ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES 57, 26–37 (1994)

Computer-Assisted Decision Making: Performance, Beliefs, and the Illusion of Control

JEFFREY E. KOTTEMANN

The University of Michigan

FRED D. DAVIS

the University of Minnesota

AND

WILLIAM E. REMUS

The University of Hawaii

Recent empirical evidence indicates that computer-assisted what-if analysis does not predictably improve decision making. Why then is what-if analysis so widely used by decision makers? We argue that what-if analysis creates an "illusion of control" which leads decision makers to overestimate its effectiveness. A between-subjects experiment was conducted using a production planning task to test this conjecture. As hypothesized, subjects falsely believed that what-if analysis improved their decision making. In fact, what-if analysis users expressed inflated confidence beliefs yet post hoc analysis revealed that they actually performed significantly worse than nonusers in the trials immediately preceding belief measurement. In light of other research linking user acceptance of computer-based technologies to users' performance perceptions, these results forewarn of sustained, but dysfunctional, use of what-if analysis.

What-if analysis allows decision makers to manipulate parameters of decision models in an attempt to make better decisions. For example, investment bankers routinely use what-if analysis to simulate the manipulation of factors such as future interest rates and portfolio diversifications in order to formulate investment strategies. Despite the popularity of what-if analysis, research comparing what-if analysis to unaided decision making has shown that what-if analysis sometimes generates positive performance effects (e.g., Benbasat & Dexter, 1982; Sharda, Barr, & McDonnell, 1988), sometimes no effects (e.g., Goslar, Green, & Hughes,

All authors contributed equally to this research. We are grateful to Izak Benbasat, Dan Stone, Ed Miller, and the reviewing editors for their helpful suggestions, and to the Division of Research, School of Business Administration, and University of Michigan for financial support. Address correspondence and reprint requests to Jeffrey E. Kottemann School of Business Administration, The University of Michigan, Ann Arbor, MI 48109-1234.

0749-5978/94 $6.00

1986; Fripp, 1985), and sometimes negative effects (e.g., Kottemann & Remus, 1987, 1991). Why is what-if analysis so widely used when research indicates only mixed effectiveness? Do decision makers overestimate the effectiveness of what-if analysis? Answers to this question are particularly important because users' impressions of performance effects are a major determinant of their adoption behavior (Davis, 1989; Davis, Bagozzi, & Warshaw, 1989). The present study assesses whether people properly appraise their predictive powers and their decision performance when using what-if analysis. Our hypothesis is that they do not.

Numerous studies have investigated factors that may unduly inflate confidence. Langer's (1975; Langer & Roth, 1975) early work on the illusion of control demonstrated that overconfidence occurs when factors ordinarily associated with improved performance in skilled situations are present in situations at least partly governed by chance—as is the case of the production scheduling task used in the present experiment. Langer found that choice, stimulus/response familiarity, competition, and active involvement can all lead to inflated confidence beliefs. Of these factors, active involvement is intrinsic to what-if analysis. Langer (1975) defined active involvement as "exerting effort while actively engaged in the task" (p. 313). In her study, Langer (1975, expt 4) investigated active involvