects of report option). Nevertheless, we are sure that many memory researchers have shared our dilemma in deciding, for instance, how to score the word "bird" when it is reported instead of "robin" in a cued or free recall task (see also Bahrick and Bahrick 1964). In more naturalistic research contexts, such as those represented by Neisser's and Fisher's studies just mentioned, perhaps all one can do is be aware of the problem and try to find a way of "correcting" for grain size in comparing performance across different conditions.

Control over grain size, however, like report option, is more than just a mere methodological nuisance that needs to be circumvented or corrected for. In most real-life memory situations, it also constitutes an important means by which rememberers regulate the accuracy of their memory reporting, and as such, is an integral aspect of the process of remembering. The challenge, then, is to find a way to systematically investigate this type of control as well.

We have recently begun to extend our framework for the strategic regulation of memory reporting to include control over grain size, and our preliminary results are illuminating. The approach we chose is similar to the one we used for investigating report option, and in fact, assumes a close relationship between these two types of control.

Consider a situation in which, say, a woman who has witnessed an accident is asked to answer a set of questions that have to do with graded dimensions, such as the time of the accident, the speed of the car, the distance to the intersection, and so forth.[5] If the witness is forced to answer each question at a specified grain size (e.g., to the nearest minute, mile per hour, foot, etc.), then the accuracy of those answers may be quite poor, particularly if they are scored in a dichotomous, correct-incorrect fashion (rather than on a continuous dimension; see Koriat and Goldsmith 1996b). However, even though the witness may not remember, say, that the accident occurred precisely at 6:17, she may be able to report that it occurred between 6:00 and 6:30, or perhaps in the early evening. What, then, will happen if the witness herself is allowed to choose the grain size for her answers? Will she be able to exploit this option in an effective manner, adopting a coarser grain of reporting when a finer grain is likely to miss the mark?

Of course, a more coarsely grained response will always be the wiser choice if accuracy (i.e., the probability of including the true value) is the sole consideration. However, it seems likely that in conveying what they remember, people are motivated not only to be accurate, but also to be informative. This idea was brought out nicely by Yaniv and Foster (1995, 424–425), who proposed that the choice of grain size for judgmental estimates "involves a trade-off between two conflicting objectives: accuracy and informativeness.... Receivers prefer estimates that are both sufficiently informative for their current decision making and appropriately accurate. For example, the prediction that the inflation rate will be "0% to 80%" would not be appreciated by receivers, although it is likely to be confirmed."

Yaniv and Foster provide evidence that when accuracy is pitted against informativeness, receivers of information often prefer a somewhat inaccurate but relatively precise estimate to a completely accurate but uninformative response (as in the inflation example). Accordingly, when people are asked to give quantitative estimates under uncertainty, they tend to consider the receiver's desire to obtain a useful response (cf. Grice 1975), and will, if necessary, sacrifice accuracy for informativeness (Yaniv and Foster 1997).

This type of accuracy-informativeness trade-off may also underlie the chocie of grain size in memory reporting. We examined this idea in a recent experiment that included two phases. In the first phase, participants were presented with a list of general-knowledge questions, and were required to answer each of them at two different grain levels. For example, "When did Boris Becker last win the Wimbledon men's tennis finals? (A) Specify a 3-year interval; (B) Specify a 10-year interval." In some cases, the finer-grained response required a specific number to be provided in the appropriate units (e.g., a specific year), whereas in other cases (such as in the preceding example), it required that a relatively narrow interval be specified. In the second phase, the participants were asked to go over their answers again, and for each item to choose the answer (at one of the two grain levels) that they would prefer to provide if they were "an expert witness testifying before a government committee" (cf. Yaniv and Foster 1995).

