# When confidence in a choice is independent of which choice is made

**ASHER KORIAT**
*University of Haifa, Haifa, Israel*

For forced-choice two-alternative general-information questions, confidence in the correctness of the answer differed reliably for different questions, regardless of which answer was chosen. Results suggested that this choice-independent confidence is mediated by the domain familiarity of the question and by its tendency to bring to mind either few or many thoughts and considerations. Ratings of the questions on familiarity and accessibility yielded strong correlations with participants' confidence in whichever of two answers they had chosen, and with estimates of the percentage of participants who were likely to have chosen either of the answers in a previous experiment. The results were interpreted in terms of confirmation bias: Because items differ in the extent to which they bring to mind few or many pertinent thoughts, selective focusing on supportive evidence should yield a positive correlation between mean confidence in one answer and mean confidence in the alternative answer, as if there is no competition between them.

This study is concerned with subjective confidence in question answering. As in many previous studies (see, e.g., Lichtenstein, Fischhoff, & Phillips, 1982; McClelland & Bolger, 1994), participants chose an answer to two-alternative general-information questions and indicated their confidence in that answer. Whereas previous studies focused on the determinants of the choice and, consequently, on the confidence in that choice, the present study examined a component of subjective confidence that correlates with properties of the items itself, no matter which answer is chosen.

The item-based approach to the study of metacognitive judgments has proved useful in the past. According to this approach, some insight into the processes underlying metacognitive judgments and their accuracy can be gained by investigating systematic differences between memory items. This approach is based on the assumption that memory questions differ reliably in characteristics that are pertinent to metacognitive judgments, and that normative data, aggregated across participants, provide clues to the processes that occur within participants. Thus, in Koriat and Lieblich's (1977) study of the tip-of-the-tongue (TOT) states, word definitions were found to differ reliably in terms of two orthogonal dimensions: their likelihood of suggesting the correct memory target, and the extent to which they evoke a feeling of knowing (FOK). The analysis of items in terms of these dimensions provided information about the processes underlying FOK and TOT states, and laid the groundwork for Koriat's (1993) accessibility model of FOK judgments. Similarly, general-information questions were found to differ reliably in two properties among participants who provided an answer, and these properties predicted the magnitude and accuracy of FOK judgments among participants who failed to provide an answer to these questions (Koriat, 1995). Also, using two-alternative general-information questions, Koriat (2008) has documented a systematic correlation between subjective confidence in the answer and the percentage of participants who chose that answer: Consensually chosen answers elicited stronger confidence than nonconsensual answers, regardless of whether they were correct or wrong, implying that confidence is diagnostic of the correctness of the answer only when the consensually endorsed answer is the correct answer, as is generally the case. When the consensual answer was the wrong answer ("deceptive"; see Fischhoff, Slovic, & Lichtenstein, 1977), this answer was endorsed with stronger confidence. Similar findings were reported by Brewer and Sampaio (2006), who also used an item-based approach to study confidence in memory for sentences.

The present study focuses on another item-based correlation that I have observed over the years across several sets of data. Consider the recent study of Koriat (2008) just mentioned: Participants answered 105 questions and expressed their confidence on a 50%–100% scale. The questions had been selected to represent a wide dispersion in terms of the likelihood that participants would choose the correct answer, so that for some questions the consensually chosen answer would be the wrong answer.

For the present study, I reanalyzed the results, calculating for each item the mean confidence for correct and wrong answers for all participants who gave each answer. Across 104 items (for 1 item, all participants gave the correct answer), mean confidence for correct answers cor-

**A. Koriat, akoriat@research.haifa.ac.il**

related .64 ($p < .0001$) with mean confidence for wrong answers. Because some of the item means were based on only a few participants, the analysis was repeated using only the 95 items for which both means were based on 6 or more participants; the resulting correlation was .70 ($p < .0001$).

These correlations illustrate the *choice-independent confidence* (CIC) effect: The confidence that is associated with an item tends to be high or low no matter which answer is chosen. A simple explanation of this effect is that it is entirely due to items for which participants have no idea about the answer and therefore choose one of the answers almost at random and attach low confidence to their choice. When I eliminated items for which mean confidence for one of the two choices was 60% or less (and both means were based on 6 or more participants), however, the correlation for the remaining 53 items was still significant: .32 ($p < .05$).

It should be stressed that there were, nevertheless, systematic differences in mean confidence between different choices. Thus, despite the deliberate inclusion of "deceptive" items, mean confidence for correct answers (70.3%) was significantly higher overall than mean confidence for wrong answers (65.2%) [$t(103) = 5.01, p < .0001, \eta_p^2 = .20$], consistent with previous studies (see Koriat, 2007, for a review). Also, mean confidence was significantly higher for consensual answers—that is, answers that were chosen by the majority of participants (70.9%)—than for nonconsensual answers (64.6%) [$t(103) = 6.74, p < .0001, \eta_p^2 = .31$]. Still, a CIC effect was obtained when mean confidence was calculated separately for participants choosing the consensual answer and for those choosing the nonconsensual answer: The correlation between the two means was .69 ($p < .0001$). Note that a similar analysis of the results of Koriat, Lichtenstein, and Fischhoff (1980, Experiment 2) yielded a correlation of .42 ($p < .001$) across the 60-item means. Thus, in addition to the choice-specific correlates of confidence that have received a great deal of attention, there are also item-specific differences: Items differ reliably in the likelihood of inducing either low or high confidence, regardless of which answer is chosen.

What are the processes underlying the CIC effect? Several researchers have proposed that confidence in an answer is influenced more strongly by evidence that speaks for it than by evidence that speaks against it (Arkes, Christensen, Lai, & Blumer, 1987; McKenzie, 1997; Sniezek, Paese, & Switzer, 1990). Thus, Koriat et al. (1980) proposed that whereas the choice of an answer is based on the balance of evidence for and against each of the alternatives (Griffin & Tversky, 1992; McKenzie, 1997, 1998), confidence in the answer chosen is based on the attempt to justify that answer, which results in overweighting of the evidence that supports it. This confirmation bias (see Nickerson, 1998) was seen to lead to overconfidence in one's answers.

Assume that participants retrieve several considerations in support of each of the two alternatives. It follows that whereas choice and confidence should depend primarily on the balance of evidence in favor of these alternatives, confidence should also vary with the *absolute* strength of the evidence in favor of the chosen alternative. If questions differ reliably in the number of considerations that they bring to mind, as was reported by Koriat, Nussinson, Bless, and Shaked (2008), then we would expect to find item-specific differences in confidence, independent of the answer chosen. Of course, this should also be the case if the bias in favor of one of the answers has already developed during the search for pertinent considerations (Sniezek et al., 1990; Tsai, Klayman, & Hastie, in press).

Indeed, several studies indicated that confidence increases with the overall amount of information available. Gill, Swann, and Silvera (1998) found that added information enhanced confidence in judgments, even when it did not improve their accuracy. Tsai et al. (in press), who used a sequential information acquisition task, found that judges' confidence in their decisions rose steadily as additional information was presented, even though accuracy did not improve. Under some conditions, in fact, additional knowledge was found to increase confidence while decreasing accuracy (Hall, Ariss, & Todorov, 2007).

## EXPERIMENT 1

Experiment 1 examined the possibility that the CIC effect is mediated by differences between items in the richness of information that they bring to mind. Information richness was assessed by asking participants to rate both the familiarity of the questions and the extent to which they brought thoughts to mind. Previous research has indicated that both familiarity and accessibility contribute to metacognitive judgments (Koriat & Levy-Sadot, 2001): FOK judgments increase with the familiarity of the cues that probe memory (Metcalfe, Schwartz, & Joaquim, 1993; Reder, 1987) and with the amount of partial clues that come to mind about the elusive memory target (Koriat, 1993, 1995). Because I assumed that the familiarity and accessibility ratings of a question disclose the richness of the information that it activates, I expected these ratings to mediate the CIC effect in part.

### Method

For the domain familiarity ratings, 15 Hebrew-speaking undergraduates received a booklet containing the 105 questions (without the alternative answers) that had been used in Koriat (2008). They were asked to rate the extent to which they were familiar with the domain of the question, regardless of the amount of knowledge they had on that domain and regardless of whether they knew the answer to that question. They rated their familiarity on a scale from 1 (*very unfamiliar*) to 10 (*very familiar*).

For the accessibility ratings, 15 additional participants were told that when they try to answer a question, various thoughts and considerations come to mind that can help them choose the answer. They were asked to rate each question on the basis of whether it brought to mind few or many thoughts and considerations, using a scale from 1 (*very few*) to 10 (*very many*). They were instructed not to attempt to retrieve specific considerations but to make a global judgment about the question's likelihood of bringing many thoughts to mind.

### Results

Mean familiarity ratings ranged from 1.8 to 7.2 and averaged 4.8 across the 105 questions. Mean accessibility ratings ranged from 2.4 to 6.9 and averaged 4.6. The correlation between the two ratings was .45 across items

**Table 1**
**Correlations Across 104 Items Between Confidence in the Correct, Wrong, Consensual, and Nonconsensual Answers in Koriat (2008), and Familiarity and Accessibility Ratings in Experiment 1**

| Answer Type | Familiarity | Accessibility |
|-------------|-------------|---------------|
| Correct | .64** | .34* |
| Wrong | .52** | .43** |
| Consensual | .64** | .34* |
| Nonconsensual | .53** | .45** |

*$p < .001$.  **$p < .0001$.

($p < .0001$). The correlations across the 104 items between mean familiarity and accessibility ratings, on the one hand, and mean confidence in correct, wrong, consensual, and nonconsensual answers in Koriat's study (2008), on the other hand, are presented in Table 1.

It can be seen that *both* confidence in the correct answer and confidence in the wrong answer were significantly correlated with familiarity and accessibility. A similar pattern was found for confidence in the consensual answer and confidence in the nonconsensual answer. Thus, the overall familiarity of a question and the extent to which it brought clues to mind seemed to contribute to the confidence one associated with a question, regardless of which answer was chosen.

## EXPERIMENT 2

In Experiment 2, I obtained additional evidence for item-specific correlates of confidence. For each question, participants estimated the number of people who would choose each of the two answers. Later, they chose an answer themselves and indicated their confidence. A pattern consistent with the CIC effect was expected for both the estimation task and the confidence task, and this pattern was expected to be mediated in part by the familiarity and accessibility ratings of the question.

### Method

Twenty Hebrew-speaking undergraduates participated in a computer-controlled experiment that included three phases. In Phase 1, participants were told that in a previous experiment 100 people had been presented with general-knowledge questions, each appearing with several possible answers (the number of answers was not specified), and had been asked to choose the correct answer. Participants were then told that the same questions would now be presented to them, each with only one of the original possible answers, and their task was to estimate how many people (0–100) had chosen that answer. Each of the 105 questions of Koriat (2008) appeared

on the screen with one of the two possible answers, and participants indicated their estimate (which will be designated *Estimate 1*). The procedure for Phase 2 was the same, except that the answer was the second possible answer. Participants were told that this answer had also been one of the possible answers used in the original experiment. The estimate participants provided was designated *Estimate 2*.

For Phase 3, participants were told that they would now perform a task similar to that performed by the 100 people: Each question would appear with two possible answers, one of which was correct, and they would have to choose the correct answer. After participants pressed the space bar, indicating that they had finished reading the question, the screen displayed the two answers; participants indicated their choice by pressing one of two keys. Decision time, defined as the interval between the space bar press and the keypress, was measured. Participants then indicated their confidence on a 50%–100% scale.

The assignment of the two answers to Phases 1 and 2 was counterbalanced across participants. Both answers appeared in Phase 3, with their order counterbalanced across participants. The order of the items was random for each participant and phase.

### Results

For each question, the means across participants were calculated for Estimate 1, Estimate 2, Confidence 1 (in Phase 3, defined as mean confidence for participants who chose the alternative presented in Phase 1), and Confidence 2 (defined as the mean for those who chose the Phase 2 alternative). The correlations between these means are included in Table 2.

First, the correlation between Confidence 1 and Confidence 2 was positive and high (.73), reproducing the CIC effect. When I eliminated items for which mean confidence for one of the two choices was 60% or less, the correlation for the remaining 78 items was still significant: .60 ($p < .0001$). Thus, questions tended to induce either low or high confidence, regardless of which answer was chosen.

Second, the correlation between the mean for Estimate 1 and the mean for Estimate 2 was also positive (.57). It was .43 for the 78 items for which confidence exceeded 60%. Both of these variables correlated with confidence in such a way that Estimate 1 also correlated significantly with Confidence 2, and Estimate 2 also correlated significantly with Confidence 1. These results suggest that both the estimation of the number of participants who would choose an answer and the indication of confidence in that answer involved a biased attempt to support that answer (see Sniezek et al., 1990). This process appears to have much in common with the inflation of conditional predictions (Koriat, Fiedler, & Bjork, 2006): The assessed probability that a certain outcome will occur, given a certain condition, tends to be markedly inflated.

**Table 2**
**Correlations Across the 105 Items Between the Means of Estimate 1, Estimate 2, Confidence 1, Confidence 2, Familiarity, Accessibility, and Decision Time (Experiment 2)**

|  | Estimate 1 | Estimate 2 | Confidence 1 | Confidence 2 | Familiarity | Accessibility |
|--|-----------|-----------|--------------|--------------|-------------|---------------|
| Estimate 2 | .57** | – | | | | |
| Confidence 1 | .63** | .58** | – | | | |
| Confidence 2 | .61** | .64** | .73** | – | | |
| Familiarity | .56** | .45** | .64** | .65** | – | |
| Accessibility | .29* | .27* | .40** | .34* | .45** | – |
| Decision time | −.27* | −.29* | −.34* | −.36* | −.14 | .05 |

*$p < .01$.  **$p < .0001$.

Third, familiarity and accessibility seemed to play a key role: Familiarity was correlated positively with Estimate 1 and Estimate 2, and with Confidence 1 and Confidence 2. A similar pattern was observed for accessibility, although the correlations were somewhat lower.

Finally, Table 2 also presents the correlations with decision time (in Phase 3), which are relevant to the proposition that the effects of information richness on confidence are due to increased fluency of processing (Gill et al., 1998; Tsai et al., in press; see also Benjamin & Bjork, 1996). The results suggest that items that allowed for a faster choice of an answer were associated with higher confidence, regardless of which answer was chosen. Such items were also associated with higher estimates of the likelihood that participants in the previous experiment had chosen each of the two answers. Thus, the CIC effect may be related to fluency of processing. The correlations of decision time with familiarity and accessibility were very low, however, and do not support the idea that the effects of amount of information on confidence are mediated by enhanced processing fluency. These low correlations might have derived from the fact that participants had seen each question twice before; however, a similar trend was observed for the results of Koriat (2008), in which decision time was measured in the same way as it was measured in Phase 3 of Experiment 2. Decision time correlated $+.02$ with the familiarity ratings and $+.17$ with the accessibility ratings collected in the present study.

## DISCUSSION

This study has documented a tendency of confidence judgments to differ reliably for forced-choice questions, regardless of which answer is chosen. The results suggested that these differences are mediated by the familiarity of the question and by its tendency to bring to mind few or many thoughts: Familiarity and accessibility ratings yielded significant and quite high correlations with confidence in whichever of the two answers was chosen, and with the estimated percentage of previous participants who were likely to have chosen each answer.

The CIC effect may derive from processes that occur during any of the several phases of the question-answering task. The finding that familiarity and accessibility enhance FOK judgments suggests that the effects of these variables on confidence may occur as soon as the question itself is encoded (Reder, 1987). Indeed, Tsai et al. (in press) proposed that added information enhances confidence by enhancing the apparent coherence of the information. Gill et al. (1998), on the other hand, suggested that added information affects confidence by enhancing the fluency with which a judgment is made (Kelley & Lindsay, 1993). Consistent with this suggestion is my finding in Experiment 2 that confidence in each of the answers increased with the speed with which an answer was chosen (see Table 2). Thus, the amount of information may affect the ease with which a decision is made, no matter which decision is ultimately reached.

The account that I proposed for the CIC effect focuses specifically on the confirmation bias that occurs *after* a choice has been made (Nickerson, 1998). This account may explain the observation that familiarity and accessibility enhanced confidence no matter which answer was chosen, presumably because participants overweighted considerations that supported their choice. A further support for postdecisional processes as a source of the CIC effect comes from the occurrence of this effect in other tasks. I reanalyzed some old data from two word-matching studies (Koriat, 1975, 1976; see also Koriat, 2008). In these studies, participants were presented with antonymic words from noncognate languages (e.g., *osklivost–krasa*) and were asked to match each of these with their English equivalents (e.g., *beautiful–ugly*). In addition, they rated their confidence in their choice. Whereas in the 1975 study the items were representative of semantic dimensions (see Slobin, 1968), in the 1976 study they were deliberately selected to include a relatively large number of consensually wrong items. For each data set, I calculated mean confidence ratings for each choice. For the data of Koriat (1975), the correlation between mean confidence in the consensual choice and mean confidence in the nonconsensual choice across the 55 items was .40 ($p < .005$). The respective correlation across the 85 items of Koriat (1976) was .60 ($p < .0001$). Thus, although it is difficult to talk about the familiarity of the "question" in this context—a familiarity that could enhance predecisional confidence—the results yielded the same type of CIC effect as was found for general-information questions.

As is often noted, understanding the processes underlying confidence in a belief is important, because confidence affects whether people translate their beliefs into behavior (Gill et al., 1998; Goldsmith & Koriat, 2008; Hall et al., 2007). Indeed, in a recent study, I presented the 105 questions of Koriat (2008) to 60 participants under free-recognition instructions (see Koriat & Goldsmith, 1996): They would win 1 point for each correct answer and lose 1 point for each incorrect answer, but they could choose not to answer a question, in which case they would neither lose nor win anything (a monetary bonus was promised, depending on the number of points won). Across 104 items, the likelihood of volunteering the correct answer correlated .55 ($p < .0001$) with the likelihood of volunteering the wrong answer. The respective correlation between volunteering the consensual and nonconsensual answers was .59 ($p < .0001$). The likelihood of volunteering the correct answer correlated .68 and .39, respectively, with the familiarity and accessibility ratings obtained in this study, and the likelihood of volunteering the wrong answer correlated .55 and .39, respectively, with these ratings. These results suggest that knowing which question is presented to participants can allow us to predict how much they will be willing to bet on their answer, regardless of which answer they finally choose.

Because the present study is correlational in nature, it is hard to reach definite conclusions about the underlying mechanism; however, the phenomenon that forced-choice items differ in their associated confidence and volunteering, regardless of the choice that is made, deserves further investigation. It suggests that although one component of subjective confidence is sensitive to the competition be-

tween the alternative answers (Griffin & Tversky, 1992), another component responds mainly to the strength of support for the chosen alternative (Koriat et al., 1980).

## AUTHOR NOTE

## REFERENCES

Arkes, H. R., Christensen, C., Lai, C., & Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behavior & Human Decision Processes*, **39**, 133-144.

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

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

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

Gill, M. J., Swann, W. B., Jr., & Silvera, D. H. (1998). On the genesis of confidence. *Journal of Personality & Social Psychology*, **75**, 1101-1114.

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

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

Hall, C. C., Ariss, L., & Todorov, A. (2007). The illusion of knowledge: When more information reduces accuracy and increases confidence. *Organizational Behavior & Human Decision Processes*, **103**, 277-290.

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 & Language*, **32**, 1-24.

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

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

Koriat, A. (2008). Subjective confidence in one's answers: The consensuality principle. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **34**, 945-959.

Koriat, A., Fiedler, K., & Bjork, R. A. (2006). Inflation of conditional predictions. *Journal of Experimental Psychology: General*, **135**, 429-447.

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

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

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

Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and experience-based metacognitive judgments: Evidence from subjective confidence. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 117-135). Mahwah, NJ: Erlbaum.

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: Cambridge University Press.

McClelland, A. G. R., & Bolger, F. (1994). The calibration of subjective probability: Theories and models 1980–94. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 453-482). Chichester, U.K.: Wiley.

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

McKenzie, C. R. M. (1998). Taking into account the strength of an alternative hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 771-792.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 851-861.

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

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, **19**, 90-138.

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

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

Tsai, C. I., Klayman, J., & Hastie, R. (in press). Effects of amount of information on judgment accuracy and confidence. *Organizational Behavior & Human Decision Processes*.