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

The Role of Metacognitive Processes in the Regulation of Memory Performance

Asher Koriat
Morris Goldsmith
University *of Haifa*

The increased interest in consciousness during the past two decades has opened once again the long-standing issue regarding the causal role of conscious experience in behavior. Is consciousness an inherent component of cognitive functioning, or is it merely an epiphenomenon that—from an information-processing standpoint—could just as well be done without. Among students of metacognition there is an implicit assumption that the subjective experience associated with monitoring one's own cognitive processing does in fact guide and regulate action. Thus, for instance, if we believe that we will not remember the name of a person who has just been introduced to us (a low *judgment of learning)*, we may take special measures to commit the name to memory. When we fail to recall that name in a subsequent encounter with the person, we will probably spend more time searching for it if we have a strong gut feeling that we should know the name (a strong *feeling of* knowing). And if the name "George" finally does come to mind, we will still hesitate to burst out with "Hello George" if we are not quite sure that that really is his name (a low *confidence judgment).*

In this chapter, we focus on the last stage of the process of remembering and examine the conditions that make one act on a retrieved piece of information. In particular, we address the question of how the metamemory processes of monitoring and control affect people's performance in tasks intended to tap memory.

THE CONTRIBUTION OF METAMEMORY PROCESSES TO MEMORY PERFORMANCE: SOME ILLUSTRATIVE EXAMPLES

When attempting to assess memory performance—for example, memory for the material studied during an academic course, or memory for a witnessed event—we may not he aware of the fact that we are often alsc tapping metamemory processes that are utilized by the rememberer in the service of achieving certain goals. Although the effects of such metamemory processes on memory performance are perhaps most clearly seen in functionally rich, naturalistic settings, they undoubtedly exist in more sterile laboratory contexts as well.

To illustrate the potential contribution of metamemory processes to memory performance and highlight the dilemma created by this contribution, consider the following example. You are grading a course examination that includes a set of short-answer questions, and find that on the exam you are currently checking, the answer to a particular question has been crossed out. Presumably, the question should be scored as an omission'. Nevertheless, being tempted to read the crossed-out answer, you find it to be perfectly correct (a situation that in our experience occurs more often than one might think). What should you do? Should you grant the student with full credit? After all, it appears that he or she "knows" the correct answer. Or should the student be penalized because he or she apparently does not "know that he or she knows" the solicited information? More generally, should poor monitoring and unwise strategic decision making be allowed to influence the student's grade on the test, and ultimately, success in the course?

This example illustrates the point that in many situations memory performance depends not only on what might be termed *memory*, but also on monitoring and control processes that fall under the rubric of *metamemory* (see also Barnes, Nelson, Dunlosky, Mazzoni, 6k Narens, in press; Nelson & Narens, 1990; and other chapters in this volume). In this case, the student apparently first thought that he or she knew the answer (monitoring) and wrote it down (control), but then changed his or her mind about the correctness of the answer (monitoring) and decided to cross it out (control). The monitoring aspect, then, involves the subjective assessment of how likely it is that an answer that comes to mind is correct, whereas the control aspect concerns the operational decision to write down the answer

or withhold it, or to cross out an answer that one has just written or leave it for inclusion in the final scoring.

Note that in the situation just described there was no explicit penalty for a wrong answer, so the student would not have risked losing points by venturing an answer. Apparently other motivations were involved apart from the desire to get the highest score possible—perhaps the motivation to be accurate, not to make a fool of oneself, and so forth. In other cases, however, the test maybe designed such that the student is explicitly required to make a strategic choice, and here metamemory processes will certainly contribute to the final score. For example, the student may be required to answer only a subset of questions of his or her choice, say, four out of six. Here it is clear that the scores will be affected by the student's ability to choose the right questions to answer, which in turn should depend on his or her ability to monitor the likelihood of providing a correct and complete answer to each question. Two students who have the same degree of knowledge can attain different scores if they differ in their monitoring ability.

Now consider a further example. Many of the standard psychometric tests of intelligence and scholastic aptitude (e.g., the Scholastic Aptitude Test and the Graduate Record Examination subject tests) use a multiple-choice format in conjunction with *formula* scoring procedures (Thurstone, 1919) that are designed to discourage guessing and also to correct for it by levying a greater penalty for incorrect answers than for omissions. It is not always clear to test administrators that performance on such tests also taps metacognitive ability; that is, the ability to make effective decisions about whether to risk providing an answer to a question or instead to omit (Budescu &. Bar-Hillel, 1993). Thus, for instance, one test-taker may tend to guess on the basis of even a small amount of partial knowledge, whereas another may feel uncomfortable providing an answer about which he or she is unsure (Abu-Sayf, 1979; Gafni, 1990). One test-taker may be effective in distinguishing between answers that are more likely or less likely to be correct, whereas another test-taker may be less effective in discriminating between what he or she knows and does not know (Angoff, 1989; Budescu & Bar-Hillel, 1993). Should differences in the operation and effectiveness of the monitoring and control processes of these test-takers be allowed to influence their scores on these tests? Exactly what types of knowledge and abilities are these tests intended to measure?