Participants were found to prefer the finer-grained answer in 41% of the cases, implying that the choice of grain level was not guided solely by the desire to be accurate, nor solely by the desire to be precise. Apparently, the participants were willing to sacrifice informativeness in their responding when the more precise answer was deemed too unreliable: questions that they chose to answer at the coarse level were much less likely to be answered correctly at the fine-grained level (21%) than were those they chose to answer at the fine level (51%). Furthermore, by sacrificing informativeness, the participants improved their overall accuracy substantially (to 59%), compared to what they would have achieved by providing the fine-grained answers throughout (32%).

These results indicate that people are fairly efficient in choosing to answer at a coarser grain size when a finer-grained answer is likely to be wrong (though the participants were still wrong about half the time when they chose to remain at the fine-grained level). Presumably, the choice of grain level is guided by confidence in the accuracy of potential candidate answers. To check this idea further, we had a second group participate under similar conditions, but also give confidence ratings for their answers at both grain levels. The results can help clarify the earlier pattern. Indeed, the same pattern emerges in terms of both confidence and accuracy: on average, participants attached a much lower confidence rating (assessed probability) at the fine-grained level (.34) to items they chose to report at the coarse-grained level than they did to items they chose to report at the fine-grained level (.77). The corresponding accuracy rates at the fine-grained level were 21% and 54%, respectively. Apparently, then, the participants could use their confidence judgments to identify low-probability answers at the fine-grained level and improve their accuracy for these items (to 59%) by choosing the coarser grain size.

How effective were the participants in monitoring the accuracy of their answers? The average within-person gamma correlation between confidence and accuracy was .46 for the fine-grained answers and .48 for the coarse-grained answers. Although these modest correlations allowed fairly effective choices of grain size to be made in the present experiment, they are substantially lower than those we have obtained in our previous work on report option (see "Some Empirical Evidence" above). Also, a substantially greater degree of overconfidence was exhibited in the present experiment (see also Yaniv and Foster 1997). If such differences in monitoring accuracy between single-valued and interval-type answers turn out to be systematic, then we may expect corresponding differences in the efficiency with which people can regulate their memory performance.

Although these results are suggestive, they clearly only scratch the surface of what remains to be explored regarding the use of grain size in the strategic regulation of memory reporting. A basic issue concerns the nature of the control process itself. Perhaps the simplest model (which has underlain the discussion thus far) is to assume that rememberers begin with as fine-grained an estimate as possible, and then widen the interval until the assessed probability of their answers exceeds some minimum criterion threshold. Our preliminary results are consistent with this possibility: the average within-person gamma correlation between confidence in the fine-grained answer and the choice of grain level was .82. Moreover, a simple threshold model can account for 87% of the participants' actual choices (compared with a 92% success rate for report option; see Koriat and Goldsmith 1996c, experiment 1).[6] Also, this model assumes that people try to maintain a "reasonable" level of accuracy through their choice of grain size (cf. the earlier described anomaly reported by Fisher 1996a,b), and indeed, the accuracy rates for the items participants chose to report at the fine-grained and coarse-grained levels were fairly similar: 51% versus 67% in the first experiment, and 54% versus 59% in the second experiment. (Considering their degree of overconfidence, the participants may have been aiming at higher accuracy rates than these.)

A further possibility, however, is that people's choice of grain size is a combined function of both these assessed likelihood and the perceived informativeness of alternative candidate answers at various grain levels. For instance, people may consider a range of potential responses, choosing the grain level that they believe will maximize the "expected utility" of their answer in terms of both accuracy and informativeness (see Yaniv and Foster 1995). One way to distinguish these and other models empirically will be to provide participants with differential incentives for accuracy and informativeness, using manipulations similar to those we employed previously in the study of report option.

