Journal of Experimental Psychology:
Learning, Memory, and Cognition
2008, Vol. 34, No. 4, 945–959

0278-7393/08/$12.00 DOI: 10.1037/0278-7393.34.4.945

Subjective Confidence in One's Answers: The Consensuality Principle

Asher Koriat
University of Haifa

In answering general-information questions, a within-person confidence–accuracy (C-A) correlation is typically observed, suggesting that people can monitor the correctness of their knowledge. However, because the correct answer is generally the consensual answer—the one endorsed by most participants—confidence judgment may actually monitor the consensuality of the answer rather than its correctness. Indeed, the C-A correlation was positive for items with a consensually correct answer but negative for items with a consensually wrong answer. Results suggest that the consensuality–confidence correlation may be mediated by 2 internal mnemonic cues that are correlated with consensuality: Consensual answers are reached faster and are selected more consistently by the same person on different occasions than nonconsensual answers. The results argue against a direct-access view of confidence judgments and suggest that such judgments will be accurate only as long as people's responses are by and large correct across the sampled items, thus stressing the criticality of a representative design.

Keywords: subjective confidence, metacognition, probabilistic mental models, consensuality, confidence–accuracy relationship

There has been increased interest in recent years in identifying the basis of metacognitive judgments and their accuracy. Some of the work has focused on judgments of learning that are made during the study of new materials (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Benjamin & Bjork, 1996; Koriat, 1997; Nelson, Narens, & Dunlosky, 2004) and on feeling of knowing (FOK) judgments made during the retrieval of information from memory (Koriat & Levy-Sadot, 2001; Metcalfe, Schwartz, & Joaquim, 1993; Reder & Ritter, 1992; Reder & Schunn, 1996). It has been proposed that although such judgments may rely on analytic inferences that draw on one's beliefs about learning and remembering, they can be based also on the online monitoring of various mnemonic cues that have some validity as diagnostics of actual memory performance. Thus, unlike direct-access views in which metacognitive judgments are assumed to be based on detecting the presence and/or the strength of memory traces (e.g., Cohen, Sandler, & Keglevich, 1991; Hart, 1965), the commonly held view in recent years is that such judgments are inferential in nature. Hence, their accuracy is not guaranteed, as would follow from the direct-access view (see Koriat, 2007).

This study focused on the subjective confidence in one's own answers and responses. Assessments of confidence have been investigated over many years in a wide range of domains, including perception and psychophysics (Baranski & Petrusic, 1994), memory (Busey, Tunnicliff, Loftus, & Loftus, 2000; Chandler, 1994; Kelley & Lindsay, 1993; Koriat & Goldsmith, 1996; Roediger & McDermott, 1995), decision making and choice (Klayman, Soll, & Juslin, 2006; Lichtenstein, Fischhoff, & Phillips, 1982), and eyewitness testimony (Bothwell, Deffenbacher, & Brigham, 1987; Read, Lindsay, & Nicholls, 1998; Sporer, Penrod, Read, & Cutler, 1995; Wells & Murray, 1984). In all of these areas, confidence measures have been also relied upon in testing various theories. In recent years, however, there has been increasing interest in the investigation of confidence judgments in their own right—in exploring their bases and the factors that affect their accuracy and inaccuracy (for reviews, see Dougherty, 2001; Erev, Wallsten, & Budescu, 1994; Juslin & Olsson, 1997; Koriat, 2007; Van Zandt, 2000).

The present study focused on confidence judgments in two-alternative general-information questions. Some of the theories on the bases of such judgments emphasize information-driven processes, assuming that confidence in the correctness of an answer is based on the content of domain-specific information retrieved from long-term memory. For example, it has been proposed that confidence in an answer depends on the amount and strength of evidence retrieved in support of that answer relative to the other answer (e.g., Griffin & Tversky, 1992; Koriat, Lichtenstein, & Fischhoff, 1980; McKenzie, 1997; Yates, Lee, Sieck, Choi, & Price, 2002). A prominent theory that stresses the contribution of information stored in long-term memory is the theory of probabilistic mental models (PMM) proposed by Gigerenzer, Hoffrage, and Kleinbölting (1991). According to this theory, both choice and confidence are based on the cues that are used to infer the answer. The correspondence between confidence and accuracy depends on the extent to which the validities of the cues that are utilized correspond to their actual, ecological validities (see below).

Other discussions, in contrast, stress the contribution to confidence judgments of the metacognitive experiences that accompany choice and retrieval, such as the ease with which the answer is retrieved or selected (Nelson & Narens, 1990; Zakay & Tuvia,

The preparation of this article was supported in part by European Commission Grant FP6-NEST: EYEWITMEM; 43460.

I am grateful to Rinat Gil for her help in conducting and analyzing the experiment and to Limor Sheffer for her advice in the statistical analyses. I am also grateful to Ulrich Hoffrage for his comments on a draft of the manuscript.

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

1998). Indeed, participants express stronger confidence in the answers that they retrieve more quickly, whether those answers are correct or not (e.g., Kelley & Lindsay, 1993; Koriat, Ma'ayan, & Nussinson, 2006; Robinson, Johnson, & Herndon, 1997). Kelley and Lindsay (1993) found that priming participants with answers to general-information questions increased the speed with which the primed answers were recalled and, in parallel, enhanced confidence in those answers. However, by and large, response latency tends to be shorter for correct than for wrong answers, so that confidence judgments that are based on response latency are likely to monitor the correctness of the answer (Barnes, Nelson, Dunlosky, Mazzoni, & Narens, 1999; Costermans, Lories, & Ansay, 1992).

What is common to the theories of confidence that have just been mentioned is that they endorse an inferential, cue-utilization view according to which confidence judgments are based on inferences from cues, whether these cues constitute knowledge stored in memory or reside in the feedback from task performance, such as the ease with which a choice of an answer is reached. In this study, I examined a phenomenon that seems to be consistent with both types of theories and to have a certain degree of generality.

Subjective Confidence in a Phonetic-Symbolism Task

The study described in this article was motivated by an intriguing phenomenon that I reported many years ago (Koriat, 1975): Participants were successful in guessing the meaning of words from several languages, some of which they had never heard of, and in addition, they were successful in monitoring their performance, endorsing with greater confidence correct responses than wrong responses. The task that I used was based on the idea of a universal phonetic symbolism: the existence of sound–meaning associations that are universally shared and that have been incorporated in the words of natural languages (Brown, 1958; Sapir, 1929; Taylor & Taylor, 1965). To test this idea, participants were presented with antonymic words from noncognate languages (e.g., *chou–mei*) and were asked to match each of these words with their English equivalents (*beautiful–ugly*). Chances of getting the translation correct are 50%, but participants have been found to perform significantly better than chance (e.g., Slobin, 1968).

I noticed that participants felt greater confidence in some of their word-matching decisions than in others. I was curious to know whether these subjective feelings are diagnostic of the accuracy of the matching because some philosophers assume that universally shared notions have the quality of self-evidence: They strike one as being right. To examine this question, in addition to the requirement to match the members of each foreign pair with the members of the corresponding English pair, participants were asked to indicate their FOK on a 4-point scale, with 1 indicating *a totally wild guess* and 4 indicating that the answer is *reasonably likely to be right.*

Overall, the percentage of correct translation was 58.1%, which was significantly higher than the 50% chance level. In addition, however, the percentage of correct translation increased monotonously with FOK ratings, from 54% for a rating of 1 to 66% for a rating of 4, and the increase was highly significant.

These results present a puzzle: How can people monitor their success in guessing the meaning of words from a language, such as Yoruba or Kanarese, which they have never heard of? Certainly, the direct-access view does not seem to offer an answer, but it is also unclear what cues could participants use that would allow them to discriminate between correct and wrong guesses.

Relating Metaknowledge Accuracy to Knowledge Accuracy

Toward a solution of this puzzle, Koriat (1976) noted an artifact (which later became a corollary of a theory; Koriat, 1993) in the 1975 study. That study, in fact, yielded two observations that at first sight appear unrelated. First, participants' knowledge was accurate: Participants were able to guess the meaning of foreign words with better than chance success. Second, their metaknowledge was accurate: They were able to monitor the correctness of their guesses. The joint occurrence of both observations raises a methodological problem. Specifically, knowledge accuracy implies that the correct translation tends to be the one that is more compelling, that is, one that is consensually endorsed. When we focus on metaknowledge accuracy, we face the problem that the correctness of the translation and its consensuality (the percentage of participants endorsing it) are correlated. Thus, it is unclear whether FOK ratings are correlated with the correctness of the translation or with some aspect of that translation that makes it attractive to most participants.

To examine this question, Koriat (1976) dissociated correctness from consensuality by including in a new study many items for which participants are likely to agree on the wrong translation. A list of items was compiled from several word-matching studies with a heavy representation of items that are consensually wrong. The same procedure as in Koriat (1975) was used. On the basis of the translation results, the items were classified ad hoc into three classes according to whether the majority of participants significantly agreed on the correct translation, significantly agreed on the wrong translation, or did not agree on either. These classes were labeled consensually correct (CC), consensually wrong (CW), and nonconsensual (NC), respectively. The results clearly indicated that FOK ratings correlated with the consensuality of the translation rather than with its correctness: For the CC class, correct answers were indeed associated with stronger FOK ratings, as in Koriat (1975). However, for the CW class, it was the wrong answers that were associated with the stronger ratings. For the NC class, FOK ratings were unrelated to the correctness of the match.

