ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES 46, 264–282 (1990)

The Effect of Choosing on Confidence in Choice

JANET A. SNIEZEK, PAUL W. PAESE, AND FRED S. SWITZER III

University of Illinois at Urbana–Champaign

Overconfidence has been widely observed in difficult judgment and choice tasks. That is, the subjective probabilities of correctness that people assign to their answers to difficult general knowledge questions are too high relative to actual proportions correct. The present study was designed to gain insight into the overconfidence phenomenon through manipulations of the choice and confidence assessment process. Results for three different measures of confidence show that manipulations of the choice process affected observed confidence levels. The effects of choosing on confidence suggest that overconfidence is most likely to be severe in spontaneous, less contemplated, choices. Suggestions are made for further research on (a) the relationship of choice to confidence, and (b) additional operationalizations of confidence.

Research on confidence that people express in their knowledge has led to the conclusion that confidence is often inappropriate—generally because it is too high (Lichtenstein, Fischhoff, & Phillips, 1982; Ronis & Yates, 1987; Wright, 1982). Support for this conclusion comes from studies of the subjective probabilities that people assign to their answers to difficult questions. The present research seeks to increase understanding of the overconfidence phenomenon by examining the process by which such subjective probabilities are formed.

This paper reports two experiments. In the first, the manner in which subjects produce choices and confidence assessments, and the framing of elicitation statements, are manipulated. In the second experiment, the target alternative of the confidence assessment is manipulated.

EXPERIMENT 1

Framing

It is well known that, in getting people to estimate unknown quantities, results often depend on the way in which the question is asked (Fischhoff

This research was supported in part by a grant from the Research Board of the University of Illinois at Urbana–Champaign. The authors thank Douglas May for his part in data collection, and Angela Fritchley, Joseph Pirc, and Susan Ryan for their assistance in data handling. Address correspondence and reprint requests to Professor Janet A. Sniezek, Department of Psychology, University of Illinois at Urbana–Champaign, Champaign, IL 61820.

0749-5978/90 $3.00

& MacGregor, 1983). Some of the most dramatic examples involve framing (see Hogarth, 1982). It is possible that the use of a "positive" elicitation frame (e.g., "What is the probability that your answer is *correct?*") may be partly responsible for overconfidence. Although not all studies explicitly report the question posed to obtain confidence assessments from people, the language used in most does suggest that such questions focus attention during the confidence assessment process on the correctness of the chosen alternative.[1]

Framing could be a source of overconfidence in various ways. At the simplest level, a general tendency to agree with or support the proposition under consideration would lead to overconfidence if the elicitation statement concerned only correctness of the chosen alternative. As suggested by Koriat, Lichtenstein, and Fischhoff (1980), overconfidence may be due to a biased search for confirming evidence prior to choice, as well as selective attention to confirming evidence following choice. Thus, framing could lead to overconfidence because of biases either in the generation of evidence prior to choice or in the evaluation of evidence following choice.

In this study we obtain confidence assessments with two versions of elicitation instructions: the typical positive or "correct" frame and also a negative or "wrong" frame. It must be noted that comparison of confidence levels with positive and negative frames in a task with multiple trials does not provide a test of the relative influence of bias in information search versus bias in information evaluation on overconfidence. Our general expectation is that by using confidence assessment instructions that focus on the probability of one's chosen answer being *wrong,* biases toward confirming evidence in generating and/or evaluating evidence should be reduced. Thus, confidence is predicted to be lower when the "wrong" frame is used.

The test between positive and negative frames in elicitation statements is intended to increase our understanding of the process by which information is generated and evaluated during confidence assessment. Further insight can be gained from direct manipulation of the choice process.

Choice

The act of choosing an alternative may itself affect one's confidence that a particular answer is correct. Koriat *et al.* (1980) suggest that information search is directed toward evidence supporting one's initially preferred alternative, thereby determining the eventual choice. Thus, accord-

[1] A review of confidence assessment instructions in 10 studies published between 1975 and 1987 suggests that they all used the "chosen alternative correctness" focus.

ing to this view, the initial (implicit) selection of one alternative *prior to choice* biases the search for information, and this biased search process leads to overconfidence. It follows that, if judges do not have to choose between alternatives, information search should not be biased toward either alternative and overconfidence should therefore be reduced. This analysis suggests that overconfidence in judgment may be mediated by the process of choosing between alternatives.

In contrast to the previous process, another possibility is that information search may be biased as a result of *having chosen,* when judges are asked to assess the chances that their chosen answers are correct. In comparison to no-choice situations in which the judged alternative has been selected by some other means, overt choice may promote personal commitment to the selected alternative. Furthermore, as commitment to a judged alternative increases with choice, the desire to be accurate is also likely to increase, thereby motivating judges to search for or selectively attend to additional evidence supporting their choices (see Festinger, 1957). Thus, overconfidence in judgment may be the result of *having chosen* between alternatives. This possibility does not preclude the earlier suggestion that *having to choose* contributes to overconfidence by biasing information search prior to choice.

Confidence assessments were elicited in this study in one choice condition and two no-choice conditions. In the *choice* condition, each choice between alternatives was followed immediately by an assessment of confidence. In contrast, judges in the *arbitrary cue* condition assessed their confidence that an arbitrarily preselected alternative was correct. This procedure directed judges' attention to one alternative (as would choice), but did not require the overt act of first choosing between alternatives. In the *uncued* condition, judges assessed the probability that they *would* choose the correct answer, without making an explicit choice. Unlike the other two conditions, attention in the uncued condition was not directed to one of the alternatives by choice or preselection. Fischhoff, Slovic, and Lichtenstein (1977) found no differences between conditions similar to our choice and arbitrary cue conditions; however, they examined only complete certainty responses.

There are two components of the choice process that distinguish between these three conditions: (a) the self as the agent of selection and (b) direction of attention to one of the available alternatives. The choice condition has both of these components. In the arbitrary cue condition, attention is directed to one of the alternatives, but the self is not the agent of selection. The uncued condition has neither of these components. In all three conditions, implicit selection from the set of alternatives is necessary to enable confidence assessment. Thus, if the implicit selection of one alternative biases the search for information prior to choice, overconfidence is likely to occur in all of these conditions. But if overconfidence is attributable to selective attention during information evaluation, then only the choice and arbitrary cue conditions should exhibit overconfidence. If information search following choice is biased because of personal commitment to one's choices, overconfidence should be greater when the self is the agent of selection; i.e., overconfidence should be higher in the choice condition than in the arbitrary cue and uncued conditions.

In general, we predicted that biased information search *following choice* would have the strongest effects on confidence in judgment. Specifically, we expected that when explicit choices are made, information search will be biased toward evidence supporting one's chosen alternative. (This same prediction was made by Ronis and Yates (1987), but was not yet published at the time this study was conducted.) In the absence of overt choice, however, we expected information search to be considerably less biased toward any particular alternative. Thus, confidence in judgment was predicted to be higher in the choice condition than in the two no-choice conditions.

In contrast to previous research, confidence in general knowledge in this study is assessed by not only subjective probabilities but also by alternative measures of confidence in general knowledge. The two additional measures are the judge's own post-task estimates of (a) his or her total number of correct answers (estimated score) and (b) the percentile rank of his or her score for the subject sample (see Sniezek & Henry, 1989). Multiple operationalizations of the dependent variable provide information about the generalizability of an effect of the manipulated variable. If selection or framing effects are similar for all three variables, then there would be evidence that the consequences of these manipulations extend from the level of confidence in one's choice to a single item to the level of confidence in one's overall performance. In contrast, significant effects on only the probability measure would suggest that any effects of the manipulations of framing and selection are only transitory and do not carry over to measures of the judge's more general confidence in his or her knowledge.

Method

Design. Two factors were manipulated: selection and framing. The selection factor consisted of three levels: *choice, arbitrary cue,* and *uncued.* The choice condition represented the typical situation in which the subject first chooses among alternatives, and then assesses confidence in the chosen alternative. Like the choice condition, the arbitrary cue condition directs the subject's attention to one of the alternatives prior to confidence assessment. But the arbitrary cue condition differs from the choice condition in that one alternative is arbitrarily selected (by the experimenter) prior to confidence assessment. By arbitrary we mean that options were arbitrarily precircled in an alternating sequence, and not randomly precircled (as in Ronis & Yates, 1987). The uncued condition uses no cueing of one of the alternatives through either choice or experimenter selection among alternatives.

The second factor, framing, had two levels. The attribute being estimated by participants was the probability of being correct in the positive condition, and the probability of being wrong in the negative condition. The selection and framing factors were crossed in a 3×2 factorial design.

Participants. Participants were 104 undergraduate students from introductory courses in organizational behavior and marketing at the University of Illinois at Urbana–Champaign who chose the research option for extra course credit. Eighty of the participants were randomly assigned to one of the six conditions, resulting in 14 subjects in the choice-positive and arbitrary cue-positive conditions and 13 subjects in each of the other conditions. Twenty-three subjects who participated in the choice-positive ($n = 16$) and choice-negative ($n = 7$) conditions for pilot work were added to the respective aforementioned groups. An additional sample of 25 subjects was tested, under the choice-positive condition, for the purpose of comparing calibration on the item subjective probability and estimated score confidence measures.

Task materials. Using information contained in *Webster's Dictionary* (1979), 50 items were written on a wide variety of general knowledge topics: history, recreation, literature, physical measurements, home management, animals, biographical data, and medicine. A pilot sample of 12 obtained a mean score of 28 or 56%. The fact that this mean was just above the random guessing level of 25 or 50% indicated that the items were in the difficulty range in which overconfidence is typically observed (see Lichtenstein & Fischhoff, 1977).

The items were typed on seven pages that were randomly ordered. A separate page with a practice item and instructions appropriate to the experimental condition was attached to the front. A post-task questionnaire page was placed at the end.

Procedure. Upon arrival at the laboratory, each subject was given a randomly selected form and seated at least 1 m from other subjects.

Instructions. All participants were given the same set of 50 two-alternative items and asked to provide a probability estimate for each. The instructions for all conditions included a sample item, and the information that there were 50 items, each with two alternatives—one of which was correct and the other of which was wrong. Instructions on how to complete each item and provide the probability estimate were unique to each condition.

In the *choice-positive frame* condition, the item instructions were "Read each item, then try to determine which of the two alternatives is correct. Answer by circling either alternative A or alternative B." The confidence assessment instructions were: "Then indicate the probability that *the chosen alternative is correct* by circling one number between .50 (a 50% chance of being correct) and 1.00 (a 100% chance of being correct)." A numerical scale prefaced by: "The probability that *the chosen alternative is correct* is" with values from .50 to 1.00 in increments of .10 followed.

Instructions for the *arbitrary cue-positive frame* condition were "*Do not answer the items.* Just read each item, then try to determine the probability that the circled alternative (which will be A on odd-numbered items, B on even-numbered items) is *correct* by circling one number between .00 (no chance of being correct) and 1.00 (a 100% chance of being correct)." The numerical scale had the same preface as in the choice condition, but ranged from .00 to 1.00 in increments of .10.

The *uncued-positive frame* condition had instructions that read "*Do not answer the items.* Just read each item, then try to determine the probability that you would answer the item correctly. Indicate the probability that *you would be correct* by circling one number between .50 (a 50% chance that you would be correct) and 1.00 (a 100% chance that you would be correct)." The scale was identical to the scale in the choice condition, but it was prefaced "The probability that *you would be correct* is." Following completion of the probability assessments, participants were given (without forewarning) a second set of the items (in a new order) and instructed: "Read each item, then try to determine which of the two alternatives is correct. Answer by circling either alternative A or alternative B."

For all the selection conditions, the *negative frame* conditions were formed by substituting "wrong" for "correct" throughout the confidence assessment instructions and response scale preface, and the scale ranged from 0 to .5 instead of .5 to 1.0.

The post-task questionnaire was used to obtain the alternative confidence assessment data, i.e., participants' aggregate performance estimates. In the choice condition with positive frames, participants were asked "Overall, how many of the 50 circled alternative are correct? Estimate 'total accuracy' and write it in the blank." The same question was asked in the arbitrary cue condition. But since 25 of the precircled alternatives were correct, and subjects were informed of this, this question could not serve as an alternative measure of confidence in this condition. Its purpose was to check on subjects' acceptance of the information supplied in the instructions for this condition. For the uncued condition, ". . . would you get correct?" was substituted for ". . . are correct?"

The negative frame conditions replaced "correct" and "total accuracy" by "wrong" and "total number of errors," respectively.

Results

Five indices were computed for each subject, including mean subjective probability, mean over/underconfidence, calibration score, resolution score, and proportion correct. As in Ronis and Yates (1987), choice was inferred from probabilities in the arbitrary cue condition.[2] Thus the mean over/underconfidence, calibration, and proportion correct measures for this condition are those that would be obtained for this condition if choices were consistent with probabilities for the arbitrarily circled alternative. In addition, the subject's estimated score (estimated number of correct answers) and the subject's estimated percentile rank (relative to the other subjects' scores) were used as dependent variables in the analyses. For Experiment 1, a 3 (Selection) $\times$ 2 (Framing) analysis of variance was conducted on each variable. The means and standard deviations on these measures for each condition are shown in Table 1.

A significant main effect of selection was found on nearly all the dependent measures. There was a significant selection effect on the mean subjective probability measure ($F(2,98) = 7.32$, $p < .01$). Bonferroni t tests of the differences in means in the selection conditions were conducted. Mean probability was significantly higher ($p < .05$) in the arbitrary cue conditions (both positive and negative) than in the choice conditions. There was no significant difference between the arbitrary cue and uncued conditions, nor between the uncued and choice conditions. There was also no framing main effect, nor was there a significant framing by selection interaction. There was also a significant effect of selection on proportion correct ($F(2,98) = 6.57$, $p < .01$). The only significant pairwise difference on this measure was between the choice and arbitrary cue conditions ($p < .05$). Again, there was no main effect for framing, nor was there an interaction between selection and framing. The effects of selection on mean probability and proportion correct are shown in Fig. 1. Note that as the level of (explicit and/or implicit) selection decreases, overconfidence becomes more extreme while accuracy gets worse.

The effect of selection on estimated score was found to be significant ($F(1,73) = 3.90$, $p = .05$) in the direction of greater confidence in the uncued than the choice conditions. (The arbitrary cue condition—in which the subjects predicted their own score—was excluded from the

[2] The circled alternative was assumed to be the subject's choice if the probability was $>.5$. If the probability was $<.5$, the other, uncircled, alternative was assumed to be the subject's choice. Circled alternatives assigned probabilities of .5 were randomly divided into chosen and unchosen halves.

TABLE 1
MEANS AND STANDARD DEVIATIONS ON PERFORMANCE MEASURES AS A FUNCTION OF SELECTION AND FRAMING

Condition[a]	C-N	C-P	UC-N	UC-P	AC-N	AC-P
Proportion correct						
Mean	.57	.58	.54	.56	.52[b]	.53[b]
SD	.079	.055	.057	.074	.055	.042
Mean subjective probability						
Mean	.67	.67	.69	.71	.72	.75
SD	.076	.051	.055	.074	.086	.094
Mean over/underconfidence						
Mean	.101	.089	.153	.150	.219	.216
SD	.073	.077	.064	.092	.086	.090
Estimated score						
Mean	25 (.50)	20 (.40)	28 (.56)	28 (.56)	24 (.48)[c]	18 (.36)[c]
SD	9.77	9.09	10.3	15.4	10.8	7.90
Score over/underconfidence						
Mean	−.07	−.18	.02	.00	—	—
SD	.238	.181	.231	.308	—	—
Calibration						
Mean	.049	.041	.079	.075	.121	.092
SD	.034	.025	.043	.049	.054	.046
Resolution						
Mean	.030	.034	.026	.019	.028	.016
SD	.022	.022	.027	.014	.021	.009
Estimated percentile rank						
Mean	52	46	59	47	66	62
SD	17.1	18.8	22.6	24.4	14.3	17.8

[a] C-N, choice, negative frame; C-P, choice, positive frame; UC-N, uncued, negative frame; UC-P, uncued, positive frame; AC-N, arbitrary cue, negative frame; AC-P, arbitrary cue, positive frame.

[b] Subjects' answers in these conditions were inferred from their subjective probabilities.

[c] These refer to subjects' estimates of the number of experimenter-circled answers that are correct.

analysis of estimated score. Because the items were answered by the experimenter in this condition, the measure was not meaningful in the arbitrary cue condition.) There was no main effect for framing, nor was there a significant interaction between selection and framing, on the estimated score measure.

A significant main effect for selection was found also for the mean over/underconfidence measure ($F(2,98) = 20.79, p < .001$). Arbitrary cue subjects showed more overconfidence than the other groups. The Bonferroni tests indicated significant differences ($p < .05$) among all pairs of

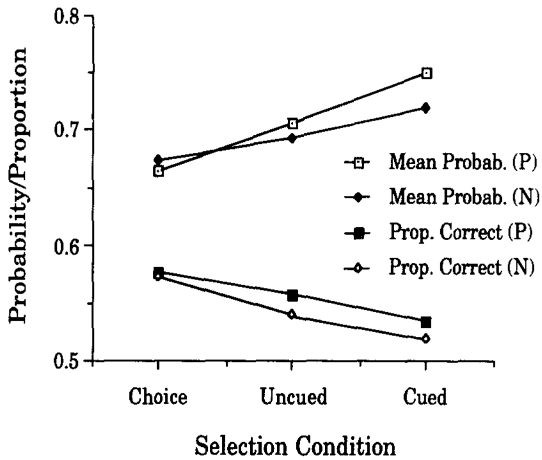

FIG. 1. Mean assigned probability and proportion correct as a function of selection condition and framing (P, positive; N, negative).

the three levels of the selection variable. Similar results were obtained for the calibration measure and for the subjects' estimated percentile rank. A main effect for selection was observed ($F(2,98) = 21.32, p < .001$), such that the arbitrary cue subjects were less well calibrated than either the uncued or the choice participants ($p < .05$). The percentile rank estimates given by arbitrary cue subjects were significantly higher (mean = .64) than rank estimates in the two other selection conditions (weighted mean = .50; $F(2,98) = 5.48, p < .01$). Again, no framing effects or interactions were found on these latter measures.

Correlation coefficients were computed between the confidence measures. The correlations of mean subjective probability with estimated score and estimated percentile rank measures were essentially zero (.016 and −.004, respectively). However, the correlation between the two post-task measures, estimated score and estimated percentile rank, was significant ($r = .449$; $p < .001$). The correlation between proportion correct and mean subjective probability was .245 ($p < .05$). The correlations of proportion correct with estimated score and estimated percentile rank were nonsignificant (−.151 and −.062, respectively).

The mean subjective probability and estimated score measures were also compared in terms of the implied level of confidence. Inspection of the Table 1 data reveals overconfidence on the former measure, but not on the latter. An additional sample of 25 was added to the 30 cases in the choice-positive condition to test for differences in confidence level between the two measures. For the entire sample of 55, the mean subjective probability was .667, and the mean estimated score was 20.5 or a proportion of .41. The difference between the mean probability and proportion was significant (paired t (54) = 11.41, $p < .001$). This sample included many subjects (n = 31) who gave estimated scores below chance (<25). Because this might have exaggerated the difference between the means, the analyses were repeated after these subjects were eliminated. For the reduced sample (n = 24), the mean subjective probability was .58, and the mean estimated score was 29.2 or a proportion of .58. This difference between the mean probability and mean proportion was significant (paired t (23) = 6.65, $p < .001$).

Discussion

We hypothesized that overconfidence is due in part to elicitation instructions that direct information search and/or attention to supportive evidence by asking individuals to state the probability of the chosen answer being correct. Confidence in choice was therefore predicted to be lower when the probability of being wrong was elicited. The results of Experiment 1 did not support this prediction, as observed confidence levels did not differ by framing condition. Judges were overconfident (on the subjective probability measure) as in previous research, but this overconfidence in choice was nearly invariant across elicitation frames. These results suggest that the well-known tendency to *overestimate* the likelihood of one's choices being correct is balanced by an equally strong tendency to *underestimate* the likelihood of one's answers being wrong.

We hypothesized that choice contributes to overconfidence because personal commitment to chosen alternatives leads judges to search for additional evidence supporting their choices. However, the observed effects of selection on confidence (as measured by mean subjective probability) were in the direction *opposite* to that predicted. Instead of being more overconfident, judges who made explicit choices were significantly less overconfident than those who did not make explicit choices. Although judges were overconfident in all conditions, confidence was more appropriate in the choice condition than in the two other conditions. Ronis and Yates (1987) recently observed the same result after making the same prediction posited in this study. However, Ronis and Yates argued that methodological artifacts might account for this outcome, and therefore did not provide a conceptual explanation of their results. Since this result was replicated in the present study, it seems advisable to consider what process might account for such a result.

The following explanation can be thought of as a heuristic process model of confidence assessment that expands on the propositions of Koriat *et al.* (1980). We propose that the appropriateness of confidence depends on the extent to which judges search for evidence concerning *each* of the alternatives. In the choice and uncued conditions, judges searched for information regarding both alternatives, and this search process was biased toward confirming evidence. Thus, in these conditions, confidence in the initially preferred answer was reduced in relative proportion to the amount of evidence in support of the other alternative (and more so when explicit choices were made because of greater involvement in the task). In the arbitrary cue condition, on the other hand, attention was focused on the preselected alternative. As a result, subjects in this condition searched their memories only for evidence concerning the preselected option. Lacking evidence in support of the "other" alternative, confidence in the arbitrary cue condition was based primarily on the relative amount of available evidence supporting the preselected option, and overconfidence was therefore more extreme. In instances in which the judge assigns probability less than .50, and is therefore "choosing" the uncircled alternative, we would have to assume that overconfidence occurs because there is so little available evidence supporting the circled alternative.

Another possibility is that subjects in the arbitrary cue condition did not believe that the preselected answers had been circled arbitrarily. Such expectations could lead to increases in overconfidence such as those observed in the present study. That is, subjects may have suspected an "unseen" manipulation on the part of the experimenter, such as the precircling of correct (or incorrect) answers. In the Ronis and Yates (1987) study, the "random cue" subjects showed a strong tendency to "choose" the precircled alternative. This may have occurred in their study because the random pattern provided by the experimenters was interpreted as providing cues about the correct answers. The alternating sequence used in the present study is less likely to be perceived as a valid cue to accuracy. The mean "estimated score" (i.e., number of experimenter circled alternatives thought to be correct) was 21 (or 42%) for the arbitrary cue condition, making it doubtful that our subjects thought that all or none of the 50 arbitrarily circled alternatives were correct. Further, a belief that all (or none) of the circled answers were correct would have also produced systematic relations between precircled answers and actual choices within subjects, with subjects choosing nearly all (or none) of the preselected options. No systematic relations were observed: the number of the 50 items for which the arbitrarily circled alternative was "chosen" ranged from 7 to 32, with a mean of 20 and standard deviation of 8.

Insufficient consideration of evidence about the uncircled option by those in the arbitrary cue condition may also explain why accuracy was lower in this condition. The lower proportion correct in the arbitrary cue condition than in the choice condition is consistent with the direction of the difference in the comparable conditions in the Ronis and Yates (1987) study (.646 for randomly cued vs .664 for choice). In addition, the respective mean subjective probabilities in the Ronis and Yates study (.812 for randomly cued and .762 for choice) showed the same direction of difference as between the arbitrary cue and choice conditions in the current study. Taken together, the opposite direction of differences between the choice and arbitrary cue subjects on subjective probability compared to proportion correct, in both the present study and the Ronis and Yates study, demonstrates that *the selection manipulation simultaneously makes people less correct, but more confident.*

Although subjects in the arbitrary cue condition were the most overconfident, subjects in the uncued condition were also significantly more overconfident than those in the choice condition. The following post hoc explanation is offered (in speculative spirit) to account for this latter difference. It seems likely that evidence concerning both alternatives was considered by subjects in both the choice and the uncued conditions. However, because choice was explicit in the former condition and implicit in the latter, participants in the uncued condition may have been less motivated at the time of confidence assessment to undertake a thorough search for information. As a result of this shallow search for information, in comparison to those in the choice condition, uncued subjects were less aware of how difficult the choice items actually were. This may be one reason why overconfidence was more extreme in the uncued condition than in the choice condition. Future research should be directed at obtaining more direct evidence of the cognitive processes responsible for the selection effects.

