Journal of Experimental Psychology: General
2002, Vol. 131, No. 1, 73–95

0096-3445/02/$5.00 DOI: 10.1037//0096-3445.131.1.73

Strategic Regulation of Grain Size in Memory Reporting

Morris Goldsmith, Asher Koriat, and Amit Weinberg-Eliezer
University of Haifa

To increase their report accuracy, rememberers may either withhold information that they feel unsure about or provide relatively coarse information that is unlikely to be wrong. In previous work (A. Koriat & M. Goldsmith, 1996c), the authors delineated the metacognitive monitoring and control processes underlying the decision to volunteer or withhold particular items of information (report option) and examined how these processes are used in the strategic regulation of memory accuracy. This article adapts that framework to address control over the *grain size* (precision–coarseness) of the information that people report. Results show that rememberers strategically regulate the grain of their answers to accommodate the competing goals of accuracy and informativeness. The metacognitive processes underlying this regulation are elucidated.

Q: Please tell me 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 sweat suit burst through the gate of the yard and ran like crazy down the alley. He seemed to be carrying a bag or something over his shoulder.
Q: What color was the bag?
A: I'm not sure. I think it was a reddish color—red or maybe orange—with some sort of striped pattern.
Q: Do you remember what time it was?
A: Around 5 o'clock, maybe 5:30.
Q: Could you be more specific?
A: Umm [thinks for a moment]. Between 5:15 and 5:30.
Q: Did you see anything else?
A: No, nothing that seems important.

Although fictional, this short transcript illustrates some of the vast flexibility that people generally have in recounting past events from memory. Unlike in most traditional laboratory experiments, there is no official "list" of input items that must be reproduced. 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. Although in some situations one may find oneself pressed to report a particular piece of information or to give a more specific answer, even then it is the rememberer who ultimately decides whether to acquiesce to such pressure.

Our interest in personal control over memory reporting was initially motivated by an attempt to understand the heated debate between proponents of the traditional, laboratory-based study of memory and those who favor the ecological study of memory in naturalistic settings (see, e.g., the January 1991 issue of *American Psychologist* [Fowler, 1991]). Our analysis of this controversy (Koriat & Goldsmith, 1994, 1996a, 1996b) disclosed two fundamental aspects that had previously been overlooked: (a) the memory property of interest and (b) the role of personal control over memory reporting. As far as memory property is concerned, the recent wave of naturalistic research discloses an unparalleled preoccupation with the accuracy of memory, that is, the extent to which memory reports can be trusted. This is in contrast to the traditional approach, which has focused almost exclusively on memory quantity, that is, on the amount of information that is retained and can be recovered. We have argued that these different foci reflect a divergence between two different guiding metaphors of memory: a *storehouse* conception that treats memory as something that can be "counted" and a *correspondence* conception that treats memory as something that can be "counted on." This latter conception is clearly seen in much of the recent work on eyewitness testimony, autobiographical memory, false memories, reconstructive memory, and so forth (for a recent review, see Koriat, Goldsmith, & Pansky, 2000).

With regard to personal control over memory reporting, here too the treatment has been quite different in laboratory and naturalistic research. Personal control has not figured prominently in traditional quantity-oriented memory research, perhaps because of its seeming incompatibility with the desire for strict experimental control (Banaji & Crowder, 1989; Nelson & Narens, 1994). Thus, the common approach is to limit personal control as much as possible (e.g., by using forced-report techniques; Erdelyi & Becker, 1974) or to attempt to "correct" for it by using techniques such as those provided by the signal-detection methodology (Banks, 1970) or standard correction-for-guessing formulas (Budescu & Bar-Hillel, 1993). Another approach is simply to ignore

Morris Goldsmith, Asher Koriat, and Amit Weinberg-Eliezer, Department of Psychology, University of Haifa, Haifa, Israel.

This research was supported by a grant to Morris Goldsmith and Asher Koriat from the Israel Science Foundation, founded by the Israel Academy of Sciences and Humanities. Facilities for conducting the research were provided by the Institute of Information Processing and Decision Making, University of Haifa, and by the Max Wertheimer Minerva Center for Cognitive Processes and Human Performance. We are grateful to Ainat Pansky for her help and valuable comments. We also thank Gilat Linn and Meni Yeari for their help in running the experiments.

Correspondence concerning this article should be addressed to Morris Goldsmith, Department of Psychology, University of Haifa, Haifa 31905, Israel. E-mail: mgold@research.haifa.ac.il

personal control, assuming that it does not have much effect on performance anyway (see Roediger, Srinivas, & Waddil, 1989).

In contrast, in everyday memory research (as in most real-life memory situations), not only are people generally given much more control over their memory reporting, but such control has been shown to have a substantial impact on memory performance. For instance, it is established wisdom in eyewitness research (e.g., Hilgard & Loftus, 1979; Neisser, 1988) that open-ended, free-narrative types of questioning are superior to more constrained or directed types of questioning in eliciting accurate testimony from witnesses (this wisdom has in fact been incorporated into standard interviewing techniques; see Fisher & Geiselman, 1992).

Open-ended memory questioning offers rememberers at least two means by which they can enhance the accuracy of what they report. The first, *report option,* involves choosing either to volunteer or to withhold particular items of information (i.e., to respond “don’t know” or “don’t remember”). The second, *control over grain size,* involves choosing the level of detail (precision) or generality (coarseness) at which to report remembered information. Both of these are intrinsic aspects of real-life remembering. Thus, as we have argued previously (Goldsmith & Koriat, 1999; Koriat & Goldsmith, 1996b), rather than constituting a mere methodological nuisance that should be eliminated or corrected for, both report option and control over grain size constitute important topics of study in their own right, with underlying dynamics and performance consequences that deserve systematic investigation. In what follows, we first describe a theoretical framework that was developed for the study of report option and then show how this framework can be extended to address control over memory grain size.

A Framework for Control of Report Option

Our work on personal control so far has focused on report option, that is, on how one uses the option to volunteer or withhold answers in the service of enhancing one’s accuracy. As a framework for addressing this question, we proposed a simple model (Koriat & Goldsmith, 1996c) that essentially merges the logic of signal-detection theory with concepts and tools from the study of metamemory. We assume that when reporting information from memory, people invoke monitoring and control processes to screen out information that is likely to be wrong. Thus, in addition to a memory-retrieval mechanism, the model includes a *monitoring* process that assesses the correctness of potential memory responses and a *control* mechanism that determines whether or not to volunteer the best available candidate answer (for similar proposals, see Barnes, Nelson, Dunlosky, Mazzoni, & Narens, 1999; Klatzky & Erdelyi, 1985). The control mechanism operates by setting a confidence threshold (response criterion) on the monitoring output: The answer is volunteered if its assessed probability of being correct passes the threshold but is withheld otherwise. The threshold is set on the basis of competing incentives for both quantity and accuracy, that is, the gain for providing correct information relative to the cost of providing incorrect information.

To test the model and flesh out its implications for both memory accuracy and memory quantity performance, we conducted several studies in which report option was manipulated either between or within participants (Koriat & Goldsmith, 1994, 1996c). Under forced-report conditions, participants were presented with a set of memory questions and were forced to answer each of them, even if they had to guess. Under free-report conditions, in contrast, they were allowed to choose which items to answer and which to withhold. An explicit payoff schedule was used to motivate accurate responding (e.g., a bonus of 25 cents for each correct answer volunteered but a penalty of 25 cents for each incorrect answer volunteered).

Memory performance was scored for both quantity and accuracy. Memory quantity performance was defined as the *input-bound* percentage of questions that were correctly answered (i.e., conditional on the number of input items), whereas memory accuracy performance was defined as the *output-bound* percentage of answers that were correct (i.e., conditional on the number of output answers). Note that whereas the input-bound quantity measure taps the likelihood that an input question can be successfully answered, the output-bound accuracy measure taps the likelihood that a reported answer is correct. Thus, the output-bound accuracy measure uniquely reflects the dependability of the reported information, that is, the extent to which each reported item can be counted on to be correct.

The results provided good support for the model, revealing the role of report option in the strategic regulation of memory accuracy. First, when given the option of free report, participants enhanced their memory accuracy performance substantially relative to forced report. Second, however, this improvement came at the cost of a reduction in memory quantity performance (i.e., a quantity–accuracy trade-off). Third, participants were sensitive to the level of accuracy incentive, enhancing their accuracy even further when a stronger incentive (e.g., a 10:1 penalty-to-bonus ratio) was provided. This improvement came at a further cost in quantity performance.

In addition, elicitation of confidence judgments (assessed probability correct) during the forced-report phase (when report option was manipulated within participants; Koriat & Goldsmith, 1996c) allowed us to uncover the mechanisms mediating the effects of report option and accuracy incentive on memory accuracy and quantity performance. First, participants were relatively successful in monitoring the correctness of the candidate answers and appeared to base the decision to volunteer or withhold each answer almost entirely on their confidence. Second, the participants’ control policies were sensitive to the specific level of accuracy incentive: Participants who were given a high accuracy incentive were more selective in their reporting, adopting a stricter criterion for volunteering an answer than those given a more moderate incentive. Finally, the extent of the quantity–accuracy trade-off depended on monitoring effectiveness: When monitoring effectiveness was good, a higher level of accuracy could be achieved at a lower cost in quantity than when monitoring effectiveness was poor.

In sum, our work on report option indicates that people use the option of free report to regulate the accuracy of their memory reporting, sacrificing quantity for accuracy in a strategic manner. They do so by monitoring the correctness of the information that comes to mind and controlling their responding accordingly. Thus, a full understanding of memory performance in real-life situations requires specification of the metacognitive processes underlying the use of report option and how the operation of these processes is affected, for the better or for the worse, by various factors.

Control Over Grain Size

Let us now turn to control over memory grain size. As discussed earlier, report option is only one of the means by which people can regulate their memory performance in real-life settings. In addition, people can also regulate the level of detail or generality at which information is reported (e.g., reporting an assailant's height as "5 feet 11 inches," "around 6 feet," or "fairly tall"). Thus, for example, Neisser (1988), in testing students' memory for events related to a course he taught, noted that in response to open-ended questions, the participants tended to provide answers at "a level of generality at which they were not mistaken" (p. 553). Also, Fisher (1996), in assessing participants' freely reported recollections of a filmed robbery, was surprised to find that the accuracy of people's reports after 40 days was no lower than on immediate testing, even though the same number of propositions were volunteered. The anomaly was resolved by considering the grain size of the reported information: Propositions volunteered after 40 days were as likely to be correct as those provided soon after the event (about 90% accuracy in both cases), but this equivalence was achieved by rememberers providing information that was more coarse (as rated by two independent judges) at the later testing than in the earlier reports.

Clearly, such person-controlled differences in the grain size of the reported information can pose a troubling methodological problem. Here too, as with report option, the traditional remedy has been to take control away from the participant by using recognition testing or by using stimulus materials, such as word lists or questions requiring single-word answers, that greatly limit the scope of the problem. This, in fact, is what we did in our studies focusing on the effects of report option, and, somewhat ironically, this approach also characterizes most of the traditional and current work on memory for gist versus verbatim detail (see General Discussion). Control over grain size, however, as with report option, is more than just a mere methodological nuisance that needs to be circumvented or corrected for. The challenge is to find a way to systematically investigate this type of control as well. In this article, we take an approach that is similar to the one we used for report option and, in fact, assumes a close relationship between these two types of control.

Consider a situation in which a witness is asked to answer a set of questions that have to do with quantitative values, such as the time of an accident, the speed of a car, the height of an assailant, and so forth.[1] If the witness is forced to answer each question at a specified grain size (e.g., to the nearest minute, mile per hour, or inch), then the accuracy of those answers may be quite poor. However, even though the witness may not remember, say, that the accident occurred precisely at 5:17, she may be able to report that it occurred between 5:00 and 5:30, or perhaps, in the late afternoon. 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, increasing the (output-bound) accuracy of her memory report? On what basis will she choose an appropriate grain size for her answers?

We propose that the considerations and mechanisms underlying the choice of grain size in memory reporting are similar to, although perhaps more complex than, those underlying the exercise of report option. Let us assume that the witness would like to fulfill her vow to "tell the whole truth and nothing but the truth." How should she proceed? On the one hand, a very coarsely grained response (e.g., "between noon and midnight") will always be the wiser choice if accuracy (i.e., the probability of including the true value: telling nothing but the truth) is the sole consideration. However, such a response may not be very informative, falling short of the goal to tell the whole truth. On the other hand, whereas a very fine-grained answer (e.g., 5:23 p.m.) would be much more informative, it is also much more likely to be wrong. Thus, control over grain size would seem to involve an accuracy–informativeness trade-off similar to the accuracy–quantity trade-off observed with regard to the control of report option.

This idea of an accuracy–informativeness trade-off was brought out nicely by Yaniv and Foster (1995, 1997) in the context of judgment and decision making. Focusing on graininess in judgmental estimation under uncertainty, they proposed that the optimal grain size for estimates of uncertain quantities

> 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 & Foster, 1995, pp. 424–425)

Yaniv and Foster (1997) suggested that when people are asked to provide quantitative estimates for the purpose of decision making, they tend to consider the recipient's desire to obtain a useful response (cf. Grice, 1975) and will often sacrifice accuracy for informativeness.

Similarly, returning to our hypothetical witness, unless she has a phenomenally precise memory of the entire event, she too will presumably have to aim for a compromise between accuracy and informativeness in choosing a grain size for her answers. Perhaps the simplest strategy for doing so would be to provide as finely grained (i.e., precise) an answer as possible, as long as its assessed probability of being correct passes some preset criterion. Thus, she might try to answer the question to the nearest minute; to the nearest 5 min, 10 min, or 15 min; and so forth until she is, say, at least 90% sure that the specified answer is correct. This model is similar to the one underlying our work on report option: The rememberer attempts to provide as much information as possible, as long as its assessed probability of being correct is high enough. As with report option, the criterion level of confidence may be raised or lowered, depending on the relative incentives for accuracy and informativeness in the particular situation. Of course, the process might be more complex. For instance, in answering each question, the witness could monitor the probability that candidate answers at various grain sizes are correct, judge the informativeness (or utility) of the answer at each grain size, and then choose the answer with the highest subjective expected utility.

Although these two models (and perhaps others) differ in their specifics, they share a common conception of the choice of grain

[1] It is methodologically convenient to operationalize grain size in terms of the range or interval width used in reporting quantitative information (see Yaniv & Foster, 1995, 1997). We provisionally assume that other forms of control over grain size (e.g., vague linguistic qualifiers such as "reddish" vs. "red") should operate according to similar principles, although ultimately this assumption will need to be tested (see the General Discussion section).

size as being based on two metacognitive processes: (a) a monitoring process that assesses the probability that answers at different grain sizes are correct and (b) a control process that uses the monitoring output, together with other information (e.g., the perceived informativeness of the answers or the relative incentives for accuracy and informativeness), to determine the most appropriate grain size for a particular answer.

In what follows, we report three experiments that examined the control of grain size within this general metacognitive framework. Our goals in this study were modest. We do not purport to specify and test a process model that captures the full complexity of the control of grain size in real-life memory reporting. Instead, we used a rather constrained experimental paradigm, similar to the one we used earlier for investigating the use of report option, in an attempt to answer several basic questions: Do people try to achieve a compromise between accuracy and informativeness when reporting information from memory? How effective are they in choosing an appropriate grain size for the information that they report? Are people able to monitor the correctness of their answers at different grain sizes? Do they use this monitoring to guide the choice of grain size and, if so, how? Is the relative informativeness of the answers also taken into account, and, if so, can the choice of grain size be affected by explicit incentives for informativeness versus accuracy? We hope that the initial answers to these questions obtained here will provide a valuable first step toward future work that brings additional cognitive, metacognitive, and psychosocial aspects of control over grain size into the laboratory for controlled experimental investigation.

Experiment 1

Experiment 1 adapted the two-phase paradigm used by Koriat and Goldsmith (1994). In the first phase of the experiment, participants were presented with a list of 40 general-knowledge questions pertaining to various types of quantitative information. The participants were required to give their best answer to each item using two different bounded intervals (grain sizes), the widths of which were specified by the experimenter. An example item is "When did Boris Becker last win the Wimbledon men's tennis finals? A) Provide a 3-year interval; B) Provide a 10-year interval." The two alternative interval widths were tailored for each item, such that one of the alternatives (the coarse-grained answer) specified a relatively wide interval, whereas the other alternative (the fine-grained answer) specified a more narrow interval (or, in some cases, a specific value, e.g., year). In the second phase, the participants were asked to go over their answers and, for each item, to indicate which of the two answers (i.e., which of the two grain sizes) they would prefer to provide, assuming that they were "an expert witness testifying before a government committee."

Will participants consider both accuracy (i.e., the probability that their answer is correct) and informativeness (i.e., the precision of the provided information) in choosing the grain size for their answer in the second phase of the experiment? A comparison of performance in Phases 1 and 2 sheds some initial light on this question and allows us to examine the performance consequences of the participants' grain choices.

Method

Participants

Forty Hebrew-speaking undergraduate students from the University of Haifa, 26 men and 14 women, participated in the experiment for payment (NIS 20, approximately $5).

Materials

A 40-item general-knowledge test (in Hebrew) was developed in which the answer to each question was a quantitative value on either an integer or continuous scale (e.g., "When did . . . ?" "How old was . . . ?" "How long is . . . ?" and "How many . . . ?"). Blanks were provided next to each question for providing an answer at two different grain sizes. For instance:

> When did the "Berlin Wall" that divided East and West Berlin fall?
> A) 2-year interval ______ – ______
> B) 10-year interval ______ – ______

For simplicity, the participants were instructed to treat the interval as specifying the arithmetic difference between the two endpoints of their answer (e.g., 1974–1984 would be considered a 10-year interval, although it is in fact an 11-year interval with respect to the scoring). When the fine-grained alternative required a specific answer (16 items), a single blank was provided.

The specified intervals for the fine-grained and coarse-grained alternatives differed for each item. These intervals were chosen on the basis of pretesting to yield about 30% correct performance for the fine-grained alternatives and about 70% correct performance for the coarse-grained alternatives. On average, the coarse-grained intervals were wider than the corresponding fine-grained intervals by a factor of five. Two item orders (one the reverse of the other) and the order of the grain alternatives for each item (fine grain first or coarse grain first) were counterbalanced across participants.

Procedure

The experiment was administered individually or in small groups and lasted about 45 min. Each participant was given a copy of the knowledge test and an accompanying instruction leaflet. Participants read the instructions and proceeded through the experiment at their own pace. The experimenter was present at all times for clarifications.

The experiment was divided into two phases. In Phase 1, the participants were required to provide the best answer they could for each item at both grain sizes, even if they had to guess. In Phase 2, they were instructed to go back over their answers from Phase 1 and choose one answer for each item (i.e., at one of the two alternative grain sizes). The instructions gave the following rationale to guide choices:

> Assume that you are an expert witness who has been called to testify before a government committee. As an expert, you are requested to provide an answer to each question that you answered before. Of the two answers that you wrote down before, which do you choose to provide to the committee?

The participants marked their choices by circling one of the two alternative answers for each item. They were not allowed to change any answers.

Results and Discussion

From 1,600 observations (40 items for each of the 40 participants), 11 were discarded as a result of minor procedural problems such as the participant deviating from the specified grain size, omitting an answer, or using illegible handwriting. The remaining observations were included in the analyses.

Did the participants strategically control the grain size of their answers, weighing accuracy against informativeness? If participants were concerned solely with maximizing the accuracy of their answers, they would be expected to cling exclusively to the coarse-grained answers, which were almost always more likely than the fine-grained answers to be correct. On the other hand, if they were concerned only with maximizing the informativeness of their answers, they would be expected to volunteer only fine-grained answers. Inspection of the data from Phase 2 (see Table 1) indicates that participants chose to volunteer the fine-grained answer 41% of the time (range: 10%–70%) and the coarse-grained answer 59% of the time. This suggests that neither informativeness nor accuracy was the participants' sole consideration; rather, they apparently tried to achieve a balance between these two competing goals.[2]

What were the performance consequences of the participants' choice of grain size? In assessments of accuracy, specific answers that matched the true value or interval answers that contained the true value were counted as correct. Accuracy scores, defined as the proportion of correct answers, were then calculated, for each participant, for the fine-grained answers in Phase 1, the coarse-grained answers in Phase 1, and the answers that were chosen to be volunteered in Phase 2 (at the chosen grain size). These scores averaged .32 and .75 for the fine-grained and coarse-grained answers, respectively (see Table 1), and .59 (SD = .12) for the answers volunteered in Phase 2. Thus, the participants were more accurate than they would have been by volunteering only fine-grained answers, $F(1, 39) = 347.55$, $p < .0001$, but less accurate than they would have been by volunteering only coarse-grained answers, $F(1, 39) = 143.57$, $p < .0001$.

Did the participants tend to choose a grain size that would enhance their accuracy with minimal loss of informativeness? To minimize the accuracy–informativeness trade-off, we would expect participants to provide the coarse-grained answer primarily when the fine-grained answer was unlikely to be correct. Table 1 presents mean proportions correct for the fine-grained and the coarse-grained answers in Phase 1, calculated separately for the items that were volunteered at the fine grain size in Phase 2 and those that were volunteered at the coarse grain size. The data show that items that participants chose to volunteer at the fine grain size were much more likely to be correct at the fine grain size (.50) than were items that participants chose to volunteer at the coarse grain size (.21), $F(1, 39) = 79.25$, $p < .0001$. Thus, it appears that participants did tend to provide coarse-grained answers primarily when the fine-grained answer was likely to be wrong, although their performance was far from optimal. Also, providing coarse-grained answers for these items led to the achieved accuracy increase (.67 − .21 = .46) being greater than the increase that would have been achieved by choosing to answer the other items at the coarse grain size (.88 − .50 = .38), $F(1, 39) = 8.14$, $p < .01$.[3]

In sum, the results of this experiment indicate that (a) participants attempt to find a compromise between accuracy and informativeness in choosing a grain size for their response and (b) their choice of grain size is not arbitrary; rather, they tend to provide the coarse-grained answer (sacrificing informativeness for accuracy) when the fine-grained answer is relatively unlikely to be correct and when the gain in accuracy from doing so is relatively large. Viewed from the perspective of our proposed framework, these results imply that participants are able to discriminate between answers that are more likely and less likely to be correct and base their choice of grain size, among other things, on this assessment. These implications were examined more fully in the following experiment.

Table 1
Experiment 1: Mean Proportions Correct for Fine-Grained and Coarse-Grained Answers in Phase 1, Calculated Separately for Items Volunteered at Each Grain Size in Phase 2, Along With Mean Percentages of Grain Choices and Overall Proportions Correct

Volunteered grain size (Phase 2)	Grain size of answer (Phase 1): Fine		Coarse		Percentage of choices (Phase 2)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Fine	.50	.20	.88	.12	41	16
Coarse	.21	.09	.67	.12	59	16
Overall	.32	.11	.75	.10		

Experiment 2

The results of Experiment 1 suggest that participants exercise control over grain size in a strategic manner, choosing to provide a more coarsely grained answer when the fine-grained answer is relatively unlikely to be correct. The aim of Experiment 2 was to examine the processes that mediate this choice.

As discussed earlier, we propose that the choice of grain size in memory reporting is based on metacognitive monitoring and control processes similar to those shown to underlie the use of report option (Koriat & Goldsmith, 1996c). However, the monitoring and control processes underlying the choice of grain size may be somewhat more complex. With regard to the monitoring process, rememberers may need to evaluate the probability that answers at various grain sizes, rather than a single specified grain size, are correct. With regard to the control process, again there may be a need to take into account the assessed probabilities of answers at different grain sizes and perhaps the differences in these probabilities. Moreover, the informativeness of the answers at different grain sizes may need to be assessed (a separate monitoring pro-

[2] One might argue that some of the variability in grain choices stems from "demand characteristics," that is, from participants' perception that they were expected to choose both fine- and coarse-grained answers. Although this may be true, it is doubtful that demand characteristics account for all of the variability, and certainly they cannot account for the systematicity that characterizes the pattern of choices in all three experiments.

[3] In response to a reviewer's comment, we note that although this interaction may be interpreted as reflecting a ceiling effect, it is precisely this ceiling on accuracy performance that might deter participants from providing the coarse-grained answer when the fine-grained answer also has a relatively high likelihood of being correct.

cess) and then taken into account in conjunction with the output of the accuracy-monitoring process.

The present experiment attempted to shed light on some of these aspects of the underlying monitoring and control processes. Participants completed the same general-knowledge memory test as in Experiment 1, again in a two-phase procedure in which they first answered each item at both of two specified grain sizes and then chose which grain size to provide to aid a "government committee." To obtain information about monitoring, we also asked the participants to assess the probability that their answer at each grain size was correct. This design allowed us to determine the extent to which the participants were effective in monitoring the correctness of their answers at the different grain sizes and to examine whether and how the monitoring judgments were taken into account in the choice of grain size.

As a first stab at characterizing the nature of the process underlying control of memory grain size, two basic types of models are compared. The first involves a *satisficing* strategy (Simon, 1956, 1990) in which the rememberer aims at providing the more informative answer as long as it has a sufficiently high assessed probability of being correct. In our experimental paradigm, this strategy can be implemented through a simple threshold model based on confidence in the fine-grained answer only: The fine-grained answer is volunteered if its assessed probability of being correct passes a preset criterion level; otherwise, the coarse-grained answer is volunteered.

The second type of model, a *relative-utility* model, assumes that the rememberer chooses a grain size that maximizes the expected subjective utility of each reported answer. For the purpose of evaluating this model, we make the broad assumption that the expected subjective utility of an answer is a monotonic function of the following product: Assessed Probability Correct × Perceived Informativeness. Although we cannot determine the (expected) subjective utility of the answers in this experiment (but see Experiment 3), the model yields the qualitative prediction that, all else equal, choice of grain size should depend on the relative disparity between the assessed probability correct of the fine-grained and coarse-grained alternatives: Participants should tend to provide the fine-grained answer when its assessed probability is relatively close to that of the coarse-grained answer but tend to provide the coarse-grained answer when its assessed probability is relatively high in comparison with that of the fine-grained answer. This is in contrast to the prediction derived from the satisficing model, that choice of grain size should depend on the assessed probability of the fine-grained answer only. Of course, the most basic question, common to both models, is whether there is any relationship at all between confidence in the correctness of the answers and the choice of grain size.

Method

Participants

Thirty-two Hebrew-speaking undergraduate students from the University of Haifa, 17 men and 15 women, participated in the experiment for payment (NIS 20, approximately $5).

Materials

Experiment 2 involved the same general-knowledge test as Experiment 1. For this experiment, however, an additional blank was added next to each answer at each grain size for recording the participant's confidence in the answer. (These blanks were positioned in a single column at the edge of the page so that they could be easily detached from the test form between phases; see the Procedure section.) Again, the two item orders and the order of the grain alternatives for each item were counterbalanced across participants.

Procedure

The procedure for this experiment was identical to that of Experiment 1, except for the elicitation of confidence judgments (assessed probability correct) in Phase 1: In addition to providing the answer to each question at each of the two grain sizes, the participants were also asked to "estimate the chances that the answer contains the true value" on a scale ranging from 0% to 100%. The confidence judgments were cut away from the test form by the experimenter before proceeding to Phase 2. The instructions for Phase 2 were the same as in Experiment 1.

Results and Discussion

As in the previous experiment, a small number of observations (24 of 1,280 items) were omitted from the analyses because of missing answers or confidence judgments, incorrect grain sizes, and so forth.

Overall Performance

To discount the possibility that the requirement to provide confidence judgments in this experiment substantially changed the participants' responding, we first examine whether the basic pattern of results observed in Experiment 1 was replicated here. The pertinent results are presented in Table 2.

As in the previous experiment, here too the participants tended to volunteer answers at both grain sizes in Phase 2, although they were somewhat more conservative in their choices than were the participants in Experiment 1, volunteering fine-grained answers about 30% of the time (1 participant volunteered no fine-grained answers), as compared with 41% in Experiment 1. This difference might have stemmed from the somewhat poorer Phase 1 performance in this experiment in regard to both fine-grained answers (28% correct vs. 32% correct in Experiment 1) and coarse-grained answers (65% correct vs. 75% correct in Experiment 1). The accuracy rates for the answers that the participants chose to provide in Phase 2 in the two experiments were similar (55% correct

Table 2
Experiment 2: Mean Proportions Correct for Fine-Grained and Coarse-Grained Answers in Phase 1, Calculated Separately for Items Volunteered at Each Grain Size in Phase 2, Along With Mean Percentages of Grain Choices and Overall Proportions Correct

Volunteered grain size (Phase 2)	Grain size of answer (Phase 1): Fine		Grain size of answer (Phase 1): Coarse		Percentage of choices (Phase 2)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Fine	.55	.23	.83	.16	30	21
Coarse	.21	.17	.59	.15	70	21
Overall	.28	.15	.65	.14		

in Experiment 2 vs. 59% correct in Experiment 1). Also, with regard to the items that participants chose to volunteer at each grain size, it again appears that participants tended to provide the coarse-grained answer when the fine-grained answer was relatively likely to be wrong and when the gain in accuracy from doing so was relatively large. First, the items provided at the fine grain size were much more likely to be correct at the fine grain size (55%) than were items provided at the coarse grain size (21%), $F(1, 31) = 50.55$, $p < .0001$ (the corresponding percentages in Experiment 1 were 50% and 21%). Second, the achieved increase in accuracy for the items provided at the coarse grain size ($.59 - .21 = .38$) was greater than the increase that would have been achieved by providing the other items at the coarse grain size ($.83 - .55 = .28$), $F(1, 31) = 7.53$, $p < .01$. On the whole, we conclude that the results of Experiment 2 replicated the pattern of performance observed in Experiment 1.

We now turn to the confidence data, which can shed some light on the mechanisms underlying the participants' choices of grain size. We begin by examining the effectiveness of the memory monitoring process.

Monitoring Effectiveness

The effectiveness of memory monitoring can be evaluated in terms of *calibration* (absolute correspondence) and in terms of *resolution* (relative correspondence). Calibration refers to the overall correspondence between the assessed and actual probabilities of being correct. The calibration data are presented graphically in Figure 1, plotted separately for the fine-grained and coarse-grained answers, according to the procedure commonly used in calibration research (Lichtenstein, Fischhoff, & Phillips, 1982). The probability assessments for the answers in Phase 1 were grouped into 12 levels (.0, .01–.10, .11–.20, . . . , .91–.99, 1.0). Proportion correct is plotted against mean assessed probability correct for the answers

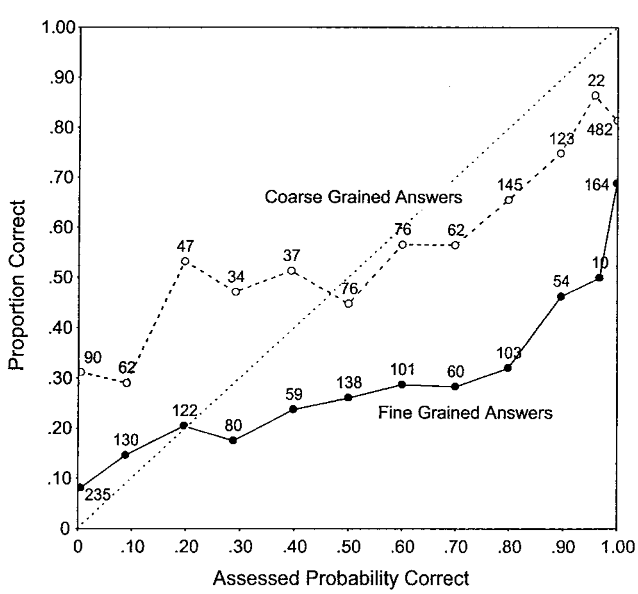

Figure 1. Calibration curves for the fine-grained and coarse-grained answers in Experiment 2 (Phase 1). The number of confidence judgments in each category appears beside each data point.

in each category, computed across participants. Perfect calibration is indicated by the diagonal line (e.g., 60% of the answers with a mean assessed probability correct of .60 should in fact be correct, and so forth).

The plots show a clear positive relationship between mean assessed probability correct and actual proportion correct for answers at both grain sizes. The general pattern of deviation from the diagonal is consistent with that of previous calibration studies (Erev, Wallsten, & Budescu, 1994), indicating overconfidence for answers with high assessed probabilities and underconfidence for answers with low assessed probabilities. This pattern is more pronounced for the coarse-grained answers than for the fine-grained answers. Overall, the mean assessed probabilities were .45 for the fine-grained answers and .71 for the coarse-grained answers, whereas the corresponding proportions correct averaged .28 and .65, respectively. Thus, the participants were overconfident for both grain sizes but more so for the fine-grained answers. Indeed, in terms of answers given with 100% confidence, only 69% of fine-grained answers and 81% of coarse-grained answers were in fact correct.

In comparing the two plots, it can be seen that they are very similar in shape. However, the actual proportion correct at the coarse grain size averages about .26 higher than the proportion correct at the fine grain size for each assessed-probability-correct category ($SD = .075$). That is, given equal confidence in a fine-grained or coarse-grained answer, the coarse-grained answer is much more likely to be correct. One might speculate that participants were perhaps deficient in taking into account the a priori probabilities of providing correct answers at the different grain intervals (cf. Tversky & Kahneman, 1974; but see Juslin, Winman, & Olsson, 2000, for possible alternative statistical explanations). Such a deficiency would have negative consequences for the control process, biasing it toward provision of fine-grained answers. This would occur under the satisficing model, because of overconfidence for the fine-grained answers with respect to the confidence threshold, or under the relative-utility model, because of bias in the calculation (and hence comparison) of the subjective expected utility of fine-grained and coarse-grained answers.

The second aspect of monitoring effectiveness, monitoring resolution, reflects people's ability to distinguish between correct and incorrect answers irrespective of the absolute levels of confidence judgments. As we have argued previously (Koriat & Goldsmith, 1996c), it is this aspect of monitoring that is most crucial for the strategic regulation of memory accuracy. We calculated resolution as the within-subject Kruskal–Goodman gamma correlation (see Nelson, 1984) between the assessed probability correct for each answer and whether or not the answer was correct. This measure averaged .48 and .47 for the fine-grained and coarse-grained answers, respectively. Interpreted in probabilistic terms, this level of monitoring implies that if a participant were presented with two answers, one correct and one incorrect, he or she would have about a 75% chance of selecting the correct one (Nelson, 1984). Thus, although far from perfect, the participants' judgments do distinguish between answers that are more likely and less likely to be correct at each grain size and, hence, may be useful in deciding which answers to provide at the fine grain size and which to provide at the coarse grain size.

Table 3
Experiment 2: Means of Within-Subject Pearson Correlations (N = 40 Items) Between Choice of Grain Size and Various Measures of Confidence (N = 32)

Measure	Coarse-grained confidence		Coarse minus fine confidence disparity		Fine–coarse confidence ratio		Choice of fine-grained answer	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Fine-grained confidence	.72***	.02	−.41***	.06	.64***	.04	.57***	.02
Coarse-grained confidence	—		.30***	.06	.09	.07	.41***	.03
Coarse minus fine confidence disparity			—		−.79***	.03	−.23***	.05
Fine–coarse confidence ratio					—		.35***	.04

*** $p < .001$ (significant difference from zero).

Control Process

We now turn to analyses that can shed some light on the nature of the process underlying control of grain size. We begin with an examination of the pattern of simple correlations between confidence and choice of grain size before moving on to more systematic logistic regression analyses.

Confidence–choice correlations. Table 3 presents mean intercorrelations between several measures of confidence and the choice to volunteer the fine-grained answer. These measures include the assessed probability correct for the fine-grained answer (P_{FINE}), the assessed probability correct for the coarse-grained answer (P_{COARSE}), and two derived measures: confidence disparity ($P_{\mathrm{COARSE}} - P_{\mathrm{FINE}}$) and confidence ratio ($P_{\mathrm{FINE}}/P_{\mathrm{COARSE}}$). The derived measures were included as alternative indexes of the relative disparity between confidence in the fine-grained and coarse-grained alternatives. Pearson correlations[4] between the measures were calculated within participants and then averaged.

Clearly, confidence is generally predictive of choice of grain size. The strongest predictor is confidence in the fine-grained answer: The more confident one is in the fine-grained answer, the more likely one is to provide it. This result is consistent with the satisficing model outlined earlier, in which choice of grain size is determined solely or primarily by confidence in the fine-grained answer. Note, however, that confidence in the coarse-grained answer is also positively correlated with choice of the fine-grained answer. That is, the higher one's confidence in the coarse-grained answer, the more likely one is to provide the fine-grained answer. Of course, this positive correlation might stem solely from the very high correlation between confidence in the coarse-grained answer and confidence in the fine-grained answer to each item, a possibility addressed later in the logistic regression analyses.

Turning to the two relative confidence measures, both the magnitude of the discrepancy between confidence in the coarse- and fine-grained answers and the ratio of the two confidence assessments were rather weak predictors of choice of grain size. Consistent with the relative expected-utility model described earlier, participants were more likely to provide the fine-grained answer when its assessed probability correct was relatively close to that of the coarse-grained answer, in terms of either the absolute difference or the proportional difference between the two assessments. Of course, these correlations too might have stemmed solely from the correlation between the relative confidence measures and confidence in the fine-grained answer.

In sum, the overall pattern of intercorrelations among measures of confidence and choice of grain size in this experiment seems most consistent with the operation of a simple satisficing strategy in which choice of grain size is perceived not as a symmetric choice between two options (fine and coarse) but, rather, as a decision regarding whether or not to risk volunteering the fine-grained answer. To buttress this conclusion, we present a more systematic comparison of alternative models in the next section.

Logistic regression analyses. Logistic regression analyses were carried out to address several questions arising from the preceding results. First, is confidence in the fine-grained answer indeed the primary or sole determinant of the choice of grain size, as implied by the satisficing model? Second, does confidence in the coarse-grained answer make any contribution at all, and, if so, does it contribute negatively to the choice of the fine-grained answer, as implied by the relative-utility model (and suggested by the correlations involving the relative confidence measures), or is its contribution actually positive (as suggested by the raw correlation)? Third, does the relative informativeness of the fine- and coarse-grained answers also contribute to the choice of grain size?

Table 4 presents the results of a series of logistic regression analyses, each using one or more predictors to derive the probability of choosing the fine-grained answer. The analyses were conducted across all participants and items.[5]

For the purpose of comparison, Analyses 1–4 correspond to the four simple correlations that were examined earlier in the within-

[4] For the purpose of comparison, we also calculated the (ordinal-scale) within-subject Kruskal–Goodman gamma correlation between each measure and choice of the fine-grained answer. These correlations averaged .83 for fine confidence, .82 for coarse confidence, −.43 for the coarse–fine confidence disparity, and .57 for the fine–coarse confidence ratio. All of the correlations were significantly different from zero, and significantly different from each other, except for the two raw correlations (fine confidence and coarse confidence).

[5] Because the analyses were conducted across participants, it is conceivable that individual differences in overall knowledge or overall confidence

Table 4
Results of Logistic Regression Analyses Predicting Choice of Fine Grain Size in Experiment 2 (Phase 2)

Analysis (model)	Standardized regression coefficient: Fine-grained confidence	Coarse-grained confidence	Coarse minus fine confidence disparity	Fine–coarse confidence ratio	Informativeness: $\mathrm{Log}_{CW} - \mathrm{Log}_{FW}$	Model statistic[a]: G	R^2_L
1	.58**					416.9	.272
2		.46**				265.2	.173
3			−.24**			72.7	.047
4				.33**		141.4	.092
5	.47**	.14*				423.1	.276
6	.62**		.11*			423.1	.276
7	.73**			−.28**		436.4	.284
8					.14**	21.5	.014
9	.47**	.13*			.11**	436.7	.285
10	.71**			−.26**	.09**	446.9	.291

Note. All models (G statistics) are significant at $p < .01$. Models 5 and 6 are statistically equivalent. Log_{CW} = logarithm of coarse grain width; Log_{FW} = logarithm of fine grain width.
[a] See Footnote 6 for explanation.
* $p < .05$. ** $p < .01$.

subject analyses, yielding the same pattern of results: Confidence in the fine-grained answer was the strongest single predictor of choice of grain size,[6] followed by confidence in the coarse-grained answer and, least of all, the two relative confidence measures. Also, as before, confidence in the coarse-grained answer was positively related to choice of the fine-grained answer, raising the question of whether this relationship was partly or fully spurious (i.e., stemming from the high positive correlation between fine- and coarse-grained confidence).

This question was addressed in Analysis 5, which included confidence in both the fine-grained and coarse-grained answers as predictors in the same regression model. This analysis again yielded a strong contribution of fine-grained confidence to choice of grain size. Surprisingly, however, when confidence in the fine-grained answer was partialed out, a small but significant positive influence of confidence in the coarse-grained answer remained. That is, even when confidence in the fine-grained answer was held constant, increasing confidence in the coarse-grained answer appeared to increase the likelihood of providing the fine-grained answer. This result is clearly inconsistent with the relative-utility model, which had seemed to gain some support from Analyses 3 and 4. However, it turns out that the relationships identified in those analyses were in fact spurious: Analysis 6, statistically equivalent to Analysis 5 (a linear combination of the same predictor variables), showed perhaps more clearly that when confidence in the fine-grained answer was held constant, a larger disparity in confidence between the coarse-grained and fine-grained answers increased the likelihood of providing the fine-grained answer. A similar result was found for the proportional difference in confidence between the coarse-grained and fine-grained answers (Analysis 7).

How should this rather counterintuitive contribution of coarse-grained confidence be interpreted? We can only speculate at this point. One possibility is that the decision to provide the fine-grained answer is based on a more global confidence judgment that is strengthened by confidence in both the fine-grained and the coarse-grained answers. If so, however, confidence in the fine-grained answer would still appear to be the dominant component of this global confidence judgment.

We can now offer answers to the first two questions that motivated these analyses. First, in partial support of the simple satisficing model, confidence in the fine-grained answer was found to be the primary contributor to choice of grain size. Confidence in the coarse-grained answer also appeared to contribute, but this contribution was quite small (compare the R^2_L values of Models 1 and 5). Second, the small contribution of coarse-grained confidence was actually the opposite of what one would expect under the relative-utility model: Increasing the confidence disparity between the coarse-grained and fine-grained answers (which should increase the relative expected utility of the coarse-grained answer)

might also contribute to the results. However, when such individual differences were partialed out from each analysis (by including each participant's mean accuracy score and mean confidence rating for the fine-grained and coarse-grained answers in Phase 1 as additional predictors), the pattern of results remained essentially unchanged. As an additional check, we conducted a repeated measures generalized linear model analysis using the generalized estimating equation method (which corrects for intraindividual item correlation; Liang & Zeger, 1986; Lipsitz, Laird, & Harrington, 1991). Again, the same pattern of results and significance levels was obtained.

[6] For the unfamiliar reader, and to clarify the notation, the overall goodness of fit of a logistic regression model is indexed and tested with the G statistic (Hosmer & Lemeshow, 1989), which is analogous to explained variance (SS_R) in linear regression analyses. An analogue to proportion of explained variance (SS_R/SS_T) is the R^2_L statistic (Hosmer & Lemeshow, 1989), which reflects proportionate reduction in badness of fit relative to the null (intercept-only) model. Interpretation of standardized coefficients is analogous to their interpretation in linear regression.

increased the likelihood that the fine-grained answer would be provided.

Let us now turn to the third question, which concerns the contribution of informativeness. So far, the contribution of informativeness has been inferred indirectly, from the tendency of participants to risk volunteering the fine-grained answer even though their chances of being wrong are almost always greater than for the coarse-grained alternative. Further insight into the contribution of informativeness can be gained by focusing on interitem differences in the informativeness of the fine-grained answer relative to that of the coarse-grained alternative. If the goal of providing an informative answer does indeed draw participants toward the fine-grained alternative, we would expect this pull to be particularly strong when the difference in informativeness between the fine-grained and coarse-grained answers is relatively large.

To examine this idea, we followed Yaniv and Foster's (1995) proposal for quantifying the perceived informativeness of interval-type estimates and calculated the difference in the informativeness of the answers at the two grain sizes as the difference in the natural log of the interval width of the two answers.[7] Analysis 8 shows that this difference had a weak but significant relationship to choice of grain size: The larger the difference, the more likely participants were to choose the fine-grained answer. Moreover, when this variable was included as a predictor in addition to the confidence variables (Analyses 9 and 10), its contribution remained significant, as did that of the confidence variables. The implication is that perceived informativeness and assessed probability correct (i.e., accuracy) make independent contributions to the control decision (cf. Yaniv & Foster, 1995). This finding also discounts the possibility that the positive contribution of confidence in the coarse-grained answer to choice of the fine-grained answer is spurious, stemming simply from the fact that the less informative the coarse answers (i.e., the wider the grain width), the more likely they are to have higher assessed probabilities assigned to them.

In sum, the results of this experiment suggest that control over grain size, as with control of report option, is based on confidence in the correctness of one's answers. When participants are relatively confident about the correctness of the fine-grained answer, they tend to provide it; otherwise, they prefer to provide the more coarsely grained answer. Somewhat surprisingly, higher confidence in the coarse-grained answer tends to reinforce the decision to provide the fine-grained answer, even when confidence in the fine-grained answer is held constant. This contribution, however, is small. In fact, in terms of actual predictive power, measured as the proportion of correct predictions of the participants' actual grain choices, none of the models significantly improved on the model that included fine confidence alone (Analysis 1), which yielded a hit rate of 81%. This hit rate is not only substantially better than the 50% chance rate, but, when computed separately for each participant, it is also significantly better than the 70% base rate that could be achieved by always predicting "coarse," $F(1, 31) = 15.74$, $p < .001$, and the 72% hit rate that could be achieved in logistic regression analyses using each participant's mean confidence level alone (see Footnote 5), $F(1, 31) = 20.92$, $p < .0001$.

The role of informativeness was also addressed in this experiment, in terms of interitem differences in the informativeness of the fine-grained answer relative to that of the coarse-grained answer. The contribution of this variable, although significant, was also quite small, perhaps because it tapped interitem variability in the informativeness difference between the fine-grained and coarse-grained answers ($SD = 1.3$) but not the (average) informativeness difference itself ($M = 2.1$), which should always predispose participants toward providing the fine-grained answer. In the next experiment, we addressed the contribution of informativeness in a different way, by providing and manipulating explicit payoffs for fine-grained and coarse-grained answers.

Experiment 3

We have proposed that the choice of grain size in memory reporting is guided by the attempt to achieve a balance between the opposing goals of accuracy and informativeness. The results of Experiments 1 and 2 support this proposition. On the one hand, a strong relationship was found between choice of grain size and both assessed probability correct and proportion correct, presumably reflecting the goal of accuracy. On the other hand, the goal of informativeness was evidenced by (a) the tendency of participants to risk volunteering the fine-grained answer even though their chances of being wrong were almost always greater than for the coarse-grained alternative and (b) the finding that participants were more likely to choose the fine-grained answer when the difference in the precision of the fine-grained and coarse-grained alternatives was relatively large.

Of course, the findings from both of these experiments are essentially correlational. Thus, the purpose of Experiment 3 was to strengthen our claim regarding the strategic nature of control over grain size by experimentally manipulating explicit incentives for informativeness versus accuracy. We used the same memory test as before; for half of the items (the high-informativeness-incentive condition), participants were offered NIS 5 (about $1.25) for each correct fine-grained answer but only NIS 1 (about $0.25) for each correct coarse-grained answer, whereas, for the other half of the items (the low-informativeness-incentive condition), they were offered NIS 2 (about $0.50) for each correct fine-grained answer and NIS 1 (about $0.25) for each correct coarse-grained answer. In both conditions (i.e., for all items), participants were penalized NIS 1 for each wrong answer at either grain size.[8] If participants do regulate the grain size of their answers in a strategic manner, we would expect their choices to be sensitive to differences in the relative weight given to informativeness and accuracy in the two incentive conditions, evidenced by a greater willingness to risk accuracy for informativeness for the subset of items that rewards informativeness more heavily.

[7] The interval width is calculated as (upper bound − lower bound) + 1; the interval width of specific answers is therefore equal to one. Yaniv and Foster's (1995) proposal to model perceived informativeness as a logarithmic function of grain size (interval width) is based on the well-known psychophysical law (Fechner) that human responses to changes in objective magnitudes approximate a concave function. Their results yielded a good fit between predictions that make use of this index and participants' subjective evaluations.

[8] Even without this penalty, there would be an implicit incentive for accuracy: Wrong answers do not receive any bonus. However, we believe that an explicit penalty for wrong answers is perhaps closer to the social and personal penalty paid for reporting wrong information in real life.

A second reason for introducing and manipulating explicit payoffs is that it can allow a further and somewhat more stringent evaluation of the satisficing and relative-utility models of the grain-control process. Not only can we evaluate the models in terms of the nature of the relationship between confidence and choice of grain size (as we did in Experiment 2), we can examine whether the effect of the incentive manipulation on choice of grain size is mediated in the manner specified by each model. According to the satisficing model, a relatively high reward for informativeness should induce participants to lower the confidence threshold that must be passed before a fine-grained answer will be provided. In contrast, under the relative-utility model, the higher reward should increase the subjective expected utility of the fine-grained answers relative to the coarse-grained answers, which in turn should lower the fine–coarse confidence ratio beyond which the choice of the fine-grained answer becomes preferable (as described subsequently). Thus, although both models predict that more fine-grained answers will be provided in the high-informativeness-incentive condition than in the low-incentive condition, they yield different predictions regarding how the incentive manipulation will be manifested in terms of changes in the underlying response criterion.

Finally, although the results of Experiment 2 clearly favor the descriptive adequacy of the satisficing model, such a conclusion would be strengthened further if it were to hold true under conditions that were most conducive to the use of a relative expected-utility strategy. Indeed, unlike in Experiment 2, in which the payoffs for accuracy and informativeness were not well defined, in this experiment the explicit payoff schedule allowed participants—at least in principle—to calculate the subjective expected utility of each of their answers. Given the explicit payoffs in this experiment, the subjective expected utility of providing an answer at each grain size can be calculated according to the following formulas: $EU_{FINE} = (P_{FINE} \times REWARD_{FINE}) - [(1 - P_{FINE}) \times PENALTY]$ and $EU_{COARSE} = (P_{COARSE} \times REWARD_{COARSE}) - [(1 - P_{COARSE}) \times PENALTY]$, where EU is the subjective expected utility of the answer at each grain size, P is the assessed probability that the answer is correct at each grain size, REWARD is the reward for providing a correct answer at each grain size, and PENALTY is the penalty for providing a wrong answer. Substituting in the rewards and penalties of the two incentive conditions and reducing the equations yields the following: low incentive, $EU_{FINE} = 3P_{FINE} - 1$ and $EU_{COARSE} = 2P_{COARSE} - 1$, and high incentive, $EU_{FINE} = 6P_{FINE} - 1$ and $EU_{COARSE} = 2P_{COARSE} - 1$.

Solving for $EU_{FINE} > EU_{COARSE}$, one can also calculate the (normative) fine–coarse confidence criterion ratio beyond which the fine-grained answer should be provided in each incentive condition: In the low-informativeness-incentive condition the fine-grained answer should be provided when $P_{FINE} > .67P_{COARSE}$, and in the high-informativeness-incentive condition that answer should be provided when $P_{FINE} > .33P_{COARSE}$. Arguably, such calculations are much more demanding in terms of mental-computational resources than what is required by the "fast and frugal" (Gigerenzer & Todd, 1999) satisficing model, and we are not suggesting that participants would consciously carry out such calculations in deciding which answers to provide. It is plausible, however, that to the extent that the relative payoffs for providing fine- and coarse-grained answers are made more clear, participants might be more likely to base their choices on the relative assessed probabilities of these answers, in line with the relative-utility model.

Method

Participants

Thirty-two Hebrew-speaking psychology students at the University of Haifa, 9 men and 23 women, participated in the experiment for course credit and the chance to win up to NIS 100 (approximately $25).

Materials

Experiment 3 involved the same general-knowledge test used in the previous experiments, with blanks for confidence (assessed probability correct) judgments as in Experiment 2. In addition, a payoff table was printed out that specified the payoff for each correct answer to each item at each grain size. For half of the items, the rewards for the fine-grained and coarse-grained answers were specified according to the high-informativeness-incentive condition (NIS 5 for a correct fine-grained answer and NIS 1 for a correct coarse-grained answer), whereas the other half were specified according to the low-informativeness-incentive condition (NIS 2 for a correct fine-grained answer and NIS 1 for a correct coarse-grained answer). The penalty for wrong answers (NIS 1), which was the same for all answers regardless of incentive condition, was also listed for each item as a reminder. Two forms of the payoff table were prepared. For Form A, 20 items were randomly assigned to the high-informativeness-incentive condition, and the remaining items were assigned to the low-informativeness-incentive condition. For Form B, the item-incentive assignments of Form A were reversed. Half of the participants received Form A, and the other half received Form B.

Procedure

The experiment was administered to all participants as a group in a single session that lasted about 1.25 hr. Each participant was given a copy of the knowledge test and the accompanying instruction sheet for Phase 1, which was the same as in Experiment 2 (requiring answers to all questions at both grain sizes, along with confidence judgments). After each participant had completed Phase 1, the confidence judgments were cut away from the test form (as in Experiment 2), and the test form was returned to him or her together with the instructions and payoff table for Phase 2. The instructions for Phase 2 were similar to those of the preceding experiments (allowing the participants to choose which of the two alternative answers to provide for each item), but this time the participants were instructed to make their choice of answers in light of the payoff schedule for each item. They were not explicitly told about the two payoff (informativeness-incentive) conditions, but simply that they should check the accompanying table for the specific potential payoffs (and penalty) for each item. They were also told that their actual winnings would be based on a random sample of half of the items (this was done simply to reduce the overall participant payments). Again, they were not allowed to change any answers between phases (they were given pens with different colors of ink for Phases 1 and 2).

Results and Discussion

Presumably stemming from the more limited supervision possible in a large-group administration, a somewhat larger number of observations than in the previous two experiments (93 of 1,280 items) had to be omitted from the analyses, primarily because of deviation from the specified grain sizes. One participant, who had an inordinately high proportion of such items (18 of 40), was dropped from the analyses entirely.

Table 5
Experiment 3: Mean Proportions Correct for Fine-Grained and Coarse-Grained Answers in Phase 1, Calculated Separately for Items Volunteered at Each Grain Size in Phase 2 and for Informativeness-Incentive Items, Along With Mean Percentages of Grain Choices

Informativeness incentive (Phase 2)	Volunteered grain size (Phase 2)	Grain size of answer (Phase 1): Fine *M*	Fine *SD*	Coarse *M*	Coarse *SD*	Percentage of choices (Phase 2) *M*	*SD*
Low	Fine	.65	.30	.85	.17	19	15
	Coarse	.17	.10	.58	.18	81	15
High	Fine	.45	.27	.79	.22	29	21
	Coarse	.17	.13	.56	.17	71	21
Overall	Fine	.55	.30	.82	.20	24	19
	Coarse	.17	.12	.57	.17	76	19

Performance Data

Table 5 presents the accuracy results for this experiment in terms of the mean proportions correct for fine- and coarse-grained answers, calculated separately for the items volunteered at each grain size in Phase 2. It can be seen that the same basic pattern was found here (using explicit incentives for accuracy and informativeness) as in the first two experiments, both overall and separately for each incentive condition (i.e., subset of items). Overall, 24% of the items were volunteered at the fine grain size and 76% at the coarse grain size in Phase 2 (3 participants volunteered no fine-grained answers). Items that participants chose to volunteer at the coarse grain size were much less likely to be correct at the fine grain size (.17) than were items that participants chose to volunteer at the fine grain size (.55), $F(1, 27) = 82.76$, $p < .0001$, and the accuracy increase that was achieved for these items ($.57 - .17 = .40$) was greater than the increase that would have been achieved by providing the other items at the coarse grain size ($.82 - .55 = .27$), $F(1, 27) = 13.07$, $p < .01$.

Turning now to the effects of the informativeness-incentive manipulation, as expected, participants were in fact more willing to sacrifice accuracy for informativeness when the payoff for informativeness was high: More fine-grained answers were provided for high-incentive items (29%) than for low-incentive items (19%), $F(1, 30) = 14.23$, $p < .001$, and the high-incentive fine-grained answers that were provided had a lower accuracy rate (.45) than did the low-incentive fine-grained answers (.65), $F(1, 23) = 4.59$, $p < .05$. The overall accuracy rate of the volunteered high-incentive items (.51), however, was only marginally lower than that of the volunteered low-incentive items (.56), $F(1, 30) = 2.73$, $p < .06$ (one tailed), presumably because of the equivalent accuracy of the coarse-grained answers provided in both conditions (.56 and .58 for the high- and low-incentive conditions, respectively, $F < 1$).

Further analyses examined the participants' performance in terms of the monetary payoffs that resulted from their grain choices. The participants earned an average of NIS 14.6 for their answers (based on all 40 answers; *M*s = NIS 9.2 and NIS 20.0 for the low-incentive and high-incentive items, respectively). This amount was significantly higher than the NIS 3.5 payoff that would have been earned by always volunteering the fine-grained answer (*M*s = NIS −10.9 and NIS 17.9 for the low-incentive and high-incentive items, respectively), the NIS 9.3 payoff that would have been earned by always volunteering the coarse-grained answer (*M*s = NIS 8.9 and NIS 9.6 for the low-incentive and high-incentive items, respectively), or the NIS 6.4 expected payoff from providing a random mix of fine-grained and coarse-grained answers (*M*s = NIS −1.0 and NIS 13.8 for the low-incentive and high-incentive items, respectively). Thus, overall, the participants were rather effective in choosing answers at the two grain sizes that would increase their average earnings.

However, the gain in earnings that can be attributed to the participants' adjustment of the grain-size control policy in accordance with the payoff schedule for each item was less impressive. One would expect that the participants' grain-size choices for the high-incentive and low-incentive items would be particularly well suited to the high-incentive and low-incentive payoff schedules, respectively. However, this was not the case: When the low-incentive payoff schedule was applied to the participants' chosen answers for the high-incentive items, and vice versa,[9] the resulting payoff of NIS 14.0 was about the same as the actual NIS 14.6 payoff (obtained by applying the proper payoff schedule to the participants' answers in each incentive condition; $F < 1$). More direct results regarding adjustment of the response criteria are reported subsequently.

Monitoring Effectiveness

The results for monitoring effectiveness replicated the overall pattern observed in Experiment 2. With regard to calibration, participants were overconfident in their fine-grained answers (mean assessed probability correct = .35 vs. mean proportion correct = .24) but not in their coarse-grained answers (mean assessed probability correct = .60 vs. mean proportion correct = .62). The shapes of the fine-grain and coarse-grain calibration plots (see Figure 2) and the consistent differences between the two plots in the proportions correct across the range of assessed probabilities ($M = .27$, $SD = .072$) were quite similar to those observed in Experiment 2. In addition, monitoring resolution was again moderate (almost identical to that in Experiment 2), the within-subject gamma correlations averaging .47 and .49 for the fine-grained and

[9] We thank John Dunlosky for suggesting this analysis.

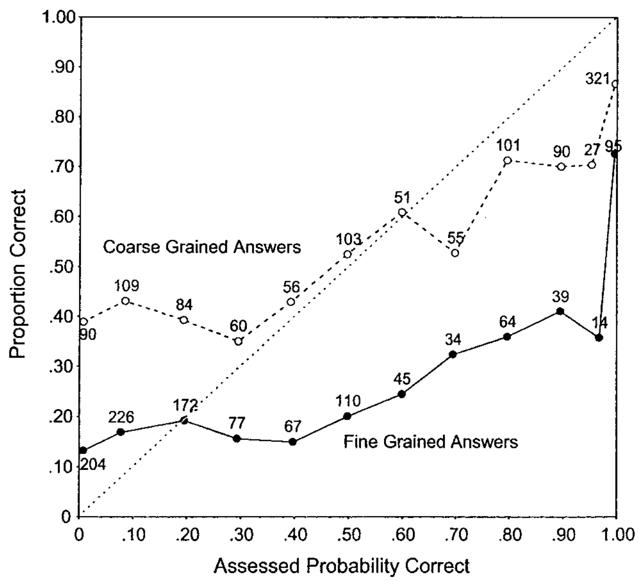

Figure 2. Calibration curves for the fine-grained and coarse-grained answers in Experiment 3 (Phase 1). The number of confidence judgments in each category appears beside each data point.

coarse-grained answers, respectively. Again, we conclude that the monitoring judgments were sufficiently diagnostic of the correctness of the answers at each grain size to be useful in controlling the grain-size choices.

Control Process

Let us turn now to an examination of the underlying control process and how this process was affected by the informativeness incentive.

Simple intercorrelations. For comparison, Table 6 presents the mean intercorrelations between the choice of grain size and the four confidence measures, calculated in the same manner as in Experiment 2, except that here the within-subject correlations were computed separately for the high-incentive and low-incentive items before averaging. We refer to several of these correlations in the context of the logistic regression analyses reported next.

Logistic regression analyses. The results from a series of logistic regression analyses, again analyzed across items and participants,[10] are summarized in Table 7. In line with the performance results just reported, the results of Analysis 1 showed that participants' grain choices were sensitive to the informativeness incentive, with the higher informativeness incentive increasing the likelihood that they would choose the fine-grained answer. This contribution was included and remained significant in all of the other regression analyses.

Was the process underlying the participants' control of grain size in this experiment different from what was found earlier when implicit incentives for accuracy and informativeness were used? It seems not. Examination of the results shows a striking similarity between the pattern obtained here and that observed in Experiment 2. Confidence in the fine-grained answer and confidence in the coarse-grained answer contributed positively to choice of the fine grain size both individually (Analyses 2 and 3, respectively; see also simple correlations in Table 6) and when included together in a single analysis (Analysis 6). Thus, once again we found that even when confidence in the fine-grained answer was held constant, increasing confidence in the coarse-grained answer increased the likelihood of providing the fine-grained answer.

If participants had been using a relative-utility comparison process, the regression model tested in Analysis 5 (using the fine–coarse confidence ratio) would have been expected to yield the best fit to the data. Although the fit of that model was significant (see also simple correlation in Table 6), it was by far inferior to the model using absolute level of fine confidence (Analysis 2). Moreover, when confidence in the fine-grained answer and the confidence-ratio variable were included together (Analysis 8), again the contribution of the confidence ratio was actually negative: Holding confidence in the fine-grained answer constant, the higher that confidence was relative to confidence in the coarse-grained answer, the less likely was the participant to choose the fine-grained answer. Similar results, of course, were obtained for the absolute (coarse − fine) confidence disparity (comparing Analyses 4 and 7).

It appears, then, that even when informativeness and accuracy are explicitly rewarded in monetary terms, participants prefer to base their choices on absolute confidence levels, with increased confidence in both the fine-grained and the coarse-grained answers increasing the likelihood of choosing to provide the fine-grained answer. Again, the primary determinant appeared to be confidence in the fine-grained answer (compare the R^2_L values of Models 2 and 7). This is nicely illustrated in Figure 3, which plots the probability of providing a fine-grained answer as a function of informativeness incentive and of confidence in the fine-grained answer (with confidence categorized in the same manner as for the earlier calibration plots; see Figures 1 and 2). The figure also shows how the effect of the informativeness-incentive manipulation can be captured as an effect on the fine-grained confidence threshold.

Relatedly, one difference from the pattern observed in Experiment 2 concerns the contribution of interitem differences in the relative informativeness of the answers at the two grain sizes (calculated in terms of the natural log of the interval width of the answer). Although this variable does, at first glance, appear to make a contribution beyond that of the explicit payoffs provided (Analysis 9), this contribution disappears when the confidence variables are also included in the analysis (Analyses 10 and 11). This would seem to suggest that the explicit payoffs in this experiment were effective in overriding the more intrinsic perceived rewards of informativeness that are captured in the interval-width measure.

Response-criterion analyses. A further comparison of the satisficing and relative-utility models can be made with respect to the effect of the incentive manipulation on the underlying response criterion. For each participant, we calculated the fit that could be obtained between each of these models and the participant's actual

[10] As in Experiment 2 (see Footnote 5), note that the same pattern of effects and significance levels was obtained in a repeated measures generalized linear model analysis using the generalized estimating equation method and when participants' Phase 1 accuracy scores and mean confidence ratings for fine- and coarse-grained answers were partialed out by including them as additional predictors in the regression models.

Table 6
Experiment 3: Means of Within-Subject Pearson Correlations (Computed Separately for the High- and Low-Incentive Items Before Averaging) Between Choice of Grain Size and Various Measures of Confidence (N = 62)

	Coarse-grained confidence		Coarse minus fine confidence disparity		Fine–coarse confidence ratio		Choice of fine-grained answer	
Measure	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Fine-grained confidence	.71***	.02	−.21***	.04	.54***	.04	.64***	.02
Coarse-grained confidence	—		.47***	.04	.01	.05	.48***	.02
Coarse minus fine confidence disparity			—		−.72***	.03	−.17***	.04
Fine–coarse confidence ratio					—		.36***	.04

*** $p < .001$ (significant difference from zero).

grain choices in Phase 2 for each incentive condition (set of items), with the criterion value for the satisficing model (i.e., the value of the confidence threshold above which the fine-grained answer will be provided) and that for the relative-utility model (i.e., the value of the fine–coarse confidence ratio above which the fine-grained answer will be provided) as open parameters.[11]

The criterion values for each model were estimated through a procedure adapted from Koriat and Goldsmith (1996c). The criterion estimate for each participant in each incentive condition was the value that would maximize the proportion of correctly predicted grain choices when applied to the participant's confidence judgments. If a range of criterion values yielded an equivalent correct-prediction rate, the midpoint of the range was chosen. The mean estimated criterion values and correct-prediction rates for each model in each incentive condition are shown in Table 8.

Looking first at the satisficing (simple threshold) model, we see that informativeness incentive had a substantial effect on the estimated criterion values (cf. Figure 3), with participants setting a lower criterion for the high informativeness incentive (.58) than for the low informativeness incentive (.74), $F(1, 30) = 14.17$, $p < .001$. Moreover, this model could account quite well for the participants' actual grain choices, with a correct prediction rate of about 90% across the two incentive conditions. Turning to the relative-utility (confidence-ratio) model, here too the informativeness incentive had the expected effect on the estimated criterion values, with participants requiring a higher ratio of fine-grained to coarse-grained confidence before they were willing to provide a fine-grained answer under the low informativeness incentive (.80) than under the high informativeness incentive (.69), $F(1, 30) = 7.88$, $p < .01$. These estimates, however, were based on a much poorer fit to the data, with an average correct-prediction rate of only 76%. Note also that the estimated confidence-ratio criterion values deviated substantially from the normative values of .67 for the low-incentive condition and .33 for the high-incentive condition. At the very least, one would expect a larger difference between the criterion values for the two incentive conditions than was observed here. Thus, examination of the underlying response criterion, and the effects of the incentive manipulation on this criterion, also favors the satisficing model over the relative-utility model.

In sum, the results of this experiment join the results of Experiment 2 in showing that participants are moderately able to monitor the correctness of their answers at different grain sizes and that they tend to control the grain size of their answers on the basis of their monitoring, taking into account the implicit or explicit incentives for accuracy and informativeness. On the whole, the results suggest that the choice of grain size depends primarily on whether or not rememberers are confident enough to provide the more informative answer (which, all else equal, they would prefer to provide) and not on the calculation and comparison of expected utilities of more informative and less informative answers.

General Discussion

The recent upsurge of interest in real-life remembering has brought with it a myriad of challenging metatheoretical and methodological issues (e.g., Fowler, 1991; Koriat & Goldsmith, 1996b). Notably, it has engendered an expanded conception of memory and memory functioning in which memory is viewed as a multifaceted tool used in the service of achieving personal and social goals (e.g., Neisser, 1988, 1996; Winograd, 1996). This conception requires the consideration of a wider range of memory and metamemory processes than does the traditional "storehouse" view and motivates an examination of these processes within a broader functional context (e.g., Barnes et al., 1999; Goldsmith & Koriat, 1999; Leippe, 1994; Neisser & Fivush, 1994; Nelson & Narens, 1994; Ross & Buehler, 1994; Yzerbyt, Lories, & Dardenne, 1998). As Neisser has eloquently argued, remembering is like "doing" (Neisser, 1996), and hence any complete theory of memory "retrieval" will need to deal with "the reason for retrieval, . . . with persons, motives, and social situations" (Neisser, 1988, p. 553).

Our work has focused on situations in which the rememberer's goals are presumably served by providing both accurate and com-

[11] As mentioned earlier, the payoff schedule used in this experiment yields "normative" fine–coarse confidence criterion values for the relative-utility model of .33 and .67 for the low- and high-incentive conditions, respectively. Nevertheless, to afford a more fair comparison with the simple threshold model (which does not allow normative criterion values to be derived), we allowed the criterion values for both models to vary.

Table 7
Results of Logistic Regression Analyses Predicting Choice of Fine Grain Size in Experiment 3 (Phase 2)

Analysis (model)	Standardized regression coefficient: Fine-grained confidence	Coarse-grained confidence	Coarse minus fine confidence disparity	Fine–coarse confidence ratio	Informativeness: $Log_{CW} - Log_{FW}$	Informativeness incentive[b]	Model statistic[a]: G	R^2_L
1						.12**	15.96	.013
2	.60**					.15**	426.7	.337
3		.53**				.09**	351.3	.278
4			−.09**			.11**	26.5	.021
5				.29**		.12**	122.6	.097
6	.36**	.29**				.13**	455.2	.360
7	.63**		.20**			.13**	455.2	.360
8	.71**			−.26**		.13**	443.8	.351
9					.17**	.12**	51.3	.041
10	.35**	.30**			.04	.13**	457.2	.361
11	.71**			−.26**	.03	.14**	445.1	.352

Note. All models (G statistics) are significant at $p < .01$. Models 6 and 7 are statistically equivalent. Log_{CW} = logarithm of coarse grain width; Log_{FW} = logarithm of fine grain width.
[a] See Footnote 6 for explanation.
[b] 0 = low, 1 = high.
* $p < .05$. ** $p < .01$.

plete memory reports. In such situations, we ask, what processes do rememberers employ in attempting to regulate the quantity and accuracy of the information that they report? One means that we have studied extensively has been the use of report option to screen out incorrect items of information from the memory report. Our studies have shown that not only adults but even young children are sensitive to situational incentives for accurate reporting and are able to regulate the accuracy and quantity of their memory reports accordingly (Koriat & Goldsmith, 1994, 1996c; Koriat, Goldsmith, Schneider, & Nakash-Dura, 2001). The principle governing this regulation is an accuracy–quantity trade-off.

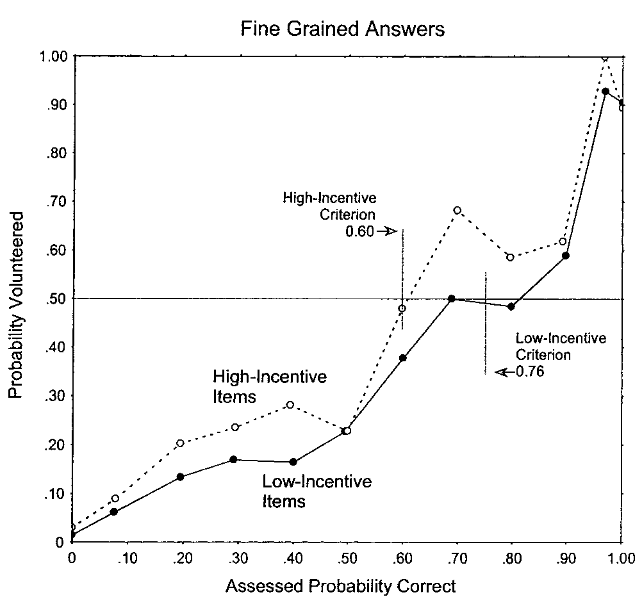

Figure 3. Probability of volunteering a fine-grained answer in Experiment 3 (Phase 2) as a function of informativeness incentive and of confidence (assessed probability correct) in the fine-grained answer. The confidence-criterion value derived by logistic regression, above which the choice of the fine-grained answer would be predicted by the logistic regression model (and below which the choice of the coarse-grained answer would be predicted), is also indicated for each incentive condition.

In this article, we have attempted to bring one further aspect of real-life memory regulation into the laboratory for experimental investigation. In many situations, rememberers are also faced with an accuracy–informativeness dilemma involving grain size: Should they provide a more informative answer that is relatively likely to be wrong or a less informative answer that is relatively likely to be right? How do rememberers resolve this dilemma?

We addressed this question using a general conceptual framework similar to the one developed earlier for the study of report option, in which metacognitive monitoring and control processes are used to regulate the accuracy and informativeness of the reported information. As discussed shortly, this conception of control of memory grain size differs substantially from more traditional treatments of memory grain size, such as memory for gist versus verbatim detail (Sachs, 1967), which have focused almost exclusively on the nature of the underlying memory representations and how these representations might constrain the information that is later retrieved (or reconstructed). Our conception does not conflict with existing approaches; rather, it adds further strategic decisional components that have hitherto been neglected. In this regard, it has more in common with work on judgmental estimates under uncertainty in the context of decision-making research (e.g., Budescu & Wallsten, 1995; Yaniv & Foster, 1995, 1997).

In discussing the potential utility of our approach, we first summarize the main findings of this study and examine their theoretical implications. We then address some limitations of the current study, presenting informal observational data that highlight aspects of the real-world control of grain size that were not

Table 8
Experiment 3: Mean Criterion Values and Percentages of Correct Predictions of Participants' Grain Choices for High- and Low-Informativeness Items, Comparing Two Models

	Fine confidence threshold model				Fine–coarse confidence ratio model			
	Low informativeness		High informativeness		Low informativeness		High informativeness	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Criterion value	.74	.15	.58	.24	.80	.15	.69	.24
Percentage of correct predictions	92.9	8.0	88.1	10.8	77.1	16.3	75.2	14.1

incorporated into the current experiments. Finally, we compare our approach to the control of memory grain size with the approach underlying the more traditional work on graininess in memory, underscoring the potentially complementary contributions of each approach to a more complete understanding of memory functioning.

The Basic Findings and Their Implications

We organize our initial discussion of the findings in terms of four basic questions and the tentative answers that can be drawn from the results.

Do Rememberers Exercise Control Over Grain Size in a Strategic Manner?

Perhaps the most basic finding of the experiments reported in this article is that when people report information from memory, they use control over grain size in a calculated manner, attempting to strike a balance between the goal of being accurate and the goal of being informative. Indeed, in choosing a grain size for their answers (Phase 2 of each experiment), participants did not blindly stick to the more informative, fine-grained answers or to the generally more accurate, coarse-grained answers. Rather, although they presumably preferred to provide the fine-grained answers, they were willing to do so only when these answers were (deemed to be) sufficiently likely to be correct. Moreover, when the incentives for informativeness versus accuracy were manipulated (Experiment 3), the participants adjusted their grain choices accordingly, lowering their confidence criterion and providing more fine-grained answers when informativeness was rewarded more heavily.

How Effective Are Rememberers in Choosing a Grain Size for Their Answers?

The results of all three experiments indicated that the participants were at least moderately successful in choosing a grain size that would enhance their accuracy with minimal loss of informativeness. First, when participants opted to sacrifice informativeness by providing the coarse-grained answer, the average increase in accuracy that was achieved for those items was relatively large (42 percentage points across the three experiments, as compared with a potential increase of 32 percentage points for the items that had instead been answered at the fine grain size). Second, the results of Experiment 3 showed that the monetary gain resulting from participants' grain choices was substantially greater than what would be achieved through an arbitrary control policy (e.g., providing all fine answers, all coarse answers, or a random mix of the two).

Performance was far from optimal, however. Optimal performance entails that the fine-grained answer be provided whenever it is correct and that the coarse-grained answer be provided otherwise. Evaluated in this manner, all of the fine-grained answers that were correct in Phase 1 of the experiment should have been volunteered in Phase 2, but none of the fine-grained answers that were wrong in Phase 1 should have been volunteered.[12] In fact, the actual percentages across the three experiments were 55% and 25%, respectively. This level of performance, although better than chance, presumably reflects a less-than-optimal monitoring process, control process, or both. Experiments 2 and 3 provided data about each of these aspects.

How Effective Are Rememberers in Monitoring the Correctness of Their Answers at Different Grain Sizes?

Effective control of memory grain size requires that people be able to monitor the correctness of their candidate answers. In terms of monitoring resolution, the results of Experiments 2 and 3 indicate that the participants were moderately successful in discriminating between correct and incorrect answers at each grain size, within-subject gamma correlations averaging .47 and .48 for the fine-grained and coarse-grained answers, respectively. These correlations, however, were lower than those sometimes obtained in memory research (e.g., .87 for the general-knowledge recall test in Koriat & Goldsmith, 1996c, Experiment 1).

In terms of calibration, the participants were generally overconfident about answers that they believed were likely to be correct,

[12] Actually, it is not clear which is the proper choice of grain size when both the fine-grained and the coarse-grained candidate answers are wrong. In line with Yaniv and Foster's (1997) proposal that people expect choice of grain size to reflect precision of knowledge, we presume that providing the coarse-grained answer would be the more appropriate choice when one lacks knowledge. Indeed, in examining the participants' actual grain choices across the three experiments, we found that they provided the coarse-grained answer in 86% of the cases (n = 874) in which both alternative answers were wrong and in fully 100% of the cases (n = 180) in which confidence in both answers was zero.

and this was particularly so for the fine-grained answers. Thus, the levels of accuracy actually achieved in the participants' grain choices presumably fell short of the intended levels. In Experiment 2, for instance, the mean assessed probability correct for answers provided at the fine grain size was .77, but in fact only 54% of these answers were correct. A similar finding was reported by Yaniv and Foster (1997, Study 2), who had participants provide 95% confidence intervals for estimated quantitative values (e.g., the current population of the United States). They found that only 43% of these intervals actually contained the true value. In fact, to reach the targeted hit rate (95%), the provided intervals would need to be widened, on average, by a factor of 17!

Thus, it may be that people do not adequately take the interval widths into account in forming subjective probability estimates. This conclusion was also suggested by the calibration plots of Experiments 2 and 3 (see Figures 1 and 2), in which the actual proportion correct for the coarse-grained answers was consistently higher (by an average of .26) than the corresponding proportion correct for the fine-grained answers in each assessed-probability-correct category.

Why should a coarse-grained answer be more likely to be correct than a fine-grained answer when both are assigned the same assessed probability correct? As suggested earlier, it is possible that participants do not sufficiently adjust their probability assessments to accommodate the differences in the a priori probabilities of the answers at the different grain widths (cf. Tversky & Kahneman, 1974). Thus, believing that they lack knowledge regarding the question, they might tend to underrate the possibility that a coarse-grained answer is nevertheless correct (yielding the pronounced underconfidence observed for the low-confidence answers at the coarse grain size), and, conversely, believing that they possess knowledge regarding the question, they might tend to underrate the possibility that a fine-grained answer is nevertheless wrong (yielding the pronounced overconfidence observed for the high-confidence answers at the fine grain size). Such a deficiency would be expected to hinder the effectiveness of the grain control process, biasing it toward the choice of fine-grained answers.

In sum, although participants' monitoring was sufficiently diagnostic of the correctness of the answers to be useful in the control of grain size, it was still quite fallible. It should be interesting to examine whether there are fundamental differences between the monitoring processes used to evaluate quantitative-interval answers and those used to evaluate the type of discrete, nominal-scale answers that are more commonly elicited in memory research. In the latter case, for instance, source monitoring processes (Johnson, Hashtroudi, & Lindsay, 1993) and the accessibility and cue-familiarity heuristics (Koriat & Levy-Sadot, 2001) might make larger contributions.

What Is the Nature of the Control Process Underlying the Choice of Grain Size?

Experiments 2 and 3 yielded some interesting results regarding the manner in which the participants apparently decided on an appropriate grain size for their answers. Two basic models were compared, the satisficing model and the relative expected-utility model. The satisficing model assumes that, because of its greater informativeness, the fine-grained answer is treated as the default response; the only consideration is whether this response is perceived as sufficiently likely to be correct. In contrast, the relative-utility model holds that the assessed probability correct and perceived informativeness of the candidate answers at both grain sizes should jointly play a role. The tendency to volunteer the fine-grained answer should increase to the extent that its assessed probability is relatively close to that of the coarse-grained answer and to the extent that its perceived informativeness is relatively high in comparison with that of the coarse-grained answer.

On the whole, the results strongly favored the satisficing model. Confidence in the fine-grained answer was by far the most diagnostic predictor of the choice of grain size, with increased confidence in the fine-grained answer increasing the likelihood that that answer would be provided. Although confidence in the coarse-grained answer also appeared to contribute, its contribution was quite small and, surprisingly, was in a direction opposite to what was predicted by the relative-utility model. That is, rather than increasing the likelihood that the coarse answer would be volunteered, greater confidence in the coarse-grained answer actually increased the likelihood that the fine-grained answer would be volunteered (holding confidence in the fine-grained answer constant). This finding was obtained not only in Experiment 2, which relied on implicit incentives for informativeness and accuracy, but also in Experiment 3, which provided explicit monetary incentives. Clearly, such findings count against the idea that participants base the grain decision on a comparison of the expected subjective utility of the fine-grained and coarse-grained answers. However, they also suggest the need for a slight modification of the satisficing model. For instance, it may be that one's confidence in the fine-grained answer is reinforced by confidence in the coarse-grained answer so that choice of grain size also depends on a more global feeling of confidence.

With regard to the role of informativeness, here too the results suggest the need for a refinement of the satisficing model. In Experiment 2, participants were more likely to provide a fine-grained answer when its informativeness was relatively high in comparison with the coarse-grained alternative (as indexed by the difference in the natural logs of the interval widths of the two answers; see Yaniv & Foster, 1995). Similarly, in Experiment 3, participants were more likely to provide a fine-grained answer when the reward for providing a correct fine-grained answer was relatively large in comparison with the reward for providing the coarse-grained answer. Thus, it would appear that the confidence criterion for providing the fine-grained answer is sensitive to the relative utility of the fine-grained and coarse-grained answers. This could be incorporated into the satisficing model by assuming that the criterion is adjusted up or down for each question, depending on the relative payoff for informativeness.

Although clearly these conclusions should be treated with some caution (see further discussion to follow), it is interesting to speculate why people might opt to use a satisficing strategy rather than a relative-utility strategy in controlling the grain size of their answers. Simon (1956, 1990) and others (e.g., Gigerenzer & Todd, 1999) have pointed to two basic advantages of a satisficing strategy. First, it generally requires fewer mental resources than a more systematic optimization strategy. In the present case, the satisficing model requires only that one first assess the probability that the fine-grained answer is correct and then compare that assessment with a preset criterion level (which must be arrived at in some manner). In contrast, the relative-utility model requires assessment of the probabilities of both the fine-grained and the coarse-grained answers, assessment of the utility of providing a correct answer at each grain size (e.g., in terms of perceived informativeness), calculation of the expected subjective utility of each answer in terms of the product of the two preceding assessments, and, finally, comparison of the expected utility values.

A second advantage of the satisficing strategy is that it does not require that one consider and evaluate each of a potentially infinite number of potential responses before choosing the one that is best. This would seem to be of dubious value in the present study, in which participants were limited to a choice between two grain sizes that were dictated by the experimenter. In real-life contexts, however, people have a much wider range of potential grain sizes available to them. Thus, although it is interesting that the behavior of participants who were confined to choosing between two grain sizes conformed to a satisficing strategy, such a strategy should be even more likely to operate in unconstrained real-life memory reporting.

Control of Grain Size in Real-Life Memory Reporting: Some Qualitative Observations

In this section, we consider some aspects of the control of grain size in real-life memory reporting that were not addressed in the present study. The desire to bring real-world memory phenomena into the laboratory for controlled experimental investigation is often at odds with the desire to capture the full richness of the phenomena as they occur in their natural contexts (Banaji & Crowder, 1989; Gruneberg & Morris, 1992). In the present study, we tried to achieve an expedient compromise that would offer us the benefits of experimental tools and rigor while still tapping some of the fundamental features of the control of grain size in real-world settings. Clearly, however, there are features of real-life control of grain size that are neglected within our rather artificial experimental paradigm.

To gain some insights about how control over grain size might manifest itself in a somewhat less constrained memory situation, we selected 10 of the 40 items from the general-knowledge test used in this study and asked 5 new participants to answer them in whatever way they saw fit, “in accordance with the extent of your knowledge.” They were asked to “think out loud” while answering the questions, which allowed us to make some general observations about their answering processes. In particular, we were interested in seeing to what extent the participants’ protocols might resemble the type of satisficing model suggested by our experimental findings.

One basic observation from these protocols was that the participants often seemed to attempt to reconstruct the answer on the basis of remembered pieces of information and plausible inferences. For instance, when asked how old President John F. Kennedy was at the time of his assassination, a participant might begin “I know he was very young, but I think U.S. presidents are required by law to be at least 35. I would say he was a bit over 40, let’s say, between 40 and 45.” In some cases, the participant would initially home in on a tentative point estimate (e.g., “I would say he was a bit over 40”) and then widen that estimate to be more certain (“let’s say, between 40 and 45”). This is the type of interval-widening process that we had in mind as a more realistic instantiation of the satisficing model presented earlier.

In other cases, however, the control process appeared to operate in the opposite direction: The participant would begin with an initial broad estimate that was considered almost certain to be correct (e.g., “I know he was young”), and this estimate would then be narrowed, sometimes in several steps, to the answer that was ultimately provided (e.g., “between 40 and 45”). On the face of it, this type of grain-contraction process would seem to be quite different from the grain-expansion process just described. On closer inspection, though, it can be seen that both processes implement the same heuristic: Provide the most precise answer that is possible while maintaining a satisfactory level of confidence.[13] By contrast, there was no indication at all in the protocols of a process resembling a relative-utility strategy, in which participants would be expected to weigh the relative benefits of several possible answers before choosing the one that they thought was best.

Several additional observations from the protocols can help illustrate further aspects of control over grain size in real-life memory reporting. First, the participants tended to choose answers (and intermediate estimates) that were expressed in relatively natural or “round” quantitative units, such as decades or 5-year spans, 10s of meters, and so forth. Thus, although in principle the range of possible grain sizes for an answer is potentially infinite, in practice people tend to restrict themselves to a much smaller number of candidate grain sizes. Second, the participants often made use of quantitative approximatives and other types of vague linguistic expressions, for example, “around 45 years old” or “in the early 1900s.” Vague linguistic expressions have both advantages and disadvantages in conveying quantitative information (e.g., Erev, Wallsten, & Neal, 1991; Moxey & Sanford, 1993; Wallsten, 1990; Wierzbicka, 1986), and the processes underlying their use in memory reporting will need to be addressed in future work. Third, in no case did a participant offer or even consider an overly precise answer (e.g., President Kennedy’s age in months, days, or seconds at the time of his death). Thus, the idea that one tries to provide as informative an answer as possible is, at the extreme, overly simplistic. It is more likely that one tries to provide the most informative answer that is appropriate in a particular situation (Grice, 1975).

A final observation concerns the relationship between control over grain size and report option. Participants often began with a statement such as “I really have no idea” or “I really don’t know.” Sometimes, they would nevertheless begin a reconstruction-estimation process along the lines just described, ultimately ending in a relatively coarse-grained answer. In other cases, however, “I don’t know” or “I can’t remember” remained as the participant’s preferred response to the question (see Smith & Clark, 1993, for a possible metacognitive distinction between these two responses). It is interesting to note that participants opted for the “don’t know” response even though they surely could have provided a very

[13] The general pattern of responses in the protocols leads us to speculate that the use of these alternative strategies may depend on the quality of the information that is available in memory: The grain-expansion strategy might be preferred when a relatively precise answer “pops” to mind, but lacking confidence in it, one decides to “hedge” one’s bet, whereas the grain-contraction strategy might be more suited to the type of reconstructive memory processing mentioned earlier (further details of the possible interaction between memory and metamemory processes are discussed subsequently).

coarse answer that was likely to be correct. Once again, pragmatic norms of communication probably led participants to censor their answers, in this case those that seemed ridiculously uninformative (cf. the inflation-rate example from Yaniv & Foster, 1995, cited in the introduction). Perhaps, in addition to a confidence threshold, rememberers also set an informativeness threshold that must be passed before an answer will be volunteered at all. A more comprehensive model of the strategic regulation of real-life memory reporting will need to address how control over grain size and control of report option are jointly used.

Comparison With Other Approaches to Memory Grain Size

Clearly, the idea that memory performance can vary in terms of its graininess or precision is not new. What is the relationship between previous work and the approach presented in this article? As an aid to the discussion, Figure 4 depicts a rough scheme for conceptualizing and distinguishing several basic components that underlie overt memory (recall) performance. This conception is essentially an extension of the framework put forward in our earlier work on report option (cf. Koriat & Goldsmith, 1996c, Figure 1).

Consider first the traditional approach to the issue of memory grain size. A large amount of work has shown that people often remember the "gist" of an event, although they do not remember its details. Much of that research has examined gist versus verbatim memory of linguistic–textual information. The basic finding is that the general meaning of studied material is forgotten less rapidly than is more detailed information, such as the surface form or verbatim wording of that material (e.g., Begg & Wickelgren, 1974; Kintsch, Welsch, Schmalhofer, & Zimny, 1990; Reyna & Kiernan, 1994). Kintsch et al. (1990), for example, found differential forgetting rates for three different levels of textual information, with surface information (i.e., verbatim memory) becoming inaccessible within 4 days, memory for the semantic content (i.e., gist) declining at a slower rate, and judgments based on situational memory (i.e., inferences from a relevant knowledge schema) remaining highly stable over time. Other work seems to parallel the gist–verbatim distinction (see Reyna & Brainerd, 1995). For example, memory for category versus instance information has been examined in terms of the tendency to falsely recognize foils taken from the same category as the studied target words (e.g., falsely recognizing *canary* instead of *sparrow;* Dorfman & Mandler, 1994; Koutstaal & Schacter, 1997) or foils that constitute the category name itself (e.g., recognizing *toy* instead of *doll;* Brainerd, Reyna, & Mojardin, 1999). Also, studies of story recall have shown that higher level (thematic or superordinate) propositions are better retained over time than lower level (subordinate) propositions (e.g., Christiaansen, 1980; Kintsch, Kozminsky, Streby, McKoon, & Keenan, 1974). Similarly, in testing memory for university course content, Conway, Cohen, and Stanhope (1991) found little forgetting of general principles and concepts over a 12-year retention period, whereas memory of specific details declined sharply (see also Stanhope, Cohen, & Conway, 1993).

In studying these phenomena, researchers have tended to focus on such issues as how the information is initially encoded, the nature of the underlying memory representations, how these representations might change over time, and how they might be accessed and processed. For instance, schema-based theorists (e.g., Bransford & Franks, 1971; see Alba & Hasher, 1983, for a review) have interpreted such findings as indicating that as a result of abstraction and integration processes, verbatim traces of the original information are either lost or become integrated with schematic-gist information. Subsequent memory performance must then be based on reconstructive processing. In contrast, studies motivated by fuzzy trace theory (Brainerd & Reyna, 1990; Reyna & Brainerd, 1995) postulate that, during encoding, verbatim and gist traces are formed in parallel, creating a hierarchy of independent representations at varying levels of precision (for a similar proposal, see Neisser, 1986). Over time, the more precise, verbatim traces decay at a faster rate than the more general, gist traces. In support of this idea, several studies have shown that

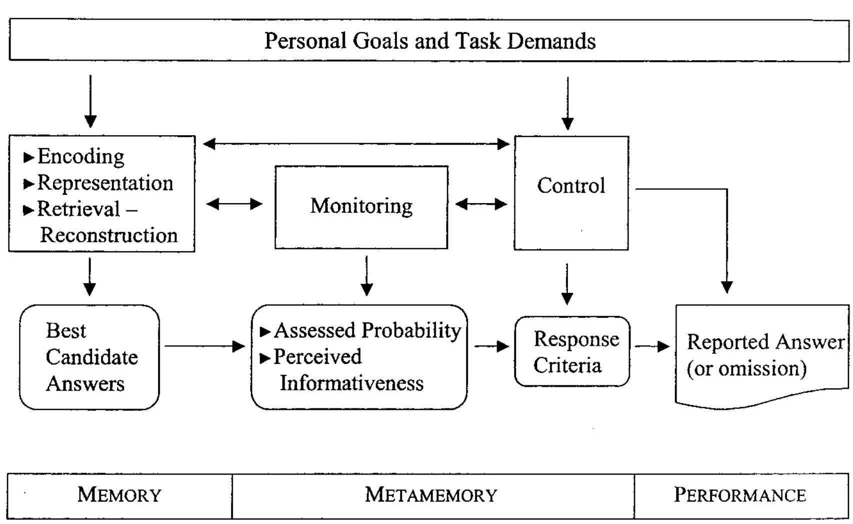

Figure 4. A scheme for conceptualizing and distinguishing cognitive and metacognitive components underlying recall memory performance, focusing on strategic regulation of grain size and report option.

memory for gist and memory for verbatim details are stochastically independent (see Reyna & Brainerd, 1995, for a review). Still other accounts have proposed that differences between verbatim and gist memory may stem from differential amounts of attention paid to different levels (types) of information during encoding (e.g., Gernsbacher, 1985; Murphy & Shapiro, 1994).

In terms of the scheme presented in Figure 4, it can be seen that most of the previous work on memory graininess (gist) has concentrated on elements contained in the left-most box, addressing the types of encoding, representation, and retrieval–reconstruction issues that are generally considered to concern "memory" per se. Our work, in comparison, is based on the assumption that between the retention and retrieval (or reconstruction) of information, on the one hand, and overt memory performance, on the other hand, reside metacognitive monitoring and control processes that are critical in determining the accuracy and informativeness of what one reports. These processes are generally ignored in memory research, perhaps because of the reluctance of researchers to give participants control over their memory performance (see Nelson & Narens, 1994). In fact, many of the experimental procedures used to study gist are based on forced-choice or old–new recognition tests that circumvent these processes. Thus, for instance, memory for gist is commonly indexed by false-recognition rates to nonstudied sentences that share the same semantic content as studied sentences (e.g., Kintsch et al., 1990; Reyna & Kiernan, 1994) or, as mentioned earlier, to nonstudied foils that are semantically related to the studied items (e.g., Brainerd et al., 1999; Dorfman & Mandler, 1994; Koutstaal & Schacter, 1997). Such procedures are well suited to identifying the nature and time course of changes to the memory representation or access to that representation. However, a different approach is needed if our aim is to understand the manner in which rememberers choose a grain size for their responses in relatively unconstrained recall and free-narrative reporting situations.

The present work explicitly focuses on recall rather than recognition memory. In this domain, it attempts to supplement the traditional approach by bringing to the fore and examining additional, metacognitive components of memory performance. Within our framework, the encoding, representation, and retrieval–reconstruction of information at different grain levels contribute the raw materials from which memory reports are ultimately forged, and the quality of this contribution surely has a substantial effect on the quality of the final product. Nevertheless, as our work demonstrates, both the accuracy and the informativeness of what people report from memory also depend on strategic regulatory processes that operate in the service of personal and situational goals. It is important, then, to understand these processes as they intervene in converting remembered information into actual memory responses (for a forerunner of this idea, see *conversion* processes in Tulving, 1983). This expanded view of the components and processes underlying memory performance has important implications. First, it implies that both cognitive and metacognitive factors must be considered in explaining differences in the graininess of people's memories. Second, it suggests that memory deficits, such as the tendency of elderly people to report gistlike memories (e.g., Earles, Kersten, Turner, & McMullen, 1999), may stem from differences in memory processes, differences in metamemory processes, or both. Attempts to remedy such deficits should vary, of course, depending on their presumed source.

Several other lines of work share our emphasis on strategic control and the role of conversion processes (although not necessarily metamemory processes) in mediating memory performance. Prominent among these is fuzzy trace theory (Brainerd & Reyna, 1990; Reyna & Brainerd, 1995), which assumes that rememberers can base their performance on memory representations at different levels of detail. According to the theory, people tend to rely on the coarsest representation that is suitable to meet task demands. Thus, for instance, Reyna and Kiernan (1994) found that changing the task from a verbatim recognition task to a gist recognition task led participants to increase acceptance rates for gist-consistent inferences. Such control is apparently limited, however. For example, in explaining false-memory phenomena, the theory assumes that false-recognition responses stem from participants' tendency to base their recognition judgments on gist, even though verbatim-identity responses are called for (Brainerd & Reyna, 1998; Brainerd et al., 1999; Reyna & Lloyd, 1997). Fuzzy trace theory has proven useful in addressing memory and reasoning performance (and the relation between the two) in a wide variety of domains (see Reyna & Brainerd, 1995). With regard to recall memory, although the theory has been applied to account for output organization in terms of the order of recalled items (Brainerd, Reyna, & Howe, 1990), it has not yet been used to address the choice of grain size in reporting information from memory. If one were to do so, we suggest that it would be necessary to include monitoring and control processes such as those considered here to guide the choice of the appropriate level of memory representation to be accessed and converted into a memory response.

A second line of research, one that also emphasizes the importance of hierarchical memory representations and conversion-control processes in reporting information from memory, is that of Huttenlocher and colleagues (Huttenlocher, Hedges, & Bradburn, 1990; Huttenlocher, Hedges, & Prohaska, 1988) on memory for elapsed time and for the dates of past events. In their model, for instance, "reports from memory depend both on what has been encoded and on an estimation process that produces reports from what has been encoded" (Huttenlocher et al., 1990, p. 196). Moreover, "the length of the temporal interval indicated in the question affects the level of specificity of information about dates which is retrieved" (Huttenlocher et al., 1988, p. 474). Biased reports occur when the interval from which the events are drawn is known by the rememberer to have distinct boundaries at the end (e.g., "sometime last semester"), allowing the selective elimination of dating errors that fall beyond the interval endpoint (see also Rubin & Baddeley, 1989). The focus of this work, however, is on the nature of the underlying memory representation (hierarchically structured and unbiased) rather than on the strategic aspects of memory reporting (for a similar approach to bias in memory reports of spatial location, see Huttenlocher, Hedges, & Duncan, 1991).

It is perhaps not surprising that more attention has been paid to strategic aspects of reporting information in work on judgment and decision making. In that domain, for example, a vast amount of work has focused on the role and validity of confidence judgments and how confidence should be expressed and taken into account in the decision-making process (e.g., Clark, 1990; Juslin et al., 2000; Keren, 1991, 1997; Koriat, Lichtenstein, & Fischhoff, 1980; Lichtenstein et al., 1982; Sniezek, Paese, & Switzer, 1990). Such work, of course, has greatly influenced the decision-theoretic approach put forward in our treatments of report option and control over grain size in memory reporting.

Students of decision making have also begun to address the role of vague (as opposed to precise) information in the decision-making process. In an interesting line of research, Wallsten, Budescu, and collaborators (e.g., Budescu & Wallsten, 1995; Wallsten, 1990; Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986) have examined the use of vague linguistic qualifiers such as "certainly," "probably," and "very likely" to indicate one's confidence (or uncertainty) that a predicted event will occur (see also Clark, 1990; Moxey & Sanford, 1993; Teigen, 1988). They have pointed out that decision makers "feel best served when representations of uncertainty are as precise as possible, but no more precise than warranted" (Wallsten, 1990, p. 35). Interestingly, when properly elicited, linguistic expressions of uncertainty were found to be as accurate as and no less useful (sometimes even more so) than numerical judgments in allowing effective decisions (see Budescu & Wallsten, 1995, for a review).

Finally, perhaps most similar to our approach, both in spirit and in substance, is work by Yaniv and Foster (1995, 1997) on the graininess of judgment under uncertainty. As noted in the introduction, they proposed that the vagueness or graininess of judgmental estimation under uncertainly involves a trade-off between the conflicting objectives of accuracy and informativeness. Because receivers of information prefer estimates that are both precise and accurate, senders of information must try to find a compromise between these two objectives. This proposal was supported by several experiments (Yaniv & Foster, 1995) in which participants (receivers of information) ranked their preferences for quantitative estimates that varied both in precision and in accuracy. The results were consistent with an additive model in which both precision and accuracy contributed to the judged quality of the estimates. In a second study, Yaniv and Foster (1997) found that, across three different methods of eliciting interval-type quantitative estimates, participants (senders of information) provided estimates that maintained relatively low error-to-precision ratios. That is, the precision or coarseness (interval width) of the estimate was highly correlated with the magnitude of the deviation of the true value from the interval midpoint (mean $r: \approx .75$). According to Yaniv and Foster (1997), the precision of a judgment (i.e., interval width) communicates to recipients the size of the error to be expected. Note that their conception of accuracy and error as being graded in magnitude for individual responses differs from the dichotomous (correct–incorrect) scoring used in our study (and in most memory research).

Although their focus and concerns are somewhat different than ours, Yaniv and Foster (1995, 1997) also have promoted the idea that the grain size of reported information reflects both a person's knowledge and the operation of strategic processes used to achieve an expedient compromise between accuracy and informativeness. In their words,

> The absolute error of a judgment reflects the judge's knowledge and is not subject to strategic behavior. In contrast, the error-to-precision ratio reflects not only knowledge but also strategic behavior. For instance, obtaining a relatively high average error-to-precision ratio in a study might suggest that greater importance was placed on informativeness. (Yaniv & Foster, 1997, p. 31)

On a substantive level, it might be interesting to see whether, in the control of memory grain size, rememberers are guided not only by the goal of answering correctly (and precisely) but also by other goals, such as communicating to their audience the order of magnitude of the error that might be expected (see also Wierzbicka, 1986).

In conclusion, although there has been a great deal of work in memory research on differences in the graininess of remembered information, that work has centered mostly on issues concerning underlying memory representations and how these representations might change over time. In the present article, we examined the grain size of memory reports from an expanded functional perspective that takes into account not only the raw memory materials but also the manner in which rememberers control their memory reporting in accordance with personal and social goals. Consideration of functional goals and how metacognitive processes are used to achieve these goals is particularly important in the type of open-ended reporting that is typical of real-life recollection, in which rememberers have great flexibility in deciding what to report and how to report it.

In our previous work (Koriat & Goldsmith, 1994, 1996c), we confined ourselves to an examination of the control of report option, that is, the decision as to whether or not to report specific pieces of information that come to mind, and showed how this decision is guided by the generally conflicting goals of accuracy and completeness. In the present study, we extended our investigation of strategic control to regulation of the grain size of the information that is reported and examined how this regulation is guided by the generally conflicting goals of accuracy and informativeness. Presumably, other goals, such as being entertaining or impressive, and considerations regarding the relevance and importance of the information also guide memory reporting (Neisser, 1996). More work is needed to complete our understanding of the various ways in which strategic control of memory reporting is used in real-life remembering.

References

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93,* 203–231.

Banaji, M. R., & 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.

Barnes, A. E., Nelson, T. O., Dunlosky, J., Mazzoni, G., & Narens, L. (1999). 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, pp. 289–313). Cambridge, MA: MIT Press.

Begg, I., & Wickelgren, W. A. (1974). A retention function for syntactic and lexical vs. semantic information in pretence recognition memory. *Memory & Cognition, 2,* 353–359.

Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review, 10,* 3–47.

Brainerd, C. J., & Reyna, V. F. (1998). Fuzzy-trace theory and children's false memories. *Journal of Experimental Child Psychology, 71,* 81–129.

Brainerd, C. J., Reyna, V. F., & Howe, M. L. (1990). Cognitive triage in children's memories: Optimal retrieval or effortful processing? *Journal of Experimental Child Psychology, 49,* 428–447.

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review, 106,* 160–179.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology, 2,* 331–350.

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.

Budescu, D. V., & Wallsten, T. S. (1995). Processing linguistic probabilities: General principles and empirical evidence. *Psychology of Learning and Motivation, 32,* 275–318.

Christiaansen, R. E. (1980). Prose memory: Forgetting rates for memory codes. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 611–619.

Clark, D. A. (1990). Verbal uncertainty expressions: A review of two decades of research. *Current Psychology: Research and Reviews, 9,* 203–235.

Conway, M. A., Cohen, G., & Stanhope, N. M. (1991). On the very long term retention of knowledge: Twelve years of cognitive psychology. *Journal of Experimental Psychology: General, 120,* 395–409.

Dorfman, J., & Mandler, G. (1994). Implicit and explicit forgetting: When is gist remembered? *Quarterly Journal of Experimental Psychology, 47(A),* 651–672.

Earles, J. L., Kersten, A. W., Turner, J. M., & McMullen, J. (1999). Influences of age, performance, and item relatedness on verbatim and *gist recall* of verb-noun pairs. *Journal of General Psychology, 126,* 97–110.

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

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review, 101,* 519–527.

Erev, I., Wallsten, T. S., & Neal, M. M. (1991). Vagueness, ambiguity, and the cost of mutual understanding. *Psychological Science, 2,* 321–324.

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

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

Fowler, R. D. (Ed.). (1991). [Entire issue]. *American Psychologist, 46*(1).

Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology, 17,* 324–363.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Goldsmith, M., & Koriat, A. (1999). The strategic regulation of memory reporting: Mechanisms and performance consequences. In D. Gopher & A. Koriat (Eds.), *Cognitive regulation of performance: Interaction of theory and application.* (Attention and Performance XVII, pp. 373–400). Cambridge MA: MIT Press.

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

Gruneberg, M. M., & Morris, P. E. (1992). Applying memory research. In M. Gruneberg & P. Morris (Eds.), *Aspects of memory* (2nd ed., Vol. 1, pp. 1–17). London: Routledge.

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

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression.* New York: Wiley.

Huttenlocher, J., Hedges, L. V., & 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., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review, 98,* 352–376.

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

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28.

Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect. *Psychological Review, 107,* 384–396.

Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica, 77,* 217–273.

Keren, G. (1997). On the calibration of probability judgments: Some critical comments and alternative perspectives. *Journal of Behavioral Decision Making, 10,* 269–278.

Kintsch, W., Kozminsky, E., Streby, W. J., McKoon, G., & Keenan, J. M. (1974). Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior, 14,* 196–214.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29,* 133–159.

Klatzky, R. L., & 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., & 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., & 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, & M. Johns (Eds.), *Basic and applied memory research: Practical applications* (Vol. 2; pp. 3–18). Hillsdale, NJ: Erlbaum.

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

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology, 51,* 483–539.

Koriat, A., Goldsmith, M., Schneider, W., & Nakash-Dura, M. (2001). The credibility of children's testimony: Can children control the accuracy of their memory reports? *Journal of Experimental Child Psychology, 79,* 405–437.

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 34–53.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 107–118.

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language, 37,* 555–583.

Leippe, M. (1994). The appraisal of eyewitness testimony. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 385–418). New York: Cambridge University Press.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73,* 13–22.

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, England: Cambridge University Press.

Lipsitz, S. R., Laird, N. M., & Harrington, D. P. (1991). Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association. *Biometrika, 78,* 153–160.

Moxey, L. M., & Sanford, A. J. (1993). *Communicating quantities: A psychological perspective.* Hove, England: Erlbaum.

Murphy, G. L., & Shapiro, A. M. (1994). Forgetting of verbatim information in discourse. *Memory & Cognition, 22,* 85–94.

Neisser, U. (1986). Nested structure in autobiographical memory. In D. C. Rubin (Ed.), *Autobiographical memory* (pp. 71–81). Cambridge, England: Cambridge University Press.

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, England: Wiley.

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

Neisser, U., & Fivush, R. (1994). *The remembered self: Construction and accuracy in the self-narrative.* New York: Cambridge University Press.

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

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy trace theory: An interim synthesis. *Learning and Individual Differences, 7,* 1–75.

Reyna, V. F., & Kiernan, B. (1994). The development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content encoding instructions, and retention interval. *Developmental Psychology, 30,* 178–191.

Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Annual Review of Psychology, 9,* 95–124.

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

Ross, M., & Buehler, R. (1994). Creative remembering. In U. Neisser & R. Fivush (Eds.), *The remembered self: Construction and accuracy in the self-narrative* (pp. 205–235). New York: Cambridge University Press.

Rubin, D. C., & Baddeley, A. (1989). Telescoping is not time compression: A model of the dating of autobiographical events. *Memory & Cognition, 17,* 653–661.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics, 2,* 437–442.

Simon, H. A. (1956). Rational choice and the structure of environments. *Psychological Review, 63,* 129–138.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41,* 1–19.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language, 32,* 25–38.

Sniezek, J. A., Paese, P. W., & Switzer, F. S. C. (1990). The effects of choosing on confidence in choice. *Organizational Behavior and Human Decision Processes, 46,* 264–282.

Stanhope, N., Cohen, G., & Conway, M. (1993). Very long term retention of a novel. *Applied Cognitive Psychology, 7,* 239–256.

Teigen, K. H. (1988). The language of uncertainty. *Acta Psychologica, 68,* 27–38.

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

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Wallsten, T. S. (1990). The costs and benefits of vague information. In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn* (pp. 28–43). Chicago: University of Chicago Press.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. H. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General, 115,* 348–365.

Wierzbicka, A. (1986). Precision in vagueness: The semantics of English "approximatives." *Journal of Pragmatics, 10,* 597–614.

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

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

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

Yzerbyt, V. Y., Lories, G., & Dardenne, B. (Eds.). (1998). *Metacognition: Cognitive and social dimensions.* Thousand Oaks, CA: Sage.

Received February 8, 2001
Revision received July 3, 2001
Accepted July 3, 2001 ■