Finally, turning to the many real-life situations in which people recount past events, the effects of metamemory processes on memory performance are even more apparent. In such situations, people generally have great freedom in deckling which pieces of information to report and which to omit, what perspective to adopt, what level of generality or detail to provide in reporting the various aspects of an event, and so on. Consider, for instance, an eyewitness to a crime. Here, too, the "scoring rule," so to speak, explicitly discourages commission errors and guessing. In fact, the oath taken by a person on the witness stand is "to tell the whole truth and nothing but the truth." Thus, the witness' testimony is likely to be mediated by metamemory processes in which he or she assesses the likelihood that various pieces of information that come to mind are correct, and decides whether or not to report them in accordance with the perceived functional incentives. When considering such testimony, we might be more inclined (than in the case of academic testing) to treat the underlying metamemory processes as being part and parcel of the person's "memory" itself. After all, the main concern in the courtroom is with the accuracy or dependability of the witness' report; that is, with the extent to which the testimony can be trusted (see Deffenbacher, 1991).

These examples bring to the fore two basic points. First, there are many situations in which memory reporting is mediated by the metamemory processes of monitoring and control, and little is known about how such processes operate or about the effects that such processes have on actual performance. Second, it is not generally realized by researcher; and test administrators that a principled decision should be made regarding whether or not to include the effects of metamemory processes in the assessment of memory performance. In fact, it is a rather complicated matter to decide how metamemory processes should be treated when assessing memory performance. We now turn to a review of work we have done that addresses both of these points.

THE ROLE OF REPORT OPTION IN MEDIATING ACCURACY-BASED AND QUANTITY-BASED MEMORY PERFORMANCE

When considering the role of metamemory processes in memory reporting, it is important to distinguish between two properties of memory, its quantity and its accuracy. In (act, our initial interest in the performance consequences of metamemory processes derived from an examination of some basic differences between quantity-oriented and accuracy-oriented memory research (Koriat & Goldsmith, 1994, 1996a, 1996b). Traditionally memory research has been guided by a *storehouse* metaphor of memory, Leading to

the evaluation of memory in terms of the amount of stored information that can be recovered. The more recent wave of naturalistic, "everyday" memory research, however, has inclined more toward a correspondence conception of memory, in which memory is evaluated in terms of its accuracy or faithfulness in representing past events. The ramifications of this shift for both memory research and memory assessment are complex and far reaching (Koriat &. Goldsmith, 1996a), and in fact, the correspondence metaphor can lead to an accuracy-oriented assessment approach that is qualitatively different from the traditional, quantity-oriented approach. For present purposes, however, we restrict our attention to memory accuracy and memory quantity performance as they are typically evaluated in a standard item-based assessment context.

In item-based memory assessment, the memory test or report is segmented into discrete items or propositions that can be dichotomously evaluated as either right or wrong, and that are generally given equal weight in computing the overall memory score. This, for instance, is the approach taken in the vast amount of memory research based on the list-learning paradigm (Ebbinghaus, 1964; Puff, 1982), as well as in much psychometric and educational testing (Cronbach, 1984). In this context, quantity-based and accuracy-based memory measures can be distinguished in terms of input-bound and output-bound measures, respectively: Quantity measures, traditionally used to tap the amount of studied information that can be recovered, are input-bound, reflecting the likelihood that each input item is correctly remembered (e.g., the percentage of studied words recalled or recognized). Accuracy measures, in contrast, evaluate the ***dependability*** of memory—the extent to which remembered information can be trusted to be correct. Hence, these measures are output bound: They reflect the conditional probability that each reported item is correct (e.g., the percentage of reported words that actually appeared in the studied list). Essentially, then, whereas input-bound measures hold the person responsible for what he or she fails to report, output-bound measures hold the person accountable only for what he or she does report.

Despite the different definitions of quantity-based and accuracy-based memory measures, there are conditions in which the two types of measures are operationally equivalent: The critical factor is report option; that is, whether or not participants are required to answer all items. When memory is tested through a ***forced-report*** procedure, memory quantity and accuracy measures are necessarily equivalent, because the likelihood of remembering each input item (quantity) is equal to the likelihood that each reported item is correct (accuracy). Accuracy and quantity measures can differ substantially, however, under *free-report* conditions, in which participants are implicitly or explicitly given the option either to volunteer a piece of information or to abstain (e.g., respond "I don't know"; Neisser, 1958). Most everyday remembering is of this sort. Also, in memory research, the most common example is the standard free-recall task, in which reporting is essentially controlled by the participant. Under free-report conditions, people tend to provide only information that they believe is likely to be correct, so that their performance is mediated by a decision process employed to avoid incorrect answers (Klatzky 6k Erdelyi, 1985; Koriat &. Goldsmith, 1994). Because the number of volunteered answers is generally smaller than the number of input items, the output-bound (accuracy) and input-bound (quantity) memory measures can vary substantially.

Report option is important however, not only because it allows memory accuracy to be operationally distinguished from memory quantity but also because it has a substantial effect on memory accuracy performance. The contribution of report option was revealed in several experiments (Koriat &. Goldsmith, 1994) that orthogonally manipulated report option (free vs. forced reporting), test format (open-ended or cued recall vs. multiple-choice recognition), and memory measure (accuracy vs. quantity). In one experiment (Koriat & Goldsmith, 1994, Experiment 1), we gave participants a 60-item general-knowledge test, in which all answers were either single-word terms (e.g., photosynthesis) or proper names (e.g., Mozart). In addition to the standard methods of free recall[1] and forced recognition, we also included the less common procedures of forced recall (in which the participants were required to answer all items) and free recognition (in which the participants were allowed to skip over items). Both quantity and accuracy scores were derived for all four methods. The results indicated that although test format was the critical factor affecting memory quantity performance, recognition superior to recall, it was report option that was the critical factor affecting memory accuracy: First, free-report accuracy performance was substantially better than forced-report performance for both the recall and the recognition test formats. Second, under free-report conditions, in which the recall and recognition participants had equal opportunity to screen their answers, the recognition and recall accuracy scores were virtually identical. This basic pattern was also obtained using a

We use the term *free recall* in opposition to *forced recall,* in order to denote the option of *free report.* In traditional usage, however, the former term has been employed in opposition to serial recall, indicating only that the individual is free lo choose the order in which items are to he recalled.

standard list-learning paradigm (Koriat & Goldsmith, 1994, Experiment 2) and when the participants were given a very strong incentive for accuracy (Koriat & Goldsmith, 1994, Experiment 3).

These results suggest that memory accuracy performance can be improved considerably when people are allowed to control their own memory reporting. Across the three experiments, the accuracy advantage of free over forced report ranged from 61 % to 89% for recall and from 15% to 38% for recognition. Furthermore, given the option of free report, people can apparently adjust their memory accuracy in accordance with the operative level of accuracy incentive: When our free-report participants were given a very high accuracy incentive (receiving a monetary bonus for each correct answer, but forfeiting all winnings if even a single incorrect answer was volunteered), they improved their accuracy performance substantially compared to performance under a more moderate incentive (in which the penalty for each incorrect answer equaled the bonus for each correct answer). In fact, fully one fourth of the high-incentive participants succeeded in achieving 100% accuracy. The improved accuracy, however, was accompanied by a corresponding reduction in quantity performance (i.e., in the number of correct answers provided or selected).

What are the implications of such findings for the role of metamemory processes in mediating memory performance? Perhaps the most basic implication is that one cannot simply ignore the operation of these processes, particularly as far as memory accuracy is concerned.

Let us go back and consider some of the earlier memory examples. Clearly, to elicit accurate testimony from witnesses, they should be allowed to tell their story under free-report conditions; that is, they should be encouraged to say "I don't remember" if they feel they do not remember. As a matter of fact, this idea has been incorporated into most witness interview guidelines (e.g., Fisher & Geiselman, 1992; Flanagan, 1981; Hilgard & Loftus, 1979), in which interrogators are generally cautioned against putting words in the witness' mouth or pressing the witness for an answer. The advice is to allow the witness to tell his or her story first in a free narrative format before moving on to more directed forms of questioning, and even then to place greater faith in the accuracy of the former testimony than the latter[2] . In the early stages of an investigation, however, one might be interested primarily

[2]Actually, this recommendation regarding the use of free-narrative questioning stems from a great deal of eyewitness research that has focused on lest format rather than on report option: Directed questioning and recognition tests are held to he more likely than open-ended questioning and recall tests to contaminate a person's memory with information contained in the questions themselves (see, e.g., Gorcnstein & Ellsworth, 1980; Hilgard ft Loftus, 1979; Loftus, 1979; Loftus <St Hoffman, 1989).

in extracting as much information from the witness as possible (to obtain potential "leads"), even if some of that information turns out to be incorrect. In that case, it may be necessary to find ways to prevent witnesses from employing their natural memory screening processes (Fisher &. Geiselman, 1992), because unless those processes are employed with perfect efficiency, the witness might unwittingly screen out correct (and crucial) information along with the incorrect answers (discussed later).

Similar concerns also arise in the more standard, quantity-oriented testing situations, such as in the academic, psychometric, and laboratory testing situations, discussed earlier. We considered one case, for instance, in which correct information was withheld (crossed out) due to faulty monitoring processes and apparently hidden motivations, such as the mctivation to be accurate, that were ostensibly extraneous to the task. How much more so should memory performance be subject to the vicissitudes of :he test-taker's monitoring and control processes in those cases in which the person is explicitly encouraged to employ such processes, as when the test-taker is allowed to choose which questions to answer (e.g., in academic testing), or to omit answers to items about which he or she is unsure (e.g., under formula scoring procedures)?

From both a theoretical and a practical standpoint, then, it is important to achieve a better understanding of the operation of metamemory processes under free-report conditions and their effects on both accuracy-based and quantity-based memory performance.

THE MONITORING AND CONTROL PROCESSES UNDERLYING FREE-REPORT MEMORY PERFORMANCE

Figure 5.1 presents a simple model of how metamemory processes are used to regulate memory accuracy and quantity performance under free-report conditions (Koriat & Goldsmith, 1996b). Essentially, die model merges the logic of signal-detection theory (e.g., Banks, 1970; Green &. Swets, 1966; Lockhart &. Murdock, 1970) with concepts and tools from the study of metamemory. Thus, in addition to an unspecified memory retrieval mechanism, the model includes a monitoring mechanism that is used to subjectively

Unfortunately, however, report option and test format are generally **confounded** in both eyewitness and traditional laboratory research (see Koriat & Goldsmith, 1994). Thus, we should stress that in our research, in which test format and report option were orthogonally manipulated, report option was the critical factor affecting memory accuracy, and in fact, test format had no effect on memory accuracy at all (Koriat & Goldsmith, 1994, 1996h).

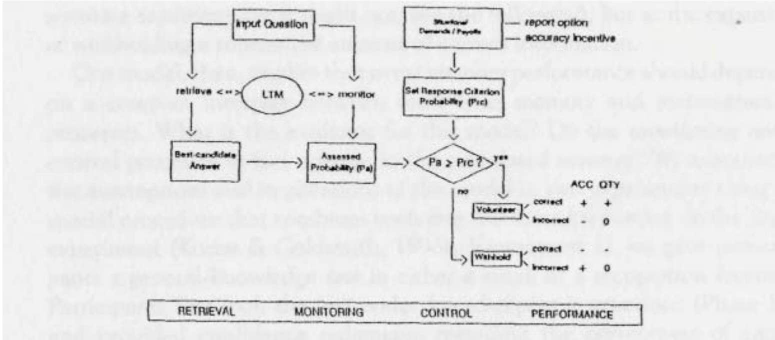

FIG. 5.1. A schematic model of how monitoring and control processes are used in the strategic regulation **of memory** performance. Effects on memory accuracy and memory quantity performance are signified by plus (increase), minus (decrease), and zero (no effect). Prom "Monitoring and Control Processes in the Strategic Regulation of Memory Accuracy, by A. Koriat and M. Goldsmith, 1996, *Psychological Review.* J03, (pp. 490-517. Copyright (e) 1996 by the American Psychological Association. Reprinted with permission.

assess the correctness of potential memory responses, and a control mechanism that determines whether or not to volunteer the best available candidate answer (see also Barnes et al., in press; Klatzky & Erdelyi, 1985). The control mechanism operates as a threshold on the monitoring output: The answer is volunteered if it passes the threshold, but is withheld otherwise. The threshold is set on the basis of the operative payoffs; that is, the gain for providing correct information relative to the cost of providing incorrect information.

Although the model's assumptions are quite simple, its implications for memory performance are not. Under the model, free-report memory performance can be shown to depend on several 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.
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 potential effects of selective reporting, borrowing from signal-detection theory, have focused on the first and fourth factors only (see, e.g., Bousfield & Rosner, 1970; Erdelyi, Finks, & Feigin-Pfau, 1989; Klatzky & Erdelyi, 1985; Roediger & Payne, 1985). Thus, the widely acknowledged prediction is for a quantity-accuracy trade-off: In general, 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). The assumption is that although people cannot increase the quantity of correct information that they retrieve (Nilsson, 1987), they can enhance the accuracy of the information that they report by withholding answers that are likely to be incorrect. Of course, the converse is also true, particularly in situations, such as in multiple-choice testing, in which there is a good chance of guessing the right answer. In such cases, by lowering the response criterion one should generally increase the number of correct (and incorrect) answers. Quite often, however, there will be explicit or implicit incentives for both quantity and accuracy. Therefore, because raising the response criterion generally increases 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.

Notwithstanding the importance of the control policy, however, it has gone largely unnoticed that both the accuracy gains and the quantity costs of selective reporting are heavily dependent on the effectiveness of the monitoring mechanism. Consider again a person on the witness stand. Assuming that the witness is not motivated to lie, can we expect him or her to be able to tell the whole truth and nothing but the truth? The answer will depend primarily on the person's monitoring effectiveness, which, we stress, is distinct both from the amount of information remembered and from the adopted control policy (response criterion level). For instance, a witness might remember very little of what happened, but if monitoring effectiveness is perfect, he or she will be able to volunteer all of the (correct) information that he or she remembers while screening out any potentially false information, yielding low quantity performance but high accuracy performance, with no quantity-accuracy trade-off. Conversely, the witness might remember a large amount of information but be relatively unable to determine which facts are correct and which are false. In that case, employing a liberal response criterion will produce a lot of correct information, but a great deal of false information as well (high quantity but low accuracy performance) whereas employing a more strict criterion might yield more

accurate testimony (or it might not; see the following), but at the expense of withholding a substantial amount of correct information.

Our model, then, implies that overt memory performance should depend on a complex interplay between underlying memory and metamemory processes. What is the evidence for ibis model? Do the monitoring and control processes in fact operate in the postulated manner? We examined the assumptions and implications of the model in two experiments using a special procedure that combines both free and forced reporting. In the first experiment (Koriat &. Goldsmith, 1996b, Experiment 1), we gave participants a general-knowledge test in either a recall or a recognition format. 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 10 times greater than the bonus).

This design enabled us to trace the links postulated by the model (see Fig. 5.1) between retrieval, monitoring, control, and memory performance (accuracy and quantity). The results accorded well with the model: First, the participants exhibited a good ability to monitor the correctness of their answers, as indicated by moderately high within-subject correlations between confidence and the correctness of the answer on the forced-report phase (.87 for recall and .68 for recognition). Second, there was a very high correlation between subjective confidence and whether or not an answer would be volunteered in the free-report phase (the gamma correlations 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. 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 (a 46% and 63% improvement in the moderate- and high-incentive conditions, respectively) . However, because the participants' monitoring was less than perfect, the increased accuracy was achieved at the cost of withholding some correct answers as well. Thus, a quantity—accuracy trade-off was observed both in comparing free- and forced-report performance (a 23% quantity decrease in the moderate-incentive condition), and in comparing performance under the two incentive conditions (a further 12% quantity decrease for the high-incentive compared to the moderate-incentive condition, after adjusting for different levels of forced-report performance).

The experiment just described examined the operation and effects of monitoring and control processes under fairly typical conditions. What should happen, however, when monitoring effectiveness is quite poor? In a second experiment (Koriat &. Goldsmith, 1996h, Experiment 2), we manipulated monitoring effectiveness 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 generally not correlated with the correctness of their answers (see also Fischhoff, Slovic, &. Lichtenstein, 1977; Gigerenzer, Hoffrage, & Kleinbolting, 1991; Koriat, 1995), whereas the other set (the "good" monitoring condition) consisted of more typical items for which the participants' monitoring was more effective. The results indicated that in both monitoring conditions the participants based their free-report control decisions on their monitoring output, but the consequences for memory performance were dramatically different in the two cases: Whereas in the good monitoring condition, the participants were able to increase their accuracy substantially when given the option of free report (75% under free report compared to 22% under forced report), in the poor monitoring condition the participant; were able to attain only a very low level of free-report accuracy (21% under tree report compared to 8% under forced report). Despite this difference in the free-report accuracy improvement, the quantity cost of the improved accuracy was about the same for both monitoring conditions (about a 5 percentage-point drop when comparing free and forced report). Thus, as expected, a much more severe quantity-accuracy trade-off was observed in the poor monitoring condition.

Monitoring effectiveness, then, emerges as a critical determinant of memory performance in the many situations in which people have the option of providing or withholding information. When rememberers' confidence judgments are reasonably diagnostic of the correctness of their answers, the option of free report can allow them to achieve high levels of accuracy. At the extreme, when monitoring is perfect, **completely** accurate performance can be achieved with no quantity cost at all. Consider again a student who wants to avoid the embarrassment of providing incorrect answers on a test, preferring to omit the answer altogether when he or she feels that he or she does not know the answer. The extent to which this inclination will impair the student's quantity score (or indeed, improve accuracy performance) will depend on his or her monitoring ability. If the

student is able to discriminate effectively between what he or she knows and what he or she does not know, the student might not lose any points at all by employing such a strategy.

In other situations, however, people's monitoring may sometimes be undiagnostic to the point of being useless. They will still control their memory reporting according to their monitoring output (for lack of any better basis), 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. Documented cases of poor monitoring are more common than one might think. Cohen (1988), for example, 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. Koriat (1995), using deceptive items such as those used here (see also Fischhoff et al., 1977), found that feeling-of-knowing (FOK) judgments after unsuccessful recall were either not correlated or even negatively correlated with subsequent recognition memory performance. Weingardt, Leonesio, and Loftus (1994) found exposure to postevent misinformation to impair the relation 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, Levin, Ghatala, & Ahmad, 1987), Korsakoff patients (e.g., Shimamura & Squire, 1986; 1988) and patients with frontal lobe lesions (e.g., Janowsky, Shimamura, &. Squire, 1989). Clearly, in all of these cases, the impaired monitoring is likely to have devastating consequences for free-report memory performance.

Of course, even when monitoring abilities are not actually impaired, differences in monitoring effectiveness may still contribute substantially to the variance in observed memory performance. Importantly, this contribution is independent of what might be called memory retention. In the experiment just described, for instance, we performed an additional comparison in which the good monitoring and poor monitoring items were matched on retention, so that forced-report quantity performance was equivalent. The basic pattern of results remained unchanged: The participants were able to attain a far superior joint level of free-report accuracy and quantity performance in the good monitoring condition than in the poor monitoring condition.

Findings 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 FOK 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 & Ritter, 1992; Schwartz & 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, whereas in fact their actual performance \vas impaired.

Such dissociations highlight a basic difference between our proposed framework for conceptualizing the strategic regulation of memory reporting and the signal-detection approach to memory. Because the application of the signal-detection methodology is essentially limited to forced-report recognition memory (Lockhart 6k Murdock, 1970), the signal-detection framework does not address the separate contributions of memory retention (or memory strength) and monitoring effectiveness to memory performance. In that framework, subjective confidence and memory strength are generally treated as synonymous (Chandler, 1994), and in fact, confidence is often used to index memory strength (see, e.g., Lockhart & Murdock, 1970; Parks, 1966). By contrast, in our proposed framework for conceptualizing free-report performance, monitoring and retention (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 any combination of these three factors.

The conceptual separation of these components of free-report performance has important implications. At the theoretical level, it calls for more serious efforts to incorporate monitoring and control processes—as well as encoding, storage, and retrieval processes—into our theories and models o: memory. At the same time, however, an acknowledgment of the potential effects of metamemory processes on memory performance poses a troubling policy issue: How should such effects be handled when assessing memory performance?

INCORPORATING MONITORING AND CONTROL PROCESSES INTO THE ASSESSMENT OF MEMORY

How have the effects of monitoring and control processes during memory reporting typically been treated in the evaluation of memory? In general, experimental psychologists have shied away from tackling the implications

of subject-controlled 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 &Narens, 1994). Thus, one approach has been to take control away from the participant, for instance by using forced-report testing techniques (Erdelyi & Becker, 1974). Another alternative is to allow participants some degree of control, but then to attempt to "correct" for it by applying such techniques as those provided by the signal-detection methodology (Banks, 1970) or formula scoring (Budescu & Bar-Hillel, 199.3). A third approach has been simply to ignore subject control altogether, assuming that it does not have much effect on performance anyway (see Roediger, Srinivas, & Waddil, 1989).

None of these approaches, however, seems completely satisfactory. First and foremost, they are all designed primarily to circumvent the contribution of subject-controlled processes to memory performance, treating this contribution as a nuisance variable rather than an integral aspect of memory functioning that should be assessed and studied. Such a strategy misses the point that metamemorial monitoring and control processes constitute a principal means by which people regulate their memory performance, and it is important to gain a better understanding of that regulation. Furthermore, it is questionable to what extent the aforementioned methods do in fact manage to yield a "pure" measure of memory performance that is untainted by the effects of metamemory processes.

Consider formula scoring, for example. This technique is usually applied in order to achieve an estimate of the test-taker's actual knowledge, cleansed from the contribution of guessing. Hence, on a 5-item multiple-choice test, for instance (in which there is a 20% baseline chance of guessing the correct answer), the test-taker might be awarded 1 point for each correct answer, but penalized one quarter point for each incorrect answer (commission error), with omissions simply ignored . Using such a procedure, have the potential contributions of monitoring and control processes been effectively neutralized? Seemingly not. First, there are possible differences in the interpretation of the instructions that can lead to the adoption of different control policies. This may be particularly true when the test-taker is not informed of the exact scoring formula, but instead is given vague guidelines that encourage him or her to guess on the basis of partial knowledge but to

This example assumes what is perhaps the most common formula scoring rule: $S = R - W/(k - 1)$, where S is the formula score, R is the number of right answers, W is the number of wrong answers (commission errors), and k is the number of response options. The basic property of this rule is that one's expected score is the -same whether one guesses the answer to an item at random or whether one omits it (see Budescu & Bar-Hillel, 1093).

avoid guessing wildly (Abu-Sayf, 1979; **Budescu &. Bar-Hillel,** 1993). In this case, the control policy that is adopted will depend to a large extent on what die person considers to be "enough" partial knowledge, and perhaps also on a variety of personality or other factors that may influence the tendency to guess, such as gender or culture (Gafni, 1990), or risk preferences (Budescu & Bar-Hillel, 1993). **Furthermore,** even when test-takers are informed of the precise scoring formula, this does not necessarily enable them to adopt the optimal control strategy for that formula (see Abu-Sayf, 1979; Koriat & Goldsmith, 1996b), nor does it preclude the possibility that extraneous motivations might also affect people's control decisions (as **when** people omit answers under forced-report instructions even though there is no objective advantage in doing so; cf. earlier example, and see Grandy, 1987). Yet both empirical studies and simulation analyses indicate that different control policies can yield substantially different levels of **performance** on such tests (e.g., Albanese, 1988; Angoff & Schrader, 1984; Cross & Frary, 1977; Frary, 1980; Slakter, 1968).

Second, even when the control policy is held constant, differences in monitoring effectiveness can also have a substantial effect on the test-taker's formula score. Metacognitive research has distinguished two distinct aspects of monitoring effectiveness, resolution and calibration (see, e.g., Koriat & Goldsmith, 1996b; Lichtenstein, Fischhoff, & Phillips, 1982; Nelson, 1984, 1996; Yaniv, Yates, &Smith, 1991). Resolution, or discrimination accuracy, is the aspect that we have considered so far, the extent to which the person is able to distinguish between answers that are more likely or less likely to be correct. Calibration, on the other hand, refers to the absolute correspondence between a person's confidence in his or her answers and the actual likelihood that they are correct. This measure relates to over or underconfidence: A person would be overconfident, say, if most of his or her subjective probability assessments are exaggerated. Clearly resolution is important: As we saw before, it is this aspect of monitoring that enables the person to choose the right answers to volunteer and to withhold. In addition, calibration may also affect the person's performance. To illustrate, assume that a test-taker decides that it is worth volunteering any answer that has a better than 25% chance of being correct, but to withhold it otherwise. This person therefore sets his or her subjective response-criterion probability at the .25 level. If this person's probability assessments are miscalibrated, however, her control policy may actually be more liberal (if overconfident) or more conservative (if underconfident) than he or she intended, and this may affect his or her ultimate score on the test.

Of course, with different testing or scoring methods, some of these contributions of metamemory processes to test performance could perhaps be neutralized. The primary issue, however, is not merely methodological. The main question that we must ask ourselves in any assessment situation is precisely what aspect or aspects of the person's performance or ability are we interested in evaluating. Rarely is the quantity of information that a person can reproduce of interest in itself. In evaluating eyewitness testimony, for instance, the quantity of information is surely important, but the accuracy of the testimony may be even more crucial. Can we depend on most or all of what the witness says to be true? Of course, if the testimony is inaccurate, we also want to know why: Is the inaccurate reporting due to a deficiency in monitoring, or in control? Although perhaps less obvious, similar questions need to be asked in the context of scholastic and psychometric testing. Is the ability to monitor one's own knowledge, for instance, to be included among those aspects of the test-taker's aptitude or achievement that the test is intended to tap? Would we want to certify (or hire the services of) a doctor, lawyer, psychologist, or engineer who was deficient in discriminating between what he or she knows and does not know? Finally, in the context of cognitive neuropsychological testing, what are the critical aspects of impaired memory functioning associated with certain forms of brain damage? To what extent does the impaired performance stem from deficient retention, deficient monitoring or deficient control (see Schacter, chapter 6, this volume) ?

In order to address such questions, it would be helpful to have available measurement techniques that incorporate metamemory processes into the assessment of memory performance, still allowing a separate evaluation of their independent contributions. One such method that we proposed (Koriat &. Goldsmith, 1996h) involves the derivation of *quantity-accuracy profiles* (QAPs). Rather than attempt to provide a single point-estimate of memory performance, QAPs provide information regarding the potential memory quantity and memory accuracy performance that can be achieved by the person under given conditions. To illustrate the method, Fig. 5.2 presents the QAPs for two participants who took the general-knowledge (recall) test in the study we described earlier (Koriat & Goldsmith, 1996b, Experiment 1). For each participant, confidence data from the initial forced-report phase were used to compute the input-bound quantity scores and the output-bound accuracy scores (plotted on the y-axis) that would result from the application of 11 different response-criterion levels (plotted on the x-axis), ranging from 0 (forced report) to 1.0. In addition, the

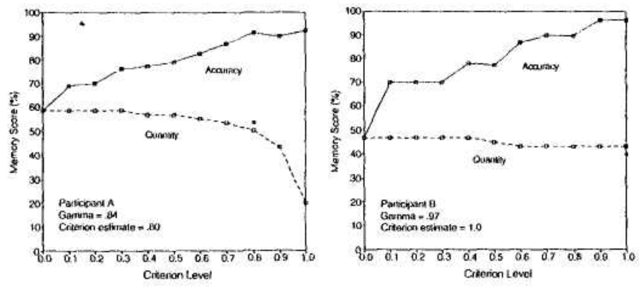

FIG. 5.2. Two illustrative quantity-accuracy profiles (QAPs). (Adapted from "Monitoring and Control Processes in the Strategic Regulation of Memory Accuracy, by A. Koriat and M, Goldsmith, 1996, *Psychological Review.* 103, (pp. 490-5 17. Copyright © 1996 by the American Psychological Association. Adapted by permission.

correlation between confidence and actual correctness of the answers on the forced-report phase (Kruskal-Goodman's gamma) was computed as a measure of monitoring effectiveness.

Finally, by examining the relation between confidence on the forced-report phase and volunteering or withholding answers on the free-report phase, a "best fit" estimate of the control policy (response criterion) adopted by the participant on the free-report phase was derived (for details, see Koriat & Goldsmith, 1996b). The actual quantity and accuracy scores achieved by the participant in the free-report phase are plotted as bullets above the estimated criterion level.

What type of information can be gleaned from these QAPs? .f we were to look only at forced-report performance as a measure of retention (or knowledge), then Participant A's performance would be superior to Participant B's. The profiles, however, offer a much more complete picture than this. First, looking at the participant's actual free-report performance, we see that although A's quantity performance is still superior to B's, B's performance is more accurate than A's. Is B's superior accuracy due simply to the use of a stricter control policy? (Both participants operated under a 10:1 penalty-to-bonus payoff scheme.) The estimated response criterion for Participant B (1.0) is indeed higher than for Participant A (.80). However, even if A had adopted the same criterion as B, B's accuracy would still be higher. Even more noticeable, however, is the substantial price in quantity performance that A would pay by raising her criterion any further. Thus, pressing A to be "absolutely sure" about what she knows before venturing an answer would markedly impair her performance. These differences

between A and B in potential accuracy performance and in the degree of trade-off between quantity and accuracy performance are reflected in the different levels of monitoring effectiveness they exhibit (gamma correlations of .84 and .97, respectively). So, who has better general knowledge, Participant A or Participant B? Although A appears to have more knowledge, B seems to be better able to discriminate between what she knows and what she does not know, yielding a higher level of potential accuracy. Also, B seems to be more cautious than A in her responding. Clearly, then, our ultimate appraisal will need to take into account the relative importance of these different aspects for the task at hand. For example, as a contributor to a brainstorming session we might prefer Participant A, with her greater amount of potential information (and her greater tendency to volunteer it), but as a key witness in a capital trial we would probably prefer Participant B, because of the high priority given to accurate testimony in that situation. An important advantage of this assessment procedure, then, is not only its ability to provide separate indices for the various aspects of memory performance (i.e., retention, monitoring, and control), but also that it forces one to make an explicit, and thoughtful decision about the weight to be attached to each of these aspects in the overall performance **evaluation.**

Of course the QAP procedure has its own limitations. First, various aspects of the procedure, such as giving people the same test twice under both forced-report and free-report instruction or the **elicitation** of confidence judgments, may not always be feasible. Second, in many practical situations (e.g., university admissions), one is interested in a single index on which to compare people's performance, and the QAP procedure complicates the derivation of such an index. This complication, however, is the price of taking the contributions of subject-controlled metamemory processes seriously, rather than simply ignoring them.

CONCLUDING REMARKS

In this chapter, we focused on some of the metamemory processes that operate during memory reporting, and showed how these processes can have substantial effects on memory performance in a variety of situations. Although a great deal of work has been directed toward an understanding of metacognitive processes and their determinants in the last decade, clearly more needs to be done to uncover and address the performance consequences of these processes. The work we have reviewed here (see also Goldsmith &. Koriat, in press) reveals some of the complexities that arise when metamemory processes are allowed to operate during memory reporting, and poses the question of how such complexities should be handled. Memory researchers, neuropsychologists, and test administrators alike will need to grapple with the issue of whether, when, and how metamemory processes should be taken into account in the assessment of people's performance.

REFERENCES

Abu-Sayf, F. K. (1979). The SCORING of multiple-choice tests: A closer look. Educational Technology, 19, 5-15.

Albanese, M. A. (1988). The projected impact of the correction for guessing on individual scores. Journal of Educational Measurement, *25*, 149-157.

Angoff, W. H. (1989). Does guessing really help? Journal *of Educational* Measurement, 26, 323-336.

Angoff, W. H., & Schroder, B. W. (1984). A study of hypotheses basic to the use of rights and formula .1-17 ,21 ,urementsMea of Educational Journal .scores

Banks, W. P (1970). Signal detection theory and human memory. Psychological Bulletin, 7-, 81-99.