The consistency in the direction of the effect of selection across all three measures of confidence demonstrates the general consequences of the manipulation. The selection factor affects not only assessments based on consideration of a single knowledge item but also assessments based on the entire set of items. Thus it appears that the selection effects are not limited to the moment of probability assessment, but extend to subsequent reflection on overall task performance.

More difficult to interpret is the lack of correspondence in calibration between the mean subjective probability and estimated score measures of confidence. On the former measure, overconfidence typical of previous research was observed, while on the latter measure, subjects expressed considerable underconfidence about their general knowledge by estimating scores poorer than those that would be obtained by chance. These results seem to suggest that observed over/underconfidence depends considerably on elicitation procedures. However, part of the difference may be explained by scale differences between the two measures. Although below chance confidence estimates were possible for the estimated score measure, the probability scale was bounded by .5 so that below chance estimates were not possible. On the basis of the below-chance probabilities (and consequently lower overconfidence) observed by Ronis and Yates (1987) with their 0 to 1.0 probability scale, it is reasonable to assume that the difference in confidence levels between the probabilities and estimated scores would have been reduced had the scales been comparable in the present study.

But the over/underconfidence discrepancy between item probabilities and estimated scores in this study might not be completely attributable to a scale artifact. When subjects with estimated scores below chance were eliminated, the mean estimated score proportion was still significantly lower than the mean subjective probability. These data raise the question of whether high probabilities can be interpreted as general *over*confidence given that people accurately estimate or *under*estimate the number of their answers that are correct. We propose that confidence in a single item is determined in a different manner than confidence in a set of items. In the single item case, the relative amount of evidence in support of the alternatives is considered to be important. In contrast, confidence in one's performance over a set of items may depend on one's judgment of task difficulty, one's ability, the effort expended, etc. From this perspective, it is quite possible for one to have low confidence in one's knowledge of a task, but be highly confident about a particular item. In future research, an attempt should be made to compare subjective probabilities and estimated scores when both have more comparable scales to determine whether the appropriateness of confidence differs for individual items and the set of the same items as a whole. Particularly useful would be the use of methods such as verbal protocols to identify the different cognitive processes underlying confidence in the part vs confidence in the whole.

It is also interesting that the estimated percentile rank measure did not show generally high levels of overconfidence. Average confidence on this measure over four of the conditions was at the 50th percentile rank. The only significant finding was higher confidence, averaging at the 64th percentile rank, in the arbitrary cue condition compared to the other conditions. The estimated score and percentile rank data suggest that subjects were not highly confident in their knowledge on either an absolute or a relative basis.

EXPERIMENT 2

In this experiment, the investigation of framing and selective attention in confidence elicitation was completed by adding an *unchosen alternative-positive frame* condition. Instructions for confidence assessment in this condition were positively framed, but directed attention to the *unchosen* alternative, such that a complementary probability (i.e., $1 - p_i$, where p_i is the subjective probability that the *chosen* alternative is correct) was elicited. If framing and selective attention have no effect on confidence assessments, then this condition should produce probability assessments that are identical to (a) those with the negative frame for the chosen alternative, and (b) the complement of those with the positive frame for the chosen alternative. That is, the same level of overconfidence would be expected for all of these conditions.

Method

Participants. Participants ($n = 13$) from the same population described in Experiment 1 were randomly assigned to the *unchosen alternative-positive frame* condition.

Procedure. In the *unchosen alternative-positive frame* condition, the item instructions were identical to those for the *choice-positive frame condition,* with the following exceptions. The confidence assessment instructions were "Then indicate the probability that *the other alternative is correct* by circling one number between .00 (no chance of being correct) and .50 (a 50% chance of being correct)." A numerical scale prefaced by "The probability that *the uncircled alternative is correct* is" with values from .00 to .50 in increments of .10 followed.

The post-task questionnaire in the *unchosen alternative-positive frame* condition asked "Overall, how many of the 50 *uncircled* alternatives are *correct?* Estimate the total and write it in the blank."

Results and Discussion

For comparison purposes, the choice-positive frame and choice-negative frame conditions from Experiment 1 were included in the analyses with the unchosen alternative-positive frame condition. A one-way ANOVA with three levels was computed separately on six dependent measures: mean subjective probability, mean over/underconfidence, calibration score, resolution score, proportion correct, and estimated score. In addition, the subject's estimated percentile rank (relative to the other subjects' scores) was collected.

Significant complement effects were found on several dependent measures. There was a significant effect on the mean subjective probability measure ($F(2,61) = 3.35$, $p < .05$). Pairwise comparisons (Bonferroni t tests) of the differences in means were conducted. Mean subjective probability in the unchosen alternative-positive frame ($M = .72$, $SD = .071$) condition was significantly higher ($p < .05$) than both the choice-positive frame ($M = .67$, $SD = .051$) and choice-negative frame ($M = .67$, $SD = .077$) conditions. There was no difference between the positive and negative frame choice conditions.

A significant complement effect was also found for the mean over/underconfidence measure ($F(2,61) = 11.08$, $p < .001$). Bonferroni comparisons indicated that the unchosen alternative subjects (M = .21, SD = .099) were significantly ($p < .05$) more overconfident than those in the choice-positive frame (M = .09, SD = .077) and choice-negative frame (M = .10, SD = .073) conditions. Again, there was no significant difference between the positive and negative frame choice conditions. The same pattern of results was obtained on the calibration ($F(2,61) = 7.88$, $p < .001$) and proportion correct measures ($F(2,61) = 4.49$, $p < .05$). The unchosen alternative subjects were more poorly calibrated (M = .09, SD = .060) and had lower proportion correct scores (M = .51, SD = .088) than those in the choice-positive frame (M = .04 and .58, SD = .025 and .055) and choice-negative frame (M = .05 and .57, SD = .034 and .079) conditions (pairwise differences were significant at the .05 level). There were no significant differences between the three conditions on the resolution, estimated score, and estimated percentile rank measures.

In Experiment 1, we offered a post hoc explanation of our results stating that confidence was lower in the *choice* conditions than in the arbitrary cue conditions because judges in the latter conditions considered evidence regarding only one alternative. Additionally, we suggested that confidence in the choice conditions was reduced in relative proportion to the amount of available evidence (generated during the process of formulating a choice) in support of the unchosen alternative. Because subjects in the unchosen alternative-positive frame condition in Experiment 2 did choose between alternatives, evidence concerning both alternatives should have been generated prior to choice. Thus, by the above reasoning, confidence in the unchosen alternative should have been reduced just as in the choice condition. The results of Experiment 2 indicate that this was not the case, however. Judges expressed more extreme overconfidence in their choices by underestimating the probability that the unchosen alternatives were correct.

Before two possible explanations of this result are considered, it must be emphasized that this result was found on only one measure of confidence, the probability measure. The lack of significant differences on the other measures suggests that this effect is not as robust as the findings of Experiment 1. One explanation is that postchoice evaluation of the unchosen alternative in this condition differs from the postchoice evaluation of the chosen alternative in the choice conditions of Experiment 1. That is, when asked to assess the probability that the unchosen alternative is correct, subjects consider evidence regarding that option. However, because the other option has already been chosen, subsequent evaluation of information about the unchosen alternative may be biased toward *disconfirming* reasons. Furthermore, should the judge encounter evidence supporting the unchosen option in this postchoice search, she or he would then feel compelled to seek additional justification for the original choice.

As a result of these biased search processes, more extreme overconfidence occurs. Note that this biased search for information following choice is the result of asking individuals to consider the correctness of the unchosen option, and would not necessarily be brought on by eliciting the probability that one's chosen answer is wrong. Obviously, this explanation needs to be tested against competing explanations in future research.

Another plausible (though less theoretically interesting) explanation for these results is the "confusion hypothesis," which refers to the possibility that the unusual nature of the judgment the subjects were asked to make interfered with the normal choice process. Subjects were asked to choose an alternative, then assess the probability that the "wrong" answer (i.e., the answer they had rejected) was correct. This essentially required subjects to reverse the normal probability scaling or perform the appropriate transformation (1 − probability of one's own answer being correct) such that *lower* numbers represented greater confidence in their answer. Confusion about this reversal or an inability to perform the transformation may have degraded performance on the choice portion of the task as well as skewing the probability assessment. Note that this explanation is consistent with the Experiment 1 findings that manipulations which raised subjective probability also lowered proportion correct and calibration.

GENERAL DISCUSSION AND CONCLUSIONS

The general conceptual thesis posited here is that confidence becomes more appropriate as the amount of cognitive processing on a choice problem increases. It is important to note that in the present research, improvement in calibration was attributable to increases in both the appropriateness of subjective probabilities and the accuracy of choices. In circumstances in which choices were associated with less processing, calibration was significantly worse. These differences in amount of processing produced the unsettling result that those who were the most confident were also the least accurate. Nevertheless, these results are instructive, for they suggest that on choice tasks requiring the serious effort of decision makers, calibration is likely to be better than on less demanding tasks. This is consistent with the finding that experts generally show better calibration than nonexperts (e.g., Keren, 1987); the development of expertise in a particular decision context requires much cognitive effort, as well as the ability to recognize the need for further analysis on a given choice problem.