The conclusion from these results is that when a representative sample of items is used, as in Koriat (1975), participants can monitor the correctness of their responses. However, as the results of Koriat (1976) indicate, they do so indirectly, by relying on some cues that are correlated with accuracy. These cues would seem to underlie the consensuality of the response—the extent to which it tends to be endorsed by the majority of people. These results also demonstrate the intimate link between metaknowledge accuracy and knowledge accuracy: Metaknowledge is accurate as long as knowledge itself is accurate.

The idea that metaknowledge accuracy derives from knowledge accuracy also underlies Koriat's (1993) accessibility model of the FOK that is experienced during retrieval. When people fail to retrieve a solicited target from memory, they can predict with some degree of accuracy whether they will be able to retrieve the elusive target at some later time or recognize it among distractors. In contrast to trace-access theories, which assume that FOK judgments are based on access to the underlying memory traces (e.g., Hart, 1965), Koriat proposed that these judgments are based on the mere accessibility of partial clues about the target—the number of clues retrieved and their ease of access. It was claimed that rememberers have no privileged access to the trace of the elusive target nor can they monitor directly the accuracy of the clues that come to mind. Rather, FOK judgments are affected by the total amount of clues that are accessible about the target regardless of their accuracy.

Why then are FOK judgments accurate by and large? According to the accessibility model, this is because memory itself is accurate in the sense that most of the information that comes spontaneously to mind during retrieval, both full and partial information, is correct (see Koriat & Goldsmith, 1996). Thus, the accuracy of metamemory derives from the accuracy of memory itself. In support of this conclusion are the results of Koriat's (1995) study. The design of that study bears some resemblance to the study of Koriat (1976), but it involved FOK judgments in attempting to recall the answer to general-information questions. Memory pointers (questions calling for a specific target—a name or a word—as a response) were divided into those that elicited primarily correct answers (CC pointers) and those that elicited primarily wrong answers (CW pointers) among participants who provided an answer. FOK ratings among participants who failed to provide an answer were found to correlate positively with subsequent recognition performance only for the CC pointers. For the CW pointers, in contrast, FOK judgments following recall failure were inversely correlated with recognition performance. This was true even for pointers that brought to mind a variety of wrong answers, rather than a single incorrect answer, among participants who provided an answer. These results were taken to suggest that FOK judgments about an elusive memory target are accurate as long as the memory pointer brings to mind partial information that is predominantly correct. Thus, FOK judgments have no privileged access to the correct elusive target. Rather, their success in predicting correct memory performance derives from their reliance on a cue that is generally correlated with correct performance.

The Confidence–Consensuality Relationship for Subjective Confidence

The present study extended investigation to confidence judgments in answering general-information questions. It examined the hypothesis that such judgments are also correlated with the consensuality of the answer rather than with its accuracy. Two-alternative forced-choice general-knowledge questions were used that were likely to differ in the proportion of participants who chose the correct answer. Participants answered each question and assessed the probability that the answer is correct using a 50%–100% scale. On the basis of the distribution of the answers across participants, the items were divided ad hoc into three classes, as in Koriat (1976): CC, CW, and NC. If confidence judgments are correlated with the consensuality of the response rather than with its correctness, we should expect the confidence–accuracy (C-A) relationship to be positive for the CC items but negative for the CW items. The NC items were expected to yield little correlation between confidence and accuracy.

What is the rationale for these predictions? It is proposed that several processes that contribute to confidence judgments may converge in producing a confidence–consensuality (C-C) relationship. To describe these processes, I refer to the distinction commonly drawn in discussions of metacognition between information-based and experience-based metacognitive judgments (see Koriat & Levy-Sadot, 1999; Koriat, Nussinson, Bless, & Shaked, in press; Strack, 1992). Information-based judgments rely on analytic, deliberate inferences in which domain-specific beliefs and knowledge in long-term memory are consulted and weighed to reach an educated judgment. Experience-based judgments, in contrast, are assumed to rely on mnemonic cues that derive online from task performance. These cues give rise directly to a sheer subjective feeling that can then serve as the basis for metacognitive judgments (Benjamin & Bjork, 1996; Kelley & Jacoby, 1996; Koriat, 1997; Schwartz, 1994). For example, the fluency with which information is processed and encoded is assumed to affect judgments of learning during study (Kelley & Rhodes, 2002). The amount of partial clues that come to mind during the search for a memory target, the consistency among these cues, and the ease with which they come to mind are assumed to affect FOK judgments (Koriat, 1993; Schreiber & Nelson, 1998). Finally, the speed with which an answer is retrieved or chosen is assumed to affect confidence in that answer (Kelley & Lindsay, 1993; Robinson et al., 1997).

The distinction between experience-based and information-based metacognitive judgments parallels the general distinction in dual process theories between System 1 and System 2 processes (see Kahneman, 2003; Koriat, Bjork, Sheffer, & Bar, 2004; Stanovich & West, 2000). Previous results suggest that participants adopt an intuitive mode of responding as a default but tend to shift to an analytic, deliberate mode when they realize that their immediate gut feelings have been contaminated by irrelevant factors or when they experience cognitive dysfluency (e.g., Alter, Oppenheimer, Epley, & Eyre, 2007; Gilbert, 2002; Jacoby & Whitehouse, 1989; Strack, 1992). Despite the differences between the two types of processes, however, it is proposed that both should yield a C-C relationship. Consider first information-based confidence judgments. As noted earlier, one of the influential theories of subjective confidence in question answering is the PMM theory (Gigerenzer et al., 1991). It assumes that the choice of an answer and the confidence in that choice are based on knowledge stored in long-term memory about the structure of the natural environment. Each general-information question may be assumed to define a target variable as well as the cues that can be used to infer that variable. For example, a question about which of two cities has a larger number of inhabitants may activate such cues as which of the cities has a famous soccer team, which of the cities is a state capital, and so on. Each such cue has a certain degree of validity, and as long as the PMM underlying choice is adapted to the natural environment, as is generally the case for typical or representative general-knowledge questions, then confidence judgments should be well calibrated. In the case of deceptive or misleading questions, however, cue validities do not correspond to ecological validities, and therefore, such questions generally engender inflated confidence because the activated cues tend to lead to an incorrect response.

Gigerenzer et al. (1991) tested their theory using questions involving the number of inhabitants in two cities. They noted that their choice of such questions was because "this content domain allowed for a precise definition of a reference class in natural environments and for random sampling from this reference class" (Gigerenzer et al., 1991, p. 513). Yet how can the PMM theory be extended to the wide variety of questions that are included in most general-knowledge tests? For such questions, there is no single reference class that cuts across all the questions, and each question possibly activates a different set of cues whose validities are hard to define. It may be proposed, however, that even for such a heterogeneous set of questions, there is one general prediction that follows from the PMM theory: Confidence judgments should correlate with the consensuality rather than with the correctness of the answers. Thus, assuming, as postulated by Gigerenzer et al., that the choice of an answer and the confidence associated with it are both determined by cue validities, then the choices consensually made by most participants can be assumed to reflect the perceived validities of the activated cues rather than their actual, ecological validities. Hence, the confidence associated with an item should increase with the proportion of participants who choose that answer regardless of whether the answer is correct or wrong.

Indeed, results consistent with this prediction have been reported recently by Brewer and his associates (Brewer & Sampaio, 2006; Brewer, Sampaio, & Barlow, 2005) on the memory for deceptive and nondeceptive sentences. Brewer et al. (2005) had participants study a list of sentences and then recall the sentences and rate their confidence in their answers. Some of the sentences (deceptive) were designed to induce schema-based inference (e.g., recalling "The absentminded professor forgot his car keys" instead of "The absentminded professor didn't have his car keys"). A strong C-A relationship was found for the nondeceptive items but a much reduced relationship for the deceptive items. It was proposed that confidence is based on the heuristic that complete recalls are accurate recalls, and because participants are not aware of making an error in the case of deceptive sentences, they use the completeness belief to assign unduly high confidence to recall errors. A subsequent study (Brewer & Sampaio, 2006) using old/new recognition testing yielded even more striking results: The C-A relationship was reversed for deceptive items. Recognition rate averaged .71 for the nondeceptive items and .31 for the deceptive items. In parallel, confidence (on a 7-point scale) was higher for correct (5.4) than for wrong responses (4.8) for the nondeceptive items, but for the deceptive items. it was actually higher for the wrong responses (5.8) than for the correct responses (5.0). This pattern conforms to the consensuality principle in that confidence ratings were higher for consensually endorsed responses regardless of their accuracy.

Thus, the first aim of this study was to test the consensuality hypothesis with regard to confidence judgments in forced-choice general-information questions. It was predicted that confidence should correlate with the consensuality of the response, so that these judgments would be accurate only for the CC items. The CW items, in contrast, should yield a negative C-A correlation.

Although Gigerenzer et al. (1991) argued that their theory "deals with spontaneous confidence—that is, with an immediate reaction, not the product of long reflection" (p. 507), other researchers assume that immediate, intuitive judgments are experience based rather than information based (Koriat, 1998). They are assumed to rely on mnemonic cues that derive from the process of making a decision rather than on domain-specific knowledge (see Koriat et al., in press). I hypothesize, however, that experience-based confidence judgments too may yield a C-C correlation. Thus, the second aim of this study was to test this hypothesis with regard to two mnemonic cues that derive online from task performance: decision latency and self-consistency.

Consider decision latency first. As mentioned earlier, it has been proposed that one of the mnemonic cues that contribute to immediate subjective confidence is the ease with which an answer is retrieved or chosen (Costermans et al., 1992; Kelley & Lindsay, 1993; Koriat et al., 2006; Nelson & Narens, 1990; Robinson et al., 1997). Indeed, several previous results suggest that (a) decision latency influences confidence judgments (cue utilization) and (b) decision latency is diagnostic of the correctness of the answer (ecological validity; Koriat et al., 2006; Robinson et al., 1997). These observations suggest that reliance on response latency as a cue for confidence contributes to the C-A correlation. I propose, however, that such is the case only for typical or representative questions for which the correct answer is the one that is more likely to be chosen. For deceptive questions, in contrast, it is the wrong answer that may be associated with shorter response latency. However, participants apply the same heuristic across all items, a heuristic that has probably been adapted to representative items: Easily retrieved or chosen answers are more likely to be correct than answers that are associated with longer latencies. Therefore, confidence judgments that are based on response latency should also yield a C-C correlation. This idea was tested by showing that response latency is correlated with the consensuality of the response rather than with its correctness and that participants appear to apply the same heuristic indiscriminately to all items: Faster choices are more likely to be correct than slower choices.

A second internal cue that may also yield a C-C correlation is cue consistency. Several theories assume that in attempting to answer a forced-choice question, participants retrieve different considerations and that their choice depends on the balance of evidence in favor of each of the answers. Subjective confidence in the chosen answer depends then on the extent to which the retrieved considerations consistently favor that answer (Juslin & Olsson, 1997; Saito, 1998; Slovic, 1966). Assuming that participants draw on a similar set of considerations in choosing the answer for a given item, then the between-participant variation in the choice made for that item should be diagnostic of the degree of within-person vacillations and conflict that occur online as a person retrieves different considerations. Hence, reliance on cue consistency as a basis of one's confidence in a choice may also be expected to yield a C-C correlation.

To test this hypothesis, I had participants answer the same questions twice. The assumption was that self-consistency in the choice made across the two encounters can serve as an index of the degree of self-consistency experienced in making a choice in each encounter (Koriat, 1976; Slovic, 1966). Therefore, degree of self-consistency across different presentations of the question should correlate with cross-participant consensus and should yield the same pattern of relationship to confidence as does cross-participant consensus. Thus, for CC items, answers that are repeated across the two presentations should tend to be more correct than nonrepeated answers. For the CW items, in contrast, it is the nonrepeated answers that have a higher chance to be correct.

Experiment

Method

Participants. Forty-one Hebrew-speaking Haifa University psychology undergraduates (33 females and 8 males) participated in the experiment for course credit.

Stimulus materials. A 110-item general-knowledge test (in Hebrew) was developed, with questions covering a broad range of topics; 105 questions were used for the experiment phase, and the other 5 were used for the preceding practice phase. Most of these questions were taken from Koriat's (1995) study of the FOK. Each question was followed by two alternative one- or two-word answers, either a concept or a name of a person or a place.

In compiling these questions, a deliberate attempt was made to include questions that differ widely in the likelihood of evoking correct and incorrect answers. The questions were compiled from many sources, taking advantage of some of the deceptive items reported in the literature (e.g., Fischhoff, Slovic, & Lichtenstein, 1977; Gruneberg, Smith, & Winfrow, 1973; Nelson, Gerler, & Narens, 1984). In addition, several pretests were carried out to assure a sufficiently large number of deceptive items. Examples of the questions can be found in Koriat (1995).

Apparatus and procedure. The experiment was conducted on an IBM-compatible personal computer. Each question appeared on the screen and remained until the participant pressed the space bar to indicate that he or she had finished reading it. Immediately after that, two alternative answers, labeled *a* and *b,* were presented beneath the question, and the participant indicated his or her answer by pressing either a left or a right key on the keyboard. Decision time, defined as the interval between the space-bar press and the choice of the answer, was measured. After the choice of an answer, the statement "confidence (50%–100%)" appeared on the screen. Participants were required to type in a number at that range, which expressed their confidence in the correctness of the answer. The instructions before the experiment indicated that 50% represented a chance level and that participants should make an effort to assess the likelihood that their answer was correct, using the full range between 50% and 100%.

After a short break, the entire procedure was repeated. The order of presentation of the questions was determined randomly for each participant and presentation. The order of the alternative answers was counterbalanced across participants.

Results

As expected, the percentage of correct answers differed extensively between items, ranging from 9.8% to 100%. The items were divided on a post hoc basis: A choice of an answer by 28 or more of the 41 participants is significantly different from chance at the .05 level by a binomial test. There were 35 items for which 28 or more of the participants chose the correct answer. These were classified as CC items. For 13 items, 28 or more of the participants chose the wrong answer, and these were classified as CW items. The remaining 57 items were classified as NC items.

Calibration. I first report the results on calibration (or absolute accuracy)—the correspondence between mean confidence and proportion correct (Lichtenstein et al., 1982). Previous discussions have emphasized the possible contribution of deceptive items to the overconfidence bias that is typically observed in the monitoring of one's own knowledge (see Hoffrage, 2004; Juslin, Winman, & Olsson, 2000; Nickerson, 1998). It has been argued that this bias may be due to the biased sampling of items by researchers—the tendency to include too many misleading items (Gigerenzer et al., 1991). Indeed, when items are sampled representatively, the overconfidence bias has been found to decrease or disappear entirely. In this study, too, confidence judgments for the CW items were considerably inflated (69.2%) in comparison with their mean percentage of correct answers (22.9%). In contrast, the CC items yielded an underconfidence bias, with confidence and accuracy averaging 80.6% and 77.8%, respectively, $t(40) = 2.53$, $p < .05$. The respective means for the NC items were 65.0% and 52.6%, $t(40) = 8.72$, $p < .0001$. This pattern of results conforms to the so-called hard–easy effect (Lichtenstein & Fischhoff, 1977): The overconfidence bias decreases as the difficulty of the questions decreases (see, e.g., Gigerenzer et al., 1991; Suantak, Bolger, & Ferrell, 1996) to the extent that easy items tend to produce a certain degree of underconfidence overall (e.g., Griffin & Tversky, 1992; Lichtenstein & Fischhoff, 1977).

Whereas previous studies focused on calibration—the overall overconfidence or underconfidence bias as it may differ for representative and nonrepresentative items (Gigerenzer et al., 1991; Juslin et al., 2000)—this study was concerned with monitoring resolution (Nelson, 1984; or discrimination accuracy, Yaniv, Yates, & Smith, 1991)—the extent to which confidence judgments discriminate between correct and wrong answers. The consensuality principle implies that the inclusion of nonrepresentative, CW items in the experimental list not only hurts calibration but should also impair resolution.

Confidence judgments for correct and wrong answers. I examine first the C-A relationship. Figure 1A presents mean confidence judgments for correct and wrong answers for the three classes of items. The means in this figure are based on 39 participants who had both correct and wrong answers for each class of items (one participant was wrong on all CW items, and one was never wrong on all CC items). The results for the CC and CW classes clearly demonstrate a crossover interaction. A two-way analysis of variance (ANOVA), Item Class (CC vs. CW) × Correctness, yielded $F(1, 38) = 22.94$, $MSE = 51.10$, $p < .0001$, for item class; $F(1, 38) = 16.57$, $MSE = 46.02$, $p < .001$, for correctness; and $F(1, 38) = 48.47$, $MSE = 70.33$, $p < .0001$, for the interaction. For the CC items, confidence was higher for correct than for wrong answers, $t(38) = 11.03$, $p < .0001$. This was true for 37 out of the 39 participants ($p < .0001$, by a binomial test). In contrast, for the CW items, confidence was higher for the wrong items, $t(38) = 2.35$, $p < .05$. This pattern was true for 29 out of the 39 participants ($p < .005$, by a binomial test). Thus, confidence is correlated with the consensuality of the response rather than with its correctness. For the NC items, there was no difference in confidence between correct and wrong answers, $t(38) = 1.14$, $p < .26$.

Note that mean confidence for the less compelling, nonconsensual response was about the same for the CC and CW items (66.5% and 65.5%, respectively) and was similar to that for both responses in the NC class (65.3% for correct responses and 64.0%, for wrong responses). This pattern suggests that the higher overall confidence for the CC class (73.4%) than for the CW class (67.9%) derives from degree of consensus being higher for the CC items than for the CW items. Thus, there were seven CC items (20%) and only

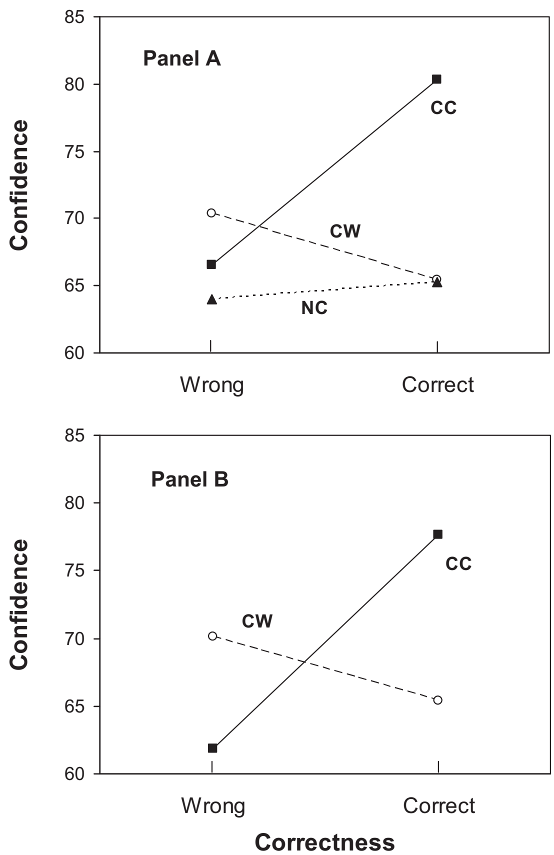

Figure 1. Mean confidence for correct and wrong answers, plotted separately for the consensually correct (CC), consensually wrong (CW), and nonconsensual (NC) items (Panel A). In Panel B, the results are plotted for the 13 CC and 13 CW items that were matched on the proportion of consensual responses.

one CW item (7.7%) with consensus above 90%. This difference may also explain why correct answers were associated with higher confidence overall (72.9%) than were wrong answers (68.5%).

To examine this interpretation of the results more closely, I selected 13 CC items such that each of them matched as closely as possible the proportion of consensual responses of one of the 13 CW items. For this set of 26 items, an Item Class (CC vs. CW) × Correctness ANOVA (based on 37 participants because 2 additional participants did not produce wrong responses in the selected CC items) yielded $F(1, 36) = 2.37$, $MSE = 62.77$, $p < .14$, for item class; $F(1, 36) = 20.54$, $MSE = 54.35$, $p < .0001$, for correctness; and $F(1, 36) = 37.10$, $MSE = 105.04$, $p < .0001$, for the interaction. Mean confidence for the CC items (69.8%) was still higher than that for the CW items (67.8%), but the difference was smaller than was observed for the entire pool of items. The crossover interaction noted earlier was replicated (see Figure 1B): For the CC items, confidence was higher for correct answers than for wrong answers, $t(36) = 8.13$, $p < .0001$, whereas, for the CW items, confidence was higher for the wrong answers than for the correct answers, $t(36) = 2.16$, $p < .05$. However, as suggested by the Class × Correctness interaction, the discrimination between consensual and nonconsensual answers was significantly better for the CC (15.8 percentage points) than for the CW items (4.8 percentage points).

It has been proposed that the confident wrong responses to deceptive items are generally based on familiarity rather than recollection (Kelley & Sahakyan, 2003). Is it perhaps the case that the relationship between confidence and consensuality is due only to participants who are less knowledgeable and base their responses mostly on familiarity or guessing? To examine this question, I divided participants at the median of the percentage of correct answers into 20 participants who exhibited low percentage and 21 who exhibited high percentage. Percentage correct averaged 52.8% and 63.4%, respectively, for the two groups. Although confidence was somewhat lower for the low (67.6%) than for the high group (71.9%), $t(39) = 2.44$, $p < .05$, both groups exhibited marked overconfidence, $F(1, 39) = 119.67$, $MSE = 23.05$, $p < .0001$, which was stronger for the low than for the high group, $F(1, 39) = 8.67$, $MSE = 23.05$, $p < .001$, for the interaction (see also Dunning, Johnson, Ehrlinger, & Kruger, 2003). Nevertheless, the results for each of the groups yielded a pattern similar to that obtained for the entire sample (see Figure 1). Thus, a two-way ANOVA, Item Class (CC vs. CW) × Correctness, carried out separately for each group (19 and 20 participants each, who had both correct and wrong answers in each item class), yielded a significant interaction for the low-accuracy group, $F(1, 18) = 23.24$, $MSE = 65.86$, $p < .0001$, as well as for the high-accuracy group, $F(1, 19) = 24.15$, $MSE = 78.00$, $p < .0001$. In fact, a three-way ANOVA, including group (low vs. high accuracy) as a between-subject factor, yielded $F < 1$ for the triple interaction. Thus, the pattern depicted in Figure 1 held true regardless of degree of knowledge.

In sum, the results are generally consistent with those of Koriat (1976) for the phonetic-symbolism task and with those of Koriat (1995) for FOK judgments about the likelihood of identifying an unrecalled memory target. These results indicate that confidence in an answer is correlated with its consensuality—the extent to which it is endorsed by the majority of participants—rather than with its correctness. Because, for most general-information questions, the consensual answer is also the correct one (as is the case for the CC items), confidence judgments generally correlate positively with accuracy.

Confidence–accuracy correlations. I calculated also the within-person C-A correlation (resolution) using gamma as a measure of the extent to which participants can discriminate between correct and wrong answers (see Nelson, 1984). The gamma correlation across all items averaged .24, significantly different from zero, $t(40) = 10.27$, $p < .0001$. This correlation, however, differed markedly for the three types of items. It was positive (.47) across the 35 CC items (using each time only participants for whom the correlation was computable), $t(39) = 9.49$, $p < .0001$, but was significantly negative (−.24) across the 13 CW items, $t(39) = 2.82$, $p < .01$. As expected, the correlation (.04) was not significant across the 57 NC items, $t(40) = 0.95$. A one-way ANOVA, comparing the three correlations (based on 39 participants) yielded $F(2, 76) = 28.15$, $MSE = 0.169$, $p < .0001$. This pattern of results is similar to that reported by Koriat (1976) for FOK ratings in the phonetic-symbolism task. In that study, the correlation across items, between the percentage of correct translations of an item and the mean FOK rating associated with it, was .63 for CC items, −.43 for CW items, and .13 for NC items. In Koriat's (1995) study, the within-person correlation between FOK judgments about nonrecallable targets and the subsequent accurate recognition of these targets averaged .31 for CC items and −.18 for the CW items, both significantly different from zero. Brewer and Sampaio (2006) also reported a within-person gamma correlation between confidence and accuracy of .26 for nondeceptive sentences and −.44 for deceptive sentences.

In sum, confidence judgments are not inherently diagnostic of accuracy. The within-person C-A correlation is positive only when the consensually favored response is the correct response. In contrast, it is negative, and significantly so, when the consensual response is the wrong response. The observation that this pattern is found consistently across several studies that differed in many procedural details testifies to the robustness of the consensuality principle: Metacognitive judgments are correlated with the consensuality of the response rather than with its accuracy.

Decision time as a mediator of the C-C correlation. The results just reported are consistent with the PMM theory of subjective confidence (Gigerenzer et al., 1991), which predicts that confidence judgments should be calibrated as long as the general-knowledge questions are drawn randomly from a reference class that is familiar to the participants. In that case, participants' mental representations will be adapted to the ecological reference class. When the questions are not representative, the cue validities activated are expected to lead to incorrect choices and to inflated confidence judgments, as was found to be the case for the CW questions.

The C-C correlation is assumed to follow from the postulate of PMM theory that choice and confidence are both determined by cue validities. In contrast, theories that stress mnemonic determinants of confidence have derived their strength precisely from demonstrations of dissociations between confidence and performance (Busey et al., 2000; Chandler, 1994; Garry, Manning, Loftus, & Sherman, 1996). How would such theories account for the C-C correlation?

One possibility is that experiential cues, such as decision latency, are correlated with the consensuality of the response so that a C-C correlation will be obtained even when confidence is based on such cues rather than on information stored in long-term memory. Thus, it was hypothesized that responses that are endorsed by a larger proportion of participants are also made faster by each person than those endorsed by a smaller proportion. Because confidence in a response correlates positively with the speed of making that response, reliance on response latency as a basis of confidence can also yield a C-C correlation.

To test this hypothesis, I examined mean decision time for correct and wrong answers for the three classes of items.[1] The results are presented in Figure 2, based on 39 participants as in Figure 1. A clear crossover interaction is evident in comparing the results for the CC and CW items. An Item Class (CC vs. CW) × Correctness (correct vs. wrong) ANOVA on decision time yielded a significant interaction, $F(1, 38) = 17.12$, $MSE = 5.42$, $p < .0005$. The main effects of class and correctness were not significant ($F < 1$ for both). The CC items yielded the typical pattern of faster decision times for correct than for wrong answers, $t(38) = 4.48$, $p < .0001$ (Koriat et al., 2006; Robinson et al., 1997). The CW items, in contrast, exhibited the opposite pattern, with wrong answers yielding faster response times than correct answers,

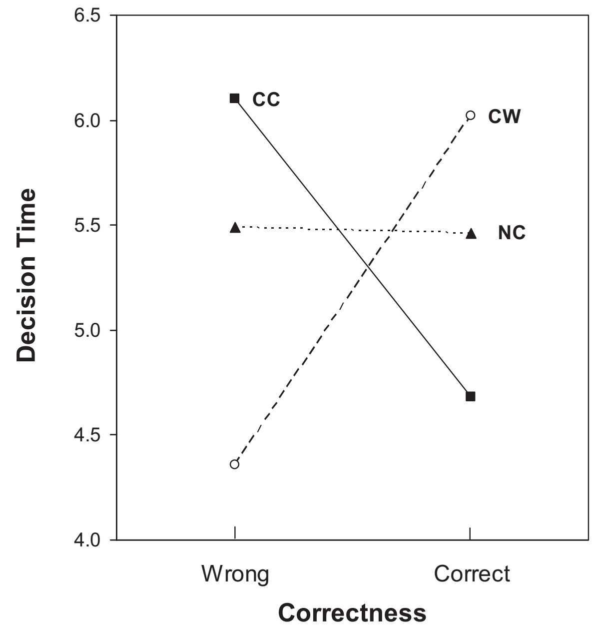

Figure 2. Mean decision time for correct and wrong answers, plotted separately for the consensually correct (CC), consensually wrong (CW), and nonconsensual (NC) items.