Barnes, A. E., Nelson, T O., Dunlosky, J., Mazzoni, G., & Narens, L (in press). An integrative system of metamemory components involved in retrieval, in D. Gopher & A. Koriat (Eds.), Cognitive regulation of performance: Interaction of theory and application: Attention and performance XVII. Cambridge, MA: MIT Press.

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

Budescu, D. & 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, hut increase confidence, in a modified recognition text. Memory and Cognition, 22, 273-280.

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

Cronbach, L. J. (1984). Essentials of psychological testing. New York: Harper & Row.

Cross, L. H., & Frary, R. B. (1977). An empirical test of Lord's theoretical results regarding formula scoring of multiple-choice tests. Journal of Educational Measurement, 14, 313-321.

Deffenbacher, K. A. (1991). A maturing of research on the behavior of eyewitnesses. Applied Cognitive Psychology, 5, 377-402.

Ebbinghaus, H. E. (1964). Memory: A contribution to experimental psychology. New York: Daver.(Original work published 1895)

Erdelyi, M. H., & Becker J. (1974). Hypermnesia for pictures: Incremental memory for pictures hut not words in multiple recall trials. Cognitive psychology, 6, 159-171.

Erdelyi, M. H., Finks, J., & 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, & 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, & Geiselman, K. 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 (Pt. 4, pp. 3-24). New York: McGraw-Hill.

Frary, R. B. (1980). The effect of misinformation, partial information, and guessing on expected multiple-choice test item scores. Applied Psychological Measurement, 4, 79-90.

Gafni, N. (1990). Differential tendencies to guess as a function of gender and lingual-cultural reference group (Rep. No. 115). Jerusalem, Israel: National Institute for Testing and Evaluation.

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

Goldsmith, M., &. Koriat, A, (in press). The strategic regulation of memory reporting: Mechanisms and performance consequences. In D. Gopher & A. Koriat (Eds,), *Cogitative regulation of performance; Interaction of theory and* application: *Attention and performance* XVII. Cambridge MA: MIT Press.

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

Grand y, J. (1987). Characteristics of examinees who leave questions unanswered on the GRE general test under right-only scoring (Rep. No. 37—38). Princeton, NJ: Educational Testing Service.

Green D. M., 61 Swets, J. A. (1966). Signal detection theory and psychophysics. New York: Wiley.

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

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

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

Klatzky, R. L., & Erdelyi, M. H. (1985). The response criterion problem in teals 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 (or the accessibility model. Journal of Experimental Psychology: General, 124, 311-333.

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

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

Koriat, A., &. Goldsmith, M. (1996b). Monitoring and control processes in the strategic regulation (if memory accuracy. Psychological Review, 103, 490-517.

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

Lockhart, R. S., & Murdock, IV 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., & Hoffman, H. G. (1989). Misinformation and memory; The creation of new memories. Journal *of* Experimental *Psychology: General*, M8, 100-104.

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

Nelson, T O- (1984)- A comparison of current measures of the accuracy of feeling-of-knowing predictions. Psychological Bulletin, 95, 109-133.

Nelson, T O. (1996). Gamma is a measure of the accuracy of predicting performance on one item relative to another item, not of the absolute performance on an Individual item: Comments on Schraw (1995). Applied *Cognitive Psychology*, *10*, 257-260.

Nelson, T O., & Narens, L. (1990). Metamemory: A theoretical framework and now findings. In G. Rower (Ed.), *The psychology of learning and motivation* (pp. 125—173). New York: Academic Press.

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

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

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

Pressley, M., levin, J. R., Ghatala, E. S., & Ahmad, M. (1987). Test monitoring in young grade school .96-111 ,43 ,*Journal of Experimented Child Psychology* .children

Puff, C. R. (Bd.) (1982). Handbook of research methods in human memory and cognition. :New York Academic Press.

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

Roediger, H. L., & Payne, D. G. (1985). Recall criterion docs mil affect recall level or hypermnesia: A puzzle for generate/recognize theories**.** Memory and Cognition**,** 13, 1-7.
Roediger, H. L., Srinivas, K., & Waddil, P. (1989). How much does guessing influence recall? Comment on Erdelyi, Finks, and Feigin-Pfau. Journal Experimental Psychology: General, 118, 253-257.
Schwartz, 15. L, & Metcalfe, J. (1992). Cue familiarity hut not target retrievability enhances feeling-of-knowing judgments. Journal of Experimental Psychology: learning. Memory, and Cognition, 18, 1074-1083.
Shimamura, A. P, & Squire, L. R. (1986). Memory and metamemory: A study of the feeling-of-knowing phenomenon in amnesic patients. Journal of Experimented Psychology : learning. Memory, and Cognition, 12.452-460.
Shimamura, A. P, & Squire, L.R. (1988). Long-term memory in amnesia: Cued recall, recognition memory and confidence ratings. Journal of Experimented Psychology : learning. Memory, and Cognition, 14, 763-770.
Slakter, M. (1968). The penalty for not guessing;. journal of Educational Measurement, 5, 141-144.
Thurstone, L L. (1919). A method fur scoring tests. Psychological Bulletin, 16. 235-240.
Weingardt, K. R., Leonesio, R. J., & Loflus, E. F. (1994). Viewing eyewitness research from a metacognitive perspective. In J. Metcalfe & A. P Shimamura (lids.), Metacognition: Knowing about knowing (pp. 157-184). Cambridge, MA; MIT Press.
Yaniv, I., Votes, J. F,& Smith, J.]•. K. (1991). Measures of discrimination skill in probabilistic judgment. Psychological Bulletin. 110. 611-617.