Finally, the relationship between the use of report option and control over grain size also needs to be clarified. Clearly, in most memory situations where people can control the grain size of their answers, they are also free to withhold the answer entirely (i.e., to reply, "I don't know"). Which form of control will they choose to employ? In a sense, report option and grain size can be conceptualized as a lying on a continuum: withholding an answer is informationally equivalent to providing a very coarsely grained response (i.e., one that covers the entire range of possible answers). However, at some point, people will probably prefer to say, "I don't know," rather than provide a ridiculously coarse answer, even though the latter would provide a bit more information. Thus, for instance, one is unlikely to hear an eyewitness report that the height of an assailant was "between four and seven feet." Ultimately, the rememberers' assessments of the informational value of different answers, and their criteria for providing coarse-grained answers rather than withholding information completely, should depend on an assortment of pragmatic (Grice 1975), situational, and even personality factors. These types of variables, then, also become of interest when a decision-theoretic view of memory is adopted.

13.3 THE STRATEGIC REGULATION OF MEMORY REPORTING: THEORETICAL AND PRACTICAL IMPLICATIONS

The general theoretical framework advanced in this chapter is based on a certain view of remembering that should be made explicit: rather than simply retrieving and producing pieces of information from a passive memory store, rememberers are seen as active agents who manage their memory reporting in accordance with personal goals and situational demands. The process of remembering is dynamic, involving a variety of subjective assessments and decisions. For instance, rememberers must assess the likelihood that answers that come to mind are correct, the accuracy and informativeness of answers at different levels of generality, the probable costs and benefits of volunteering and withholding pieces of information, and so forth. On the basis of these assessments, they must then negotiate between competing considerations—accuracy versus quantity or informativeness—in attempting to choose the most appropriate response for the situation at hand. This decision-theoretic perspective has unique implications for the study of memory, the assessment of memory, and the applications of memory research.

The implications for the study of memory are straightforward. As discussed earlier, the traditional approach has been to try to minimize the contribution of report option and grain size to memory performance, so that remembering might be studied under as sterile and (experimenter-) controlled conditions as possible. If, however, the strategic regulation of performance is viewed as an intrinsic aspect of memory functioning, then rememberer-controlled regulatory processes must be allowed to operate, so that their underlying dynamics can be examined. Of course, one must still find a way to make these processes amenable to controlled scientific study. Thus, in our work, we tried to show how at least some rudimentary aspects of the dynamics of real-life memory reporting can be systematically investigated.

More generally, the view of rememberers as "strategic regulators" of memory performance should lead to a more serious consideration of functional and socioal-psychological factors. Most basic are the goals that the rememberer is trying to achieve (Neisser 1988, 1996; Winograd 1996). Remembering does not occur in a vacuum (even in psychology laboratories), but rather is typically used as a means toward some end. As Neisser (1996) has eloquently argued, remembering is like "doing." Hence any complete theory of memory "retrieval" will need to deal with "the reason for retrieval ... with persons, motives, and social situations." (Neisser 1988, 553). Thus understanding how (and why) people regulate their memory performance requires asking questions that are normally beyond the purview of memory research, such as how people evaluate the informativeness of different answers, how they perceive and respond to demands for accuracy versus informativeness, how they arrive at an effective control policy, and so forth (see Neisser 1996 for more far-reaching implications).

Beyond its implications for the study of memory, however, the view of the rememberer advanced in this chapter also raises a difficult issue with regard to memory assessment. How can we sensibly assess a person's memory if memory performance, particularly memory accuracy, is under the person's control? It seems that what is needed is a way to incorporate person-controlled regulatory processes into the assessment of memory performance, but still allow the independent contributions of these processes to be isolated. This is not a simple problem.

In the context of our work on report option, we developed a method that is based on the idea discussed earlier, that free-report memory performance can be partitioned into three separate components: retention, monitoring, and control. Unlike the standard point estimate measures of memory performance (e.g., percentage recall), the proposed *quantity-accuracy profile* (QAP) describes the joint levels of quantity and accuracy performance that can be achieved across a range of potential control policies (including the one actually used by the person) given the person's overall level of retention and monitoring effectiveness. The profiles resemble those derived in the earlier simulation analyses (see figure 13.2), but in this case the computed functions are based on the participant's actual confidence and performance data (for details, see Koriat and Goldsmith 1996c). Two illustrative profiles at the group level comparing the good and poor monitoring conditions in the experiment reported earlier (cf. table 13.3) are presented in figure 13.4.

As a supplement to the standard free-report quantity and accuracy measures, the QAP assessment methodology discloses additional information about retention (forced-report performance), monitoring effectiveness (gamma correlations, potential accuracy levels, and quantity-accuracy trade-off rates), and control (estimated response criterion and control sensitivity). It also encourages the inclusion of functional considerations relating to the goals of the rememberer (or of the evaluator) in order to assign a relative weight to the observed levels of quantity and accuracy performance. However, the method will need to be extended to incorporate control over grain size as well.

Turning to some more specific implications, the basic approach to memory research and assessment that follows from our work can be applied to a number of topics that are currently of major theoretical and practical interest. Most prominent is the study of eyewitness testimony. Can witnesses "tell the whole truth and nothing but the truth," as they are sworn to do? Our work indicates that they probably cannot; in most cases, people will have to trade quantity for accuracy in reporting what they remember. However, the work also points out some of the variables that need to be taken into account to answer that question more definitively in any particular case:. Is the witness free to control his or her own memory reporting, or are there implicit or explicit pressures to provide an answer to all questions (report option) and/or to be precise (grain size)? How well can the witness distinguish between what he or she knows and doesn't know (monitoring effectiveness)? How motivated is the witness to be accurate rather than informative and complete in his or her testimony (accuracy incentive)?

Taking such factors into account can enrich the study of memory accuracy in forensic settings as well as other contexts. The framework might also be applied to address possible differences in the reliability of people's memory in certain special populations, such as children, the elderly, or patients suffering from various kinds of brain damage. Thus, for instance, rather than simply ask whether children are less reliable witnesses than adults (e.g., Ornstein, Gordon, and Baker-Ward 1992), one can ask whether children are as effective as adults in using their metamemory processes to regulate their memory performance. Various deficits in children's metamemory processes have been noted (see Schneider, chap. 17, this volume; Schneider and Pressley 1989). Yet data collected in our laboratory indicate that children as young as 8 to 9 years of age can exercise the option of free report to enhance the accuracy of their memory reports. These children, however, were unable to achieve the level of accuracy (and quantity) performance attained by a group of older children (11- to 12-year-olds). Is this failure due to a difference in the degree of retention, the control policy, monitoring effectiveness, or some combination of these factors? We hope to provide some answers to these questions in future work.

Yet another relevant topic is the problem of psychometric testing. Many standard aptitude and achievement tests, as well as high school and college exams, allow and even encourage the test taker to exercise discretion in choosing which questions to answer and which to pass up or put off till the end. It is not generally acknowledged, however (except by the designers of courses and materials that help students prepare for taking exams such as the SAT and GRE), that performance on such tests therefore depends on the test taker's metacognitive skills and strategies as well as on domain knowledge or aptitude per se (see, for example, Budescu and Bar-Hillel 1993). In testing people's abilities, one might try to isolate these different components, or at least take them into account in designing procedures that reflect the test developer's actual psychometric goals (Koriat and Goldsmith 1998).

Finally, one last implication concerns the prospect of finding ways to improve memory performance by enhancing the efficiency of people's memory reporting. Much as traditional mnemonic techniques have focused on the improvement of encoding and retrieval processes (e.g., Fisher, chap. 19, this volume; Fisher and Geiselman 1992; Herrmann 1993), we may envisage corresponding techniques designed to enhance monitoring effectiveness (see also Bjork 1994; Druckman and Bjork 1994; Koriat 1997) and engender an optimal control of memory reporting. In line with this idea, for example, Johnson and colleagues (e.g., Dodson and Johnson 1993; Lindsay and Johnson 1989) have found that encouraging people to make more stringent source discriminations can reduce the harmful effects of misinformation in eyewitness testimony, and Memon and colleagues (Memon et al. 1996) report that children's tendency to employ the "don't know" option correlates positively with the accuracy of their testimony. However, to provide more reliable guidance to memory trainers, interrogators, and jurors, memory researchers will need to gain a much more detailed understanding of the underlying processes than has been achieved so far. It would seem, then, that the strategic regulation of memory reporting poses some important challenges for both theoretical and applied memory research, challenges that have only begun to be addressed.