$t(38) = 2.33$, $p < .05$. For the NC items, the difference between correct and wrong answers was not significant, $t(39) = 0.14$, *ns*.

Overall, the pattern of results for decision time (see Figure 2) mimics that observed for confidence judgments (see Figure 1). The similarity between the two patterns supports the possibility that experience-based confidence judgments that are based on response latency as a mnemonic cue for subjective confidence should also yield a C-C correlation.

The relationship between decision time and accuracy. The within-person gamma correlation between decision time and accuracy is an index of the extent to which decision time can be relied upon as a cue for discriminating between correct and wrong answers. This correlation averaged only −.04 across all items but differed markedly between the three classes of items. As expected, it was negative (−.26) for the CC items (here and in the following analyses, using each time only participants for whom the correlation was computable), $t(39) = 5.34$, $p < .0001$, but was significantly positive (.20) for the CW items, $t(39) = 3.22$, $p < .005$. The correlation for the NC items (.01) was not significant, $t(40) = 0.53$. A one-way ANOVA comparing the three correlations (based on 39 participants) yielded $F(2, 76) = 26.74$, $MSE = 0.073$, $p < .0001$. Thus, although response latency is generally a valid predictor of accuracy (Costermans et al., 1992; Koriat et al., 2006; Robinson et al., 1997), such is only true for typical items, for which participants' responses are correct by and large.

The relationship between decision time and confidence. Although decision time was not correlated uniformly with accuracy,

[1] Because of a program error, the decision time for the first item (which appeared immediately after the five practice items) was not recorded. Thus, the decision latency analyses were based only on 104 items for each participant.

participants could have made use of it as a cue for accuracy had they had access to the classification of the items as deceptive or nondeceptive. The results, however, suggest that decision time influences confidence judgments in the same way for all items under the heuristic that easily reached answers are more likely to be correct, as can be seen in Figure 3. The data for this figure were generated by first splitting the decision times for each participant at the median for each class of items and then calculating mean confidence judgments for below-median (short) and above-median (long) decision times. These means are plotted in Figure 3 as a function of the actual mean decision time for short and long responses. It can be seen that the same general trend was observed for all three classes of items: Confidence decreased with increasing decision time. A two-way ANOVA, Decision Time (short vs. long) × Class (CC, CW, and NC), on confidence judgments yielded a significant effect for decision time, $F(1, 40) = 81.81$, $MSE = 59.95$, $p < .0001$. The effect of class was significant, $F(2, 80) = 163.54$, $MSE = 21.17$, $p < .0001$, consistent with the previously noted differences in mean confidence for the three classes of items (see Figure 1). The interaction was also significant, $F(2, 80) = 13.91$, $MSE = 27.70$, $p < .0001$, possibly reflecting the observation that the utilization of response latency as a cue for confidence was lower for the NC than for the CC and CW items. However, the difference in confidence between short and long decision latencies was significant for all three classes of items: For the CC items, $t(40) = 11.14$, $p < .0001$; for the CW items, $t(40) = 5.69$, $p < .0001$; and for the NC items, $t(40) = 4.03$, $p < .0005$. Thus, the same relationship between confidence and decision time holds for all three classes of items. It would seem that participants apply the same heuristic indiscriminately, a heuristic that is valid for representative items. Thus, reliance on decision latency as a cue for confidence should yield a pattern in which confidence is correlated with the consensuality of the answer rather than with its correctness.

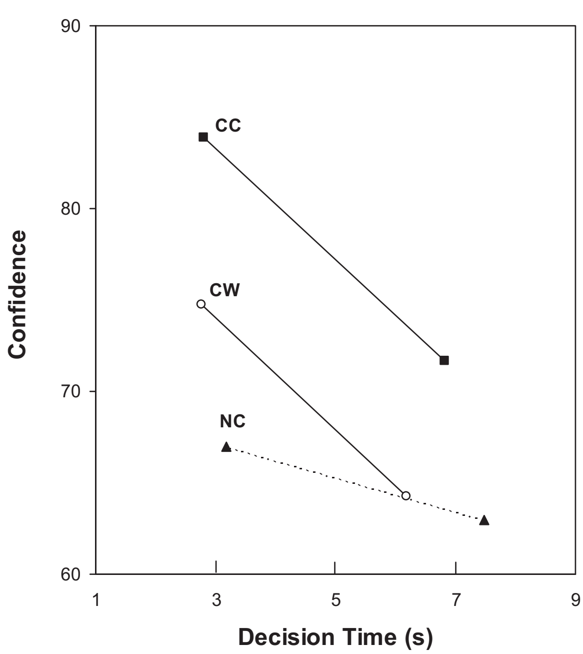

Figure 3. Mean confidence for answers with below-median and above-median decision times, plotted separately for the consensually correct (CC), consensually wrong (CW), and nonconsensual (NC) items.

Self-consistency as a mediator of the effects of consensuality on confidence. I turn next to the second internal cue that may affect confidence judgments—self-consistency. Assuming that in answering a question, different considerations are explored before settling on an answer, then confidence in the chosen answer may be expected to increase with the extent to which these considerations speak consistently for the same answer. Whereas self-consistency breeds conviction, self-inconsistency and vacillation should produce feelings of doubt (Adams & Adams, 1961) and uncertainty (Kahneman & Tversky, 1973; Saito, 1998; Slovic, 1966). The C-C correlation may stem precisely from reliance on self-consistency as a basis for confidence, assuming that inconsistency between different participants reflects inconsistency and conflict within participant.

Self-consistency was indexed by the percentage of first-presentation choices that were repeated by each participant in the second presentation of the questions, assuming that cross-presentation consistency is indicative of the consistency among the considerations that are retrieved in each presentation. This percentage averaged 83.4% across all participants and items. Mean repetition percentage was 89.8%, 83.7%, and 79.5% for the CC, CW, and NC items, respectively, $F(2, 80) = 22.00$, $MSE = 49.93$, $p < .0001$, significantly differing one from each other according to Scheffé (1953).

To examine the relationship between cross-person consensus and within-person consistency, I defined for each item a majority answer and a minority answer according to the distribution of responses made in the first presentation. I then calculated for each answer the percentage of participants who repeated that answer in the second presentation. Across all items, repetition percentage averaged 86.4% for the majority answer and 77.1% for the minority answer, $t(40) = 8.53$, $p < .0001$. Thus, answers that were consistently selected across participants tended to be consistently selected by the same person on a second encounter with the item.

I examined next mean repetition percentage for correct and wrong answers for the three classes of items. The results are presented in Figure 4, based on 39 participants as in Figure 1. The pattern displayed in this figure mimics the patterns depicted in Figures 1 and 2. A two-way ANOVA, Item Class (CC, CW, and NC) × Correctness (correct vs. wrong), yielded nonsignificant effects for item class and correctness, $F(2, 76) = 1.69$, $MSE = 233.74$, *ns,* and $F(1, 38) = 1.89$, $MSE = 303.67$, *ns,* respectively. The interaction, however, was highly significant, $F(2, 76) = 20.03$, $MSE = 218.88$, $p < .0001$. For the CC items, repetition percentage was significantly higher for correct answers than for wrong answers, $t(38) = 6.38$, $p < .0001$, whereas, for the CW items, it was higher for wrong answers than for correct answers, $t(38) = 2.16$, $p < .05$. The NC items yielded little difference between correct and wrong answers, $t(38) = 0.78$, *ns*. Thus, degree of deliberation, as reflected in the tendency to choose a different answer on each of the two occasions, may also mediate the relationship between consensuality and confidence.

The relationship between self-consistency and confidence. I also examined the relationship between self-consistency and confidence by dividing the answers that were given by a participant in Presentation 1 into those that were repeated in Presentation 2 and

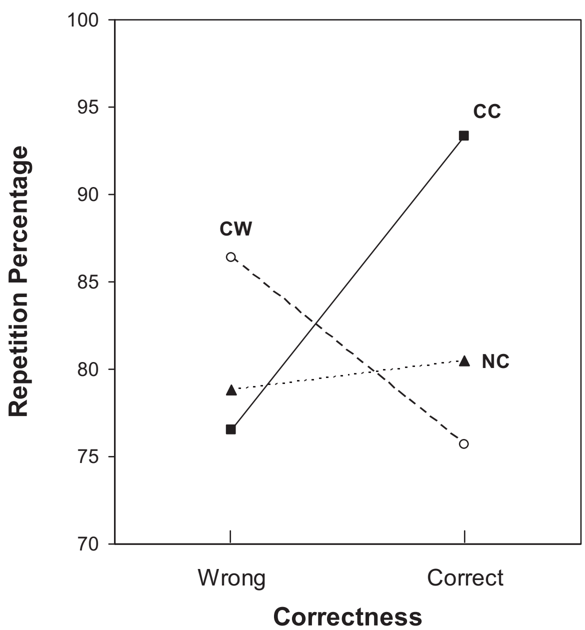

Figure 4. Repetition percentage for correct and wrong answers, plotted separately for the consensually correct (CC), consensually wrong (CW), and nonconsensual (NC) items.

those that were not repeated. Confidence for repeated answers was 72.5%, compared with 56.4% for nonrepeated answers, $t(40) = 19.44$, $p < .0001$. Confidence in the second presentation was also higher for the repeated (72.7%) than for the nonrepeated answers (56.0%), $t(40) = 18.57$, $p < .0001$.

Because inconsistent answers may sometimes reflect total lack of knowledge, the analyses were repeated using only responses with confidence judgments above 60.0% in Presentation 1. The differences just noted remained significant: Confidence judgments averaged 87.5% for the answers that were repeated in the second presentation, and 80.0% for the nonrepeated answers (using 36 participants who had both repeated and nonrepeated answers), $t(35) = 5.71$, $p < .0001$. Confidence in the second presentation was also higher (84.8%) for the repeated than for the nonrepeated (67.4%) answers, $t(35) = 9.42$, $p < .0001$.

Using the entire range of confidence judgments, confidence was higher for repeated than for nonrepeated answers for each of the three classes (including in each case only participants for whom both means were available): $t(39) = 10.42$, $p < .0001$, for CC items; $t(35) = 13.61$, $p < .0001$, for CW items; and $t(40) = 13.64$, $p < .0001$, for NC items. Saito (1998) also reported results indicating that confidence in one presentation of a set of general-information questions is correlated with the likelihood of repeating the same answer on a second presentation of the questions.

In sum, assuming that the choice of different answers in the two presentations is indicative of the degree of conflict and fluctuation that is experienced by the person on each encounter with the question, then the results suggest that, like decision time, self-consistency is also used indiscriminately across items, under the heuristic (which is valid for representative items) that answers that are consistently favored by different considerations are more likely to be correct.

The relationship between self-consistency and decision time. The results presented so far indicate (a) that both decision time and repetition percentage are correlated with the consensuality of the answer such that consensual answers are made faster and are more likely to be repeated and (b) that both decision time and repetition percentage are correlated with confidence such that faster responses and repeated answers are endorsed with stronger confidence. These findings accord with the idea that decision time and self-consistency may mediate the C-C correlation.

To complete the picture, I examined the relationship between decision latency and repetition percentage. For each participant, the responses in the first presentation were split into those that were repeated and those that were not repeated in Presentation 2. Median decision times were shorter for repeated answers (4.27 s) than for nonrepeated answers (4.53 s), but the difference was short of significance, $t(40) = 1.68$, $p < .11$. The weak link between self-consistency and decision time is somewhat unexpected because it may indicate that the C-C correlation may be produced by reliance on different cues that are only partly correlated. Note, however, that median decision latency in Presentation 2 was significantly shorter for repeated answers (2.23 s) than for nonrepeated answers (3.03 s), $t(40) = 6.03$, $p < .0001$.

Discussion

The results of this study provide further evidence in support of the claim that metacognitive judgments are diagnostic of the consensuality of the response rather than its correctness. Before discussing these results, I comment on some of the unique properties of the methodology used in this study.

Methodological Considerations

The methodology of this study followed the item-based logic underlying the work of Koriat and his associates (Koriat, 1995; Koriat & Lieblich, 1977) and Brewer and his associates (Brewer & Sampaio, 2006; Brewer et al., 2005). This logic is based on two assumptions: first, that memory questions differ reliably across participants in characteristics that are pertinent to metacognitive judgments, and second, that normative data, aggregated over participants, provide information about processes that take place within participants (e.g., Schreiber & Nelson, 1998). These assumptions imply that some insight into the mechanisms underlying metacognitive judgments can be gained by studying differences between memory items.

Indeed, in Koriat and Lieblich's (1977) study on the tip of the tongue (TOT), memory pointers (word definitions) were found to differ reliably in terms of two orthogonal dimensions, the likelihood of their suggesting or eliciting the correct target and the likelihood of their evoking a preliminary FOK or feeling of not knowing. The analysis of the items that differ in terms of both dimensions provided some insight into the processes underlying FOK and TOT states, as well as the conditions that precipitate illusions of knowing. Likewise, in Koriat's (1995) study, general-information questions were found to differ reliably across participants in properties that seem to affect both the strength of FOK judgments that they induce when recall of the target fails and the accuracy of these judgments in predicting memory performance. The work of Brewer and his associates (Brewer & Sampaio, 2006; Brewer et al., 2005) also testifies to the usefulness of exploiting differences between memory items to shed light on the bases of metamemory judgments and their accuracy.

In this study, too, I focused on a normative characteristic of items that seems to be correlated with confidence—the percentage of participants who got the answer right. The results replicated the pattern observed for the phonetic-symbolism word-matching task (Koriat, 1976) and help extend the consensuality principle to a memory task that is more representative of the kind of everyday situations in which differences in subjective convictions have behavioral implications (Fischhoff et al., 1977; Goldsmith & Koriat, 2008). In both studies it was necessary to secure a sufficient number of items for which participants significantly agreed on the wrong answer. In the present study, I took advantage of the fact that real-world knowledge exhibits many occasions in which, for one reason or another (see Fischhoff et al., 1977), recall or recognition tends to deviate reliably from veridicality. Although these items may be nonrepresentative, their inclusion was critical for dissociating correctness from consensuality.

The implication of this methodological approach is twofold. On the one hand, the results underscore the importance of a representative design, as stressed by Brunswik (1956; see Dhami, Hertwig, & Hoffrage, 2004; Gigerenzer et al., 1991). Generalizations about subjective monitoring require a sampling of stimuli that reflects the ecology toward which these generalizations are intended. In previous work on subjective confidence, this principle has been discussed primarily in connection with calibration: It was argued that the overconfidence bias that has been observed in many studies derives from a biased sampling of items by researchers (Gigerenzer et al., 1991; Juslin et al., 2000). In this study, in contrast, I focused on resolution—the ability to discriminate between correct and wrong answers. Thus, the observation that the C-A correlation was only .22 across the entire sample of items is probably not very informative for making conclusions about the real world because of the nonrepresentative sampling of items in this study.

On the other hand, however, the results highlight the theoretical benefits that ensue from a deliberate inclusion of nonrepresentative items (see Roediger, 1996). It is this inclusion that allows dissociating the effects of correctness from those of consensuality, thus providing some clues into the mechanism underlying the successful monitoring of one's own performance in many real-life situations (see Koriat, Pansky, & Goldsmith, 2007).

I now review and discuss the main findings of the present study. I first examine the evidence supporting the claim that confidence judgments are correlated with the consensuality of the answer rather than with its correctness. Then, I discuss theoretical accounts of the C-C correlation, focusing on the distinction between information-based and experience-based metacognitive judgments.

The Consensuality Principle

The results on the whole confirm the consensuality principle. This principle has now been demonstrated for the phenomenal experience accompanying the task of guessing the meaning of foreign words (Koriat, 1976) and for FOK judgments on unrecalled general-information questions (Koriat, 1995) and is also consistent with the findings of Brewer and Sampaio (2006) on sentence memory. Taken together, these results suggest that metacognitive judgments are sensitive to properties of the response that make it compelling to the majority of participants. Because, by and large, people are more likely to be right than wrong, metacognitive judgments tend to be diagnostic of the correctness of the response. However, when the wrong response is the consensual response, it is this response that will be endorsed with stronger confidence. In that case, metacognitive judgments will be counterdiagnostic of the correctness of the answer (see Benjamin, Bjork, & Schwartz, 1998). Thus, even if the consensuality principle is treated only as a descriptive principle, it depicts a powerful correlation that seems to hold across different domains and, thus, to have a significant predictive value. Of course, the results also suggest that participants have no access to the classification of an item as a CC item (representative) or a CW item (nonrepresentative, deceptive). Had they had such access, they could have improved their monitoring and control processes immensely (see Koriat & Goldsmith, 1996).

The immediate theoretical implication of this principle, however, is that it argues against trace-access (or direct-access) accounts of metacognitive judgments according to which such judgments are based on privileged access to the presence or strength of memory traces (e.g., Cohen et al., 1991; Hart, 1965; see Van Zandt, 2000). These accounts were motivated primarily by findings indicating that participants are generally accurate in monitoring their memory. However, the results of this study indicate that monitoring accuracy is not inherent to memory functioning and is not guaranteed, but varies systematically with properties of the sampled items. The C-C correlation suggests that the positive within-person C-A correlation that has been generally reported in previous studies derives from the fact that in the sample of items included in those studies, participants' actual performance was overall better than chance.

These findings also support Koriat's (1993) contention that metacognitive accuracy is a by-product of memory accuracy. Because memory is accurate by and large, metamemory tends also to be accurate. When memory goes wrong, metamemory will also go wrong. Hence, when memory is led astray for a variety of reasons, it might be futile to expect that metacognitive judgments such as FOK or subjective confidence would have privileged access to the correct target (see Koriat, 1993, 1994).

As indicated earlier, the present results are also in line with the recent findings of Brewer and his associates (Brewer & Sampaio, 2006; Brewer et al., 2005; see also Fischhoff et al., 1977) in which confidence judgments were compared between deceptive and nondeceptive sentences. However, whereas Brewer and his associates described their results in terms of the contrast between deceptive and nondeceptive items, I preferred to cast the results of the present study in terms of the consensuality of the answer. The weakness of this formulation, of course, is that consensuality does not bring to mind a specific theoretical mechanism (see below). Its advantage, however, is that consensuality is defined strictly in terms of cognitive performance, and hence, its relation to metacognitive accuracy is not trivial. In contrast, the terms *deceptive* and *misleading* are often used to imply not only a cognitive error but also a metacognitive failure. For example, Brewer and Sampaio (2006) stated, "We use the term *deceptive* to refer to items that tend to lead to memory errors without the rememberer being aware of the processes leading to the errors" (p. 541). Such a definition invites treating deceptive items as a special category in terms of the underlying process: "the confidence/accuracy relation breaks down when the memory performance involves unconscious reconstructive memory processes that lead to a memory response that is different from the original input information" (Brewer & Sampaio, 2006, p. 541). In contrast, I do not see the process underlying confidence judgments for deceptive items as differing in quality from the process underlying confidence judgments in general. The process that leads to erroneous metacognitive judgments in some cases is the same as that which is responsible for the accuracy of these judgments in other cases. Thus, describing the results in terms of consensuality has the advantage that it allows the critical variable to be specified in terms of a factor that is operationally separate from metacognitive accuracy and error.

The Processes Underlying The Confidence–Consensuality Correlation

Let us now examine the theoretical accounts of the C-C correlation. It was proposed that several mechanisms can converge in producing this correlation. These mechanisms presumably differ in their characteristics and in the conditions under which they are likely to dominate in affecting metacognitive judgments. Nevertheless, each of them, alone or in combination with the others, may contribute to the C-C correlation. Of course, future research should attempt to specify their relative contribution under different conditions, but in this study, I chose to focus on the consensuality principle as their common outcome.

I examined these mechanisms in the context of the distinction between information-based and experience-based metacognitive judgments (Kelley & Jacoby, 1996; Koriat et al., in press; Matvey, Dunlosky, & Guttentag, 2001; Slovic & Peters, 2006; Strack, 1992). Information-based judgments are assumed to rely on a deliberate inference from domain-specific beliefs and theories and on specific knowledge retrieved from memory. Indeed, several theories of subjective confidence assign a critical role to information-based processes: They assume, for example, that confidence is based on the strength of the evidence that is marshaled in favor of the chosen answer relative to the evidence in support of the alternative answers (e.g., Griffin & Tversky, 1992; Koriat et al., 1980; Yates et al., 2002). The PMM theory is an example of such theories (see Todd & Gigerenzer, 2007). It assumes that both choice and confidence are based on stored domain-specific knowledge about cues and their cue validities. The C-C correlation can be seen to ensue from the assumption that confidence and choice are both determined by the same set of cue validities. Presumably, in the case of the CC items, the hierarchy of cues used and their validities correspond to those of the ecological validities and hence lead to confidence judgments being relatively well calibrated. Such is probably not the case for CW items, which yield a large proportion of incorrect answers and unwarranted confidence levels.

The PMM theory can, in fact, be extended to account for the FOK results of Koriat (1995) as well. As noted earlier, according to Koriat's (1993) accessibility model, FOK judgments are based on the accessibility of partial clues about the target (the number of clues retrieved and their ease of access) regardless of the accuracy of these cues. For typical memory pointers, however, the majority of partial clues about the elusive target are correct, and hence their accessibility is diagnostic of the availability of the target in memory. This is clearly true for the CC items in Koriat's (1995) study, which were found to bring to mind more correct than incorrect answers among participants who did produce an answer. The CW memory pointers, in contrast, precipitated more wrong than correct answers among participants who did recall a target and presumably activated many incorrect partial clues among participants who failed to produce any answer. Hence, the FOK judgments provided by these participants were likely to be based on nondiagnostic cues. Clearly, this account, as well as the PMM account of confidence, assumes that participants do not have access to the actual ecological validities of the cues underlying their metacognitive judgments.

Turning next to experience-based judgments, these are assumed to rely on heuristics that make use of internal, mnemonic cues. These cues reside in the online feedback from the cognitive operations that are used to perform the task (Kelley & Jacoby, 1996; Koriat & Levy-Sadot, 1999). It was proposed that experience-based judgments may also contribute to the C-C correlation. This proposition was tested with regard to two potential cues—decision time and self-consistency. For each of these, I examined how it relates to consensuality on the one hand and to subjective confidence on the other hand.

The results were quite clear for both cues. With regard to decision time, a comparison of decision times for correct and wrong answers for the three classes of items (see Figure 2) yielded a pattern that mimicked closely the pattern observed for confidence judgments (see Figure 1): For the CC items, decision time was shorter for correct than for incorrect answers, as has been found to be the case by others (Costermans et al., 1992; Koriat et al., 2006; Robinson et al., 1997), but the reverse was true for the CW items. In parallel, the relationship between decision time and confidence was found to be the same across item types (see Figure 3): Confidence increased as decision time decreased for all three classes of items. These results indicate that reliance on decision latency as a cue for confidence can also produce a C-C correlation.

The same general pattern was observed for self-consistency as indexed by the tendency to choose the same answer again in the second administration of the test. The results for response repetition (see Figure 4) mimicked the pattern observed for the relationship between consensuality and confidence (see Figure 1): For CC items, rate of repetition was higher for correct than for wrong answers, whereas the opposite was true for CW items. At the same time, confidence was higher for repeated than for nonrepeated answers for each of the three classes of item.

The two internal cues—response latency and self-consistency—were related in that decision times in Presentation 1 were shorter for answers that were repeated in Presentation 2 than for those that were not repeated. Surprisingly, however, the relationship was not strong. Perhaps the observed effects of consensuality reflect the combined effects of several cues that are only partly intercorrelated. If so, there is a methodological advantage in focusing on consensuality at the empirical level while attempting to specify the mechanisms that mediate its effects.

Several questions remain open and must be addressed in future research. The first question concerns the possible mediating role of mnemonic cues in the C-C correlation. The results indicated very systematic relations between consensuality, decision time, self-consistency, and confidence, but what is the mechanism underlying these relations? For example, why is it the case that consensually favored answers are associated with greater fluency (faster response time) and self-consistency? This question requires a model that specifies the processes underlying choice and decision and perhaps the relationships between these processes and those underlying confidence. Several such models exist (e.g., Griffin & Tversky, 1992; Juslin & Olsson, 1997; Vickers, 1979), but their evaluation is beyond the scope of this article.

A second question concerns the relation between information-based and mnemonic-based processes. Previous research indicated that these two types of processes can sometimes conflict, affecting metacognitive judgments in opposite directions. For example, Koriat et al. (in press) had participants choose an answer to general-information questions, then list either one reason or four reasons in support of their choice, and finally indicate their confidence in the correctness of the answer. Participants' confidence judgments were lower after listing four reasons than after listing one reason. Presumably, the effects of the mnemonic cue of ease of retrieval (one reason is easier to retrieve than four) won over the effects of the declarative, informational content of the reasons listed (see Schwarz, 2004, for a review). In the present article however, it was argued that both types of processes may converge in producing the C-C relationship. What is the contribution of each of them to this relationship?

The observation that the C-C correlation was equally found for participants exhibiting very different levels of knowledge may suggest a greater contribution of experience-driven than of information-driven processes. This suggestion is consistent with other results that testify to the failure of beliefs and knowledge to affect metacognitive judgments (e.g., Koriat et al., 2004) and with the claim that people often choose in line with their intuitions even when other information undermines the validity of these intuitions (Simmons & Nelson, 2006).

Other observations, however, suggest a contribution of information-driven processes that cannot be accounted for by the consensuality principle. For example, confidence judgments were higher for the consensual answer for both CC and CW items, but the difference was larger for the CC items (see Figure 1) even when the two sets of items were matched in terms of the proportion of consensual responses. In addition, confidence was overall higher for the CC than for the CW items (see Figure 3). These observations suggest an effect of the actual correctness of the answer over and above that captured by its consensuality.

In fact, an important theoretical question is whether we should leave room for the possibility that one's answer is sometimes based on direct knowledge that is associated with strong confidence. Clearly, in many cases, people's choice of a particular answer may be based on the recollection of specific, idiosyncratic details. For example, in one of my experiments, one participant was completely confident that Canberra rather than Sydney is the capital of Australia (a CW item) because he had lived previously in Canberra. How does such sheer real-life knowledge affect confidence? Possibly, the answer provided in such cases is associated with a short response latency and high consistency, but can we be sure that the strong conviction is indeed based on these mnemonic cues rather than on the fact that the answer is based on sheer knowledge? In fact, Metcalfe (2000), who endorsed the idea that metacognitive judgments are generally inferential in nature, argued for the postulation of a special noetic state in which people have direct access to their knowledge. Assuming that answers that are based on sheer, direct knowledge are generally correct, this would explain the higher confidence in the CC than in the CW items.

How then do information-based and mnemonic-based processes combine in producing the C-C relationship? One possibility is that the informational content of the considerations that are explored by participants in choosing an answer (e.g., that a certain city has a famous soccer team; Gigerenzer et al., 1991) generally discloses the ecological structure of the environment as has been absorbed through past experience and that this structure has most likely helped to shape the subjective experience associated with the choice of that answer. Benjamin and Bjork (1996), for example, reviewed the learning principles that make retrieval fluency generally (but not always) a diagnostic cue for memory accuracy. In general, information that is better learned is more readily accessible and tends to come to mind with greater consistency and persistence. Thus, the C-C correlation may result from information-driven processes that affect confidence judgments directly (e.g., Koriat et al., 1980; Todd & Gigerenzer, 2007), but they may also affect these judgments indirectly through their influence on the mnemonic cues that are associated with task performance. In his analysis of the bases of judgments of learning, Koriat (1997) proposed several predictions that follow from the distinction between the direct and mediated effects of information-driven processes. It would be useful to extend his framework to the study of confidence judgments.

In sum, the present study provided additional evidence for the consensuality principle. This principle can serve to organize and summarize some of the experimental evidence on metacognitive judgments and appears to be consistent with both information-based and experience-based accounts of subjective confidence. The strength of this principle is that it pinpoints an operationally defined parameter that serves as a powerful mediator of the accuracy of metacognitive judgments. The results help reinforce the idea that intuitive feelings, such as the FOK or sheer subjective convictions, do not have privileged access to correct information. Rather, they are based on the same processes that underlie retrieval and decision, and when these processes go astray, so will the associated metacognitive judgments (Koriat, 1993). This conclusion motivates a detailed analysis of the complex processes underlying metacognitive judgments and their accuracy.

References

Adams, J. K., & Adams, P. A. (1961). Realism of confidence judgments. *Psychological Review, 68,* 33–45.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136,* 569–576.

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. *Perception & Psychophysics, 55,* 412–428.

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.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 287–313). Cambridge, MA: MIT Press.

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28,* 610–632.

Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. Reder (Ed.), *Implicit memory and metacognition* (pp. 309–338). Hillsdale, NJ: Erlbaum.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127,* 55–68.

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology, 72,* 691–695.

Brewer, W. F., & Sampaio, C. (2006). Processes leading to confidence and accuracy in sentence recognition: A metamemory approach. *Memory, 14,* 540–552.

Brewer, W. F., Sampaio, C., & Barlow, M. R. (2005). Confidence and accuracy in the recall of deceptive and nondeceptive sentences. *Journal of Memory and Language, 52,* 618–627.

Brown, R. W. (1958). *Words and things.* Glencoe, IL: Free Press.

Brunswik, E. (1956). *Perception and representative design in psychological experiments.* Berkeley: University of California Press.

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. (2000). Accounts of the confidence–accuracy relation in recognition memory. *Psychonomic Bulletin & Review, 7,* 26–48.

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

Cohen, R. L., Sandler, S. P., & Keglevich, L. (1991). The failure of memory monitoring in a free recall task. *Canadian Journal of Psychology, 45,* 523–538.

Costermans, J., Lories, G., & Ansay, C. (1992). Confidence level and feeling of knowing in question answering: The weight of inferential processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 142–150.

Dhami, M., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin, 130,* 959–988.

Dougherty, M. R. P. (2001). Integration of the ecological and error models of overconfidence using a multiple-trace memory model. *Journal of Experimental Psychology: General, 130,* 579–599.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12,* 83–87.

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.

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

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review, 3,* 208–214.

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

Gilbert, D. T. (2002). Inferential correction. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases* (pp. 167–184). New York: Cambridge University Press.

Goldsmith, M., & Koriat. A. (2008). The strategic regulation of memory accuracy and informativeness. In A. Benjamin & B. H. Ross (Eds.), *The psychology of learning and motivation* (Vol. 48, pp. 1–60). London: Academic Press.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24,* 411–435.

Gruneberg, M. M., Smith, R. L., & Winfrow, P. (1973). An investigation into response blockaging. *Acta Psychologica, 37,* 187–196.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56,* 208–216.

Hoffrage, U. (2004). Overconfidence. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgment and memory* (pp. 235–254). Hove, England: Psychology Press.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General, 118,* 126–135.

Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. *Psychological Review, 104,* 344–366.

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.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58,* 697–720.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237–251.

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language, 35,* 157–175.

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

Kelley, C. M., & Rhodes, M. G. (2002). Making sense and nonsense of experience: Attributions in memory and judgment. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 41, pp. 293–320). New York: Academic Press.

Kelley, C. M., & Sahakyan, L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory and Language, 48,* 704–721.

Klayman, J., Soll, J. B., & Juslin, P. (2006). Subjective confidence and the sampling of knowledge. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 153–182). Cambridge, England: Cambridge University Press.

Koriat, A. (1975). Phonetic symbolism and feeling of knowing. *Memory & Cognition, 3,* 545–548.

Koriat, A. (1976). Another look at the relationship between phonetic symbolism and the feeling of knowing. *Memory & Cognition, 4,* 244–248.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100,* 609–639.

Koriat, A. (1994). Memory's knowledge of its own knowledge: The accessibility account of the feeling of knowing. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 115–135). Cambridge, MA: MIT Press.

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

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

Koriat, A. (1998). Metamemory: The feeling of knowing and its vagaries. In M. Sabourin, F. Craik, & M. Robert (Eds.), *Advances in psychological science* (Vol. 2, pp. 461–469). Hove, England: Psychology Press.

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–325). Cambridge, England: Cambridge University Press.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133,* 643–656.

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

Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 483–502). New York: Guilford Press.

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.

Koriat, A., & Lieblich, I. (1977). A study of memory pointers. *Acta Psychologica, 41,* 151–164.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135,* 36–69.

Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (in press). Information-based and experience-based metacognitive judgments: Evidence from subjective confidence. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory.* Mahwah, NJ: Erlbaum.

Koriat, A., Pansky, A., & Goldsmith, M. (2007). *An output-bound perspective on false memories: The case of the Deese-Roediger-McDermott (DRM) paradigm.* Manuscript in preparation.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? The calibration of probability judgments. *Organizational Behavior and Human Performance, 20,* 159–183.

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.

Matvey, G., Dunlosky, J., & Guttentag, R. (2001). Fluency of retrieval at study affects judgments of learning (JOLs): An analytic or nonanalytical basis for JOLs? *Memory & Cognition, 29,* 222–233.

McKenzie, C. R. M. (1997). Underweighting alternatives and overconfidence. *Organizational Behavior and Human Decision Processes, 71,* 141–160.

Metcalfe, J. (2000). Feelings and judgments of knowing: Is there a special noetic state? *Consciousness and Cognition, 9,* 178–186.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 851–864.

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., Gerler, D., & Narens, L. (1984). Accuracy of feeling-of-knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General, 113,* 282–300.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation*: *Advances in research and theory* (Vol. 26, pp. 125–123). San Diego, CA: Academic Press.

Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research on metamemory: Pre-judgment recall and monitoring (PRAM). *Psychological Methods, 9,* 53–69.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2,* 175–220.

Read, J. D., Lindsay, D. S., & Nicholls, T. (1998). The relation between confidence and accuracy in eyewitness identification studies: Is the conclusion changing? In C. Thompson, D. Herrmann, J. Read, D. Bruce, D. Payne, & M. Toglia (Eds.), *Eyewitness memory: Theoretical and applied perspectives* (pp. 107–130). Mahwah, NJ: Erlbaum.

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

Reder, L. M., & Schunn, C. D. (1996). Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 45–77). Mahwah, NJ: Erlbaum.

Robinson, M. D., Johnson, J. T., & Herndon, F. (1997). Reaction time and assessments of cognitive effort as predictors of eyewitness memory accuracy and confidence. *Journal of Applied Psychology, 82,* 416–425.

Roediger, H. L., III. (1996). Memory illusions. *Journal of Memory and Language, 35,* 76–100.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803–814.

Saito, M. (1998). Fluctuations of answer and confidence rating in a general knowledge problem task: Is confidence rating a result of direct memory-relevant output monitoring or not? *Japanese Psychological Research, 40,* 92–103.

Sapir, E. (1929). A study in phonetic-symbolism. *Journal of Experimental Psychology, 12,* 225–239.

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika, 40,* 87–110.

Schreiber, T. A., & Nelson, D. L. (1998). The relation between feelings of knowing and the number of neighboring concepts linked to the test cue. *Memory & Cognition, 26,* 869–883.

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

Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology, 14,* 332–348.

Simmons, J. P., & Nelson, L. D. (2006). Intuitive confidence: Choosing between intuitive and nonintuitive alternatives. *Journal of Experimental Psychology: General, 135,* 409–428.

Slobin, D. I. (1968). Antonymic phonetic symbolism in three natural languages. *Journal of Personality and Social Psychology, 10,* 301–305.

Slovic, P. (1966). The cue-consistency and cue-utilization in judgment. *American Journal of Psychology, 79,* 427–434.

Slovic, P., & Peters, E. (2006). Risk perception and affect. *Current Directions in Psychological Science, 15,* 322–325.

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence–accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118,* 315–327.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences, 23,* 645–665.

Strack, F. (1992). The different routes to social judgments: Experimental versus informational strategies. In L. L. Martin & A. Tesser (Eds.), *The constructions of social judgments* (pp. 249–275). Hillsdale, NJ: Erlbaum.

Suantak, L., Bolger, F., & Ferrell, W. R. (1996). The hard–easy effect in subjective probability calibration. *Organizational Behavior and Human Decision Processes, 67,* 201–221.

Taylor, I. K., & Taylor, M. M. (1965). Another look at phonetic symbolism. *Psychological Bulletin, 64,* 413–427.

Todd, P. M., & Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. *Current Directions in Psychological Science, 16,* 167–171.

Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 582–600.

Vickers, D. (1979). *Decision processes in visual perception.* New York: Academic Press.

Wells, G. L., & Murray, D. M. (1984). Eyewitness confidence. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 155–170). Cambridge, England: Cambridge University Press.

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

Yates, J. F., Lee, J. W., Sieck, W. R., Choi, I., & Price, P. C. (2002). Probability judgment across cultures. In T. Gilovich & D. Griffin (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 271–291). Cambridge, England: Cambridge University Press.

Zakay, D., & Tuvia, R. (1998). Choice latency times as determinants of post-decisional confidence. *Acta Psychologica, 98,* 103–115.

Received July 5, 2007
Revision received November 29, 2007
Accepted January 4, 2008 ■