Support for our contention that overconfidence decreases as the amount of cognitive processing in choice increases is available from other research. In one study, Zakay (1985) demonstrated that postdecisional confidence assessments preceded by cognitively complex decision processes were significantly lower than confidence assessments preceded by simpler decision processes. Paese and Sniezek (in press) found that when an additional decision was made on the basis of judgment, overconfidence in the earlier judgment was reduced in comparison to conditions in which no additional decision was made. In the latter study, more extensive processing of the judgment problem was necessary to make the decision, and this processing appears to have been responsible for the reduction in overconfidence.

The present research also suggests that one means of increasing confidence in choice is to cue the decision maker to one of the possible alternatives. This idea is supported by the research of Ronis and Yates (1987) discussed earlier, as well as a study by Block and Harper (1987) in which each subject was anchored on a quantity that was either her own judgment or the judgment of another person. In the latter study, confidence in the accuracy of the anchor was inferred from the width of the interquartile range (above and below the anchor) specified by the judge, as well as the percentage of instances in which the true value actually fell in the specified range. Although such assessments of uncertainty about judgments may result from processes different than the assessment of confidence in correctness of choice (Peterson & Pitz, 1988), the results of this study are similar to those of Experiment 1. Subjects who were given the judgments of another person as anchors were significantly more overconfident (i.e., specified narrower ranges that contained the true values less often) than those who provided their own anchors. Like the precircled alternatives in Experiment 1, the judgments of another person served as arbitrary cues in the Block and Harper study. We contend that this cueing reduced the amount of processing devoted to each judgment, thereby contributing to overconfidence. Interestingly, these results suggest that those who are cued to a particular decision are likely to be more overconfident about the correctness or incorrectness of the decision than the actual decision maker.

The results of the arbitrary cueing condition in Experiment 1 have some indirect implications for confidence assessments in social settings. Given the high confidence and high overconfidence found with arbitrary cueing, it follows that confidence assessments formed in social settings may be susceptible to greater levels of overconfidence than independently formed individual confidence assessments. Given that Ronis and Yates (1987) found that people are more likely to "choose" a (randomly) cued alternative, we can expect that there will be a general tendency for person A's choice to subsequently be favored over alternatives by a person B who has not previously participated in the choice process. Additionally, B is likely to be even more overconfident about the alternative previously chosen by A than was A. This would occur not because of conformity effects, but simply because of reduced cognitive processing resulting from the cueing. In social settings (e.g., groups) in which prior independent judgments have not been made by all individual persons (Sniezek & Henry, 1990), this might be a particularly serious problem leading to poor performance but high confidence. Another sort of problem would occur on those (relatively fewer) occasions in which B chose a different alternative than A originally chose. Here too, B would be even more confident than A, but the result would be an increase in cognitive conflict in comparison to a choice situation without cueing. Therefore, if it is important to avoid overconfidence, each evaluator should independently repeat the choice process without cues regarding other persons' choices. Clearly, the validity of these speculations should be tested in future research.

REFERENCES

Block, R. A., & Harper, D. R. (1987). *Explicit anchoring reduces overconfidence in estimation*. Seattle: Psychonomic Society.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford Univ. Press.

Fischhoff, B., & MacGregor, D. (1983). Judged lethality: How much people seem to know depends upon how they are asked. *Risk Analysis*, 3, 229–236.

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.

Hogarth, R. M. (1982). *Question framing and response consistency*. San Francisco: Jossey-Bass.

Keren, G. (1987). Facing uncertainty in the game of bridge: A calibration study. *Organizational Behavior and Human Decision Processes*, 39, 98–114.

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

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how 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*. Cambridge: Cambridge Univ. Press.

May, R. S. (1986). Overconfidence as a result of incomplete and wrong knowledge. In R. W. Scholz (Ed.), *Current Issues in West German Decision Research* (pp. 13–30). New York: Lang.

Paese, P. W., & Sniezek, J. A. (in press). Influences on the appropriateness of confidence in judgment: Practice, effort, information, and decision making. *Organizational Behavior and Human Decision Processes*.

Peterson, D. K., & Pitz, G. F. (1988). Confidence, uncertainty, and use of information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 85–92.

Ronis, D. L., & Yates, J. F. (1987). Components of probability judgment accuracy: Individual consistency and effects of subject matter and assessment method. *Organizational Behavior and Human Decision Processes*, 40, 193–218.

Sniezek, J. A., & Henry, R. (1989). Accuracy and confidence in group judgment. *Organizational Behavior and Human Decision Processes*, 42, 1–28.

Sniezek, J. A., & Henry, R. (1990). Revision, weighting, and commitment in consensus group judgment. *Organizational Behavior and Human Decision Processes*, **45**, 66–84.

Webster's new twentieth century dictionary. (2nd ed.) (1979). Williams Collins Publishers.

Wright, G. (1982). Changes in realism of probability assessments as a function of question type. *Acta Psychologica*, **52**, 165–174.

Zakay, D. (1985). Post-decisional confidence and conflict experienced in a choice process. *Acta Psychologica*, **58**, 75–80.

RECEIVED: August 9, 1988