NOTES

The major part of the reported work was supported by grant 032-1106 (1990) from the Israel Foundations Trustees. We thank Amit Weinberg Eliezer for her help in collecting the data on "grain size." The work was carried out at the Institute of Information Processing and Decision Making, University of Haifa.

1. It is important to distinguish between two different indices of monitoring effectiveness, *calibration* and *discrimination accuracy* (or resolution; see, for example, Lichtenstein, Fischhoff, and Phillips 1982; Liberman and Tversky 1993; Yaniv, Yates, and Smith 1991; Yates 1982, 1990). Calibration captures the absolute correspondence between subjective probabilities and the actual proportions of correct answers. Perfect calibration, however, does not entail perfect monitoring effectiveness at the level of individual answers. For instance, even though a person may be well calibrated in that, for example, among all items assigned a probability of .60, exactly .60 are correct, such a situation in fact means that the subjective monitoring is not effective enough to differentiate the 60% correct responses from the 40% incorrect responses included in this category. Thus it is discrimination accuracy (relative correspondence) that is more critical for the effective operation of the control mechanism. When assessed probabilities are polarized between the 0 (certainly wrong) and 1.0 (certainly right) categories, perfect calibration entails perfect discrimination accuracy at the level of individual items. Note, however, that the same discrimination accuracy would be obtained even when the probability values assigned to the two categories were, say, .40 and .41, in which case calibration would be very poor.

2. This pattern, together with the role of monitoring effectiveness, could perhaps explain Roediger, Srinivas, and Waddil's observation (1989, 255) that a recall criterion effect on quantity performance is "intuitive, but remarkably little evidence for it exists." See also Koriat and Goldsmith 1996c.

3. Due to sampling error, participants in the high-incentive condition yielded a higher quantity score (57.3%) than did moderate-incentive participants (52.5%) on phase 1. When this initial difference was partialled out in an analysis of covariance, the difference in free-report quantity performance for the high-incentive and moderate-incentive conditions was significant (adjusted means: 39.1% versus 44.6%, respectively, across recall and recognition). The incentive effect on accuracy remained significant (adjusted means: 88.7% for the high incentive versus 81.4% for the moderate incentive).

4. There have been several reports in the literature indicating situations in which memory monitoring may be rather poor. For example, Cohen (1988) found that although participants were quite accurate in monitoring the recallability of studied words, their judgments of the recallability of self-performed tasks had no predictive validity whatsoever (for a somewhat different example, see Metcalfe and Wiebe 1987). Koriat (1995), using deceptive items such as those used here (see also Fischhoff, Slovic, and Lichtenstein 1977), found that feeling-of-knowing judgments after unsuccessful recall were either uncorrelated or even negatively correlated with subsequent recognition memory performance. Weingardt, Leonesio, and Loftus (1994) found exposure to postevent misinformation to impair the relationship between confidence and the accuracy of people's answers (see also Chandler 1994). Finally, there is evidence that monitoring abilities may be relatively poor in certain special populations, for example, young children (e.g., Pressley et al. 1987), Korsakoff patients (e.g., Shimamura and Squire 1986), and patients with frontal lobe lesions (e.g., Janowsky, Shimamura, and Squire 1989).

5. It is methodologically convenient to operationalize grain size in terms of the range or interval width used in reporting quantitative information (see, for example, Huttenlocher, Hedges and Bradburn 1990; Huttenlocher, Hedges, and Prohaska 1988; Yaniv and Foster 1995, 1997). We provisionally assume that other forms of control over grain size (e.g., generification or abstraction) should operate according to basically similar principles, though ultimately this assumption will need to be tested.

6. Taking into account confidence in the coarse-grained answer as well as confidence in the fine-grained answer in a logistic regression analysis did not add to the predictive power of the model.

REFERENCES

Bahrick, H. P., and Bahrick, P. O. (1964). A re-examination of the interrelations among measures of retention. *Quarterly Journal of Experimental Psychology, 16,* 318–324.

Banaji, M. R., and Crowder, R. G. (1989). The bankruptcy of everyday memory, *American Psychologist, 44,* 1185–1193.

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin, 74,* 81–99.

Bartlett, F. C. (1932). *Remembering.* Cambridge: Cambridge University Press.

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

Bousfield, W. A., and Rosner, S. R. (1970). Free vs. uninhibited recall. *Psychonomic Science, 20,* 75–76.

Britton, B. K., Meyer, B. J., Hodge, M. H., and Glynn, S. M. (1980). Effects of organization of text on memory: Tests of retrieval and response criterion hypotheses. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 620–629.

Brown, E. L., Deffenbacher, K. A., and Sturgill, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology, 62,* 311–318.

Brown, J., Ed. (1976). *Recall and recognition.* London: Wiley.

Budescu, D., and Bar-Hillel, M. (1993). To guess or not to guess: A decision-theoretic view of formula scoring. *Journal of Educational Measurement, 38,* 277–291.

Chandler, C. C. (1994). Studying related pictures can reduce accuracy, but increase confidence, in a modified recognition text. *Memory and Cognition, 22,* 273–280.

Cofer, C. N. (1967). Does conceptual organization influence the amount retained in immediate free recall? In B. Kleinmuntz (Ed.), *Concepts and the structure of memory,* pp. 181–214. New York: Wiley.

Cohen, R. L. (1988). Metamemory for words and enacted instructions: Predicting which items will be recalled. *Memory and Cognition, 16,* 452–460.

Dodson, C. S., and Johnson, M. K. (1993). The rate of false source attributions depends on how questions are asked. *American Journal of Psychology, 106,* 541–557.

Druckman, D., and Bjork, R. A., Eds. (1994). *Learning, remembering, believing: Enhancing human performance.* Washington, DC: National Academy Press.

Erdelyi, M. H. (1970). Recovery of unavailable perceptual input. *Cognitive Psychology, 1,* 99–113.

Erdelyi, M. H., and Becker (1974). Hypermnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology, 6,* 159–171.

Erdelyi, M. H., Finks, J., and Feigin-Pfau, M. B. (1989). The effect of response bias on recall performance, with some observations on processing bias. *Journal of Experimental Psychology: General, 118,* 245–254.

Fischhoff, B., Slovic, P., and Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 552–564.

Fisher, R. P. (1996a). Implications of output-bound measures for laboratory and field research in memory. *Behavioral and Brain Sciences, 19,* 197.

Fisher, R. P. (1996b). Personal communication, May, 1996.

Fisher, R. P., and Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview.* Springfield, IL: Thomas.

Flanagan, E. J. (1981). Interviewing and interrogation techniques. In J. J. Grau (Ed.), *Criminal and civil investigation handbook,* Part 4, pp. 3–24. New York: McGraw-Hill.

Gigerenzer, G., Hoffrage, U., and Kleinbolting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98,* 506–528.

Gorenstein, G. W., and Ellsworth, P. C. (1980). Effect of choosing an incorrect photograph on a later identification by an eyewitness. *Journal of Applied Psychology, 65,* 616–622.

Green D. M., and Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. L. Morgan (Eds.), *Syntax and semantics.* Vol. 3, *Speech acts,* pp. 41–58. New York: Academic Press.

Herrmann, D. J. (1993). Improving student memory. Seattle: Hogrefe and Huber.

Hilgard, E. R., and Loftus, E. F. (1979). Effective interrogation of the eyewitness. *International Journal of Clinical and Experimental Hypnosis, 27,* 342–357.

Huttenlocher, J., Hedges, L. V., and Bradburn, N. M. (1990). Reports of elapsed time: Bounding and rounding processes in estimation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 196–213.

Huttenlocher, J., Hedges, L. V., and Prohaska, V. (1988). Hierarchical organization in ordered domains: Estimating the dates of events. *Psychological Review, 95,* 471–484.

Janowsky, J. S., Shimamura, A. P., and Squire, L. R. (1989). Memory and metamemory: Comparisons between frontal lobe lesions and amnesic patients. *Psychobiology, 17,* 3–11.

Kelley, C. M., and Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32,* 1–24.

Keppel, G., and Mallory, W. (1969). Presentation rate and instructions to guess in free recall. *Journal of Experimental Psychology, 79,* 269–275.

Klatzky, R. L., and Erdelyi, M. H. (1985). The response criterion problem in tests of hypnosis and memory. *International Journal of Clinical and Experimental Hypnosis, 33,* 246–257.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General, 124,* 311–333.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126,* 349–370.

Koriat, A., and Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General, 123,* 297–316.

Koriat, A., and Goldsmith, M. (1998). The role of metacognitive processes in the regulation of memory performance. In G. Mazzoni and T. O. Nelson (Eds.), *Metacognition and cognitive neuropsychology: Monitoring and control processes,* pp. 97–118. Hillsdale, NJ: Erlbaum.

Koriat, A., and Goldsmith, M. (1996a). Memory as something that can be counted versus memory as something that can be counted on. In D. Hermann, C. McEvoy, C. Hertzog, P. Hertel, and M. Johnson (Eds.), *Basic and applied memory research.* Vol. 2, *Practical applications,* pp. 3–18. Hillsdale, NJ: Erlbaum.

Koriat, A., and Goldsmith, M. (1996b). Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. *Behavioral and Brain Sciences, 19,* 167–188.

Koriat, A., and Goldsmith, M. (1996c). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490–517.

Liberman, V., and Tversky, A. (1993). On the evaluation of probability judgments: Calibration, discrimination, and monotonicity. *Psychological Bulletin, 114,* 162–173.

Lichtenstein, S., Fischhoff, B., and Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases,* pp. 306–334. Cambridge: Cambridge University Press.

Lindsay, D. S., and Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory and Cognition, 17,* 349–358.

Lipton, J. P. (1977). On the psychology of eyewitness testimony. *Journal of Applied Psychology, 62,* 90–95.

Lockhart, R. S., and Murdock, B. B. (1970). Memory and the theory of signal detection. *Psychological Bulletin, 74,* 100–109.

Loftus, E. F. (1979). *Eyewitness testimony*. Cambridge MA: Harvard University Press.

Loftus, E. F., and Hoffman, H. G. (1989). Misinformation and memory: The creation of new memories. *Journal of Experimental Psychology: General, 118*, 100–104.

Loftus, E. F., Miller, D. G., and Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 19–31.

Loftus, G. R., and Wickens, T. D. (1970). Effect of incentive on storage and retrieval processes. *Journal of Experimental Psychology, 85*, 141–147.

Memon, A., Holley, A., Wark, L., Bull, R., and Khnken, G. (1996). Reducing suggestibility in child witness interviews. *Applied Cognitive Psychology, 10*, 503–518.

Memon, A., and Stevenage, S. V. (1996). Interviewing witnesses: What works and what doesn't? *Psycholoquy*, [on-line serial] *7(6)*, witness-memory.l.memon.

Metcalfe, J. (1993). Novelty monitoring, metacognition, and control in a composite holographic associative recall model: Implications for Korsakoff amnesia. *Psychological Review, 100*, 3–22.

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

Metcalfe, J., and Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory and Cognition*, 15, 238–246.

Neisser, U. (1988). Time present and time past. In M. M. Gruneberg, P. Morris, and R. Sykes (Eds.), *Practical aspects of memory*. Vol. 2, *Current research and issues*, pp. 545–560. Chichester: Wiley.

Neisser, U. (1996). Remembering as doing. *Behavioral and Brain Sciences, 19*, 203–204.

Nelson, T. O., and Narens, L. (1994). Why investigate metacognition? In J. Metcalfe and A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing*, pp. 1–25. Cambridge, MA: MIT Press.

Nilsson, L.-G. (1987). Motivated memory: Dissociation between performance data and subjective reports. *Psychological Research, 49*, 183–188.

Norman, D. A., and Wickelgren, W. A. (1969). Strength theory of decision rules and latency in retrieval from short-term memory. *Journal of Mathematical Psychology, 6*, 192–208.

Ornstein, P. A., Gordon, B. N., and Baker-Ward, L. (1992). Children's memory for salient events: Implications for testimony. In M. L. Howe, C. H. Brainard, & V. F. Reyna (Eds.), *Development of long-term retention*, pp. 135–158. New York: Springer.

Parks, T. E. (1966). Signal detectability theory of recognition-memory performance. *Psychological Review, 73*, 44–58.

Pressley, M., Levin, J. R., Ghatala, E. S., and Ahmad, M. (1987). Test monitoring in young grade school children. *Journal of Experimental Child Psychology, 43*, 96–111.

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

Roediger, H. L. (1980). Memory metaphors in cognitive psychology. *Memory and Cognition, 8*, 231–246.

Roediger, H. L., and Payne, D. G. (1985). Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory and Cognition, 13*, 1–7.

Roediger, H. L., Srinivas, K., and Waddil, P. (1989). How much does guessing influence recall? Comment on Erdelyi, Finks, and Feigin-Pfau. *Journal of Experimental Psychology: General, 118*, 253–257.

Schneider, W., and Pressley, M. (1989). *Memory development between 2 and 20.* New York: Springer.

Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin and Review, 1,* 357–375.

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

Shimamura, A. P., and 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.

Stern, W. (1904). Wirklichkeitsversuche. *Beriträge zur Psychologie der Aussage, 2,* 1–31. Translation by U. Neisser. In U. Neisser (Ed.), *Memory observed: Remembering in natural context,* pp. 95–108. New York: Freeman and Company, 1982.

Swets, J. A., Tanner, W. P., Jr., and Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review, 68,* 301–340.

Timm, H. W. (1983). The factors theoretically affecting the impact of forensic hypnosis techniques on eyewitness recall. *Journal of Police Science and Administration, 11,* 442–450.

Tulving, E. (1983). *Elements of episodic memory.* Oxford: Clarendon Press.

Wagenaar, W. A., and Boer, H. P. A. (1987). Misleading postevent information: Testing parameterized models of integration in memory. *Acta Psychologica, 66,* 291–306.

Weiner, B. (1966a). Effects of motivation on the availability and retrieval of memory traces. *Psychological Bulletin, 65,* 24–37.

Weiner, B. (1966b). Motivation and memory. *Psychological Monographs: General and Applied, 80,* (18, Whole no. 626).

Weingardt, K. R., Leonesio, R. J., and Loftus, E. F. (1994). Viewing eyewitness research from a metacognitive perspective. In J. Metcalfe and A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing,* pp. 157–184. Cambridge: MIT Press.

Winograd, E. (1996). Contexts and functions of retrieval. *Behavioral and Brain Sciences, 19,* 209–210.

Yaniv, I., and Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General, 124,* 424–432.

Yaniv, I., and Foster, D. P. (1997). Precision and accuracy of judgmental estimation. *Journal of Behavioral Decision Making, 10,* 21–32.

Yaniv, I., Yates, J. F., and Smith, J. E. K. (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin, 110,* 611–617.

Yates, J. F. (1982). External correspondence: Decompositions of the mean probability score. *Organizational Behavior and Human Performance, 30,* 132–156.

Yates, J. F. (1990). *Judgment and decision-making.* Englewood Cliffs, NJ: Prentice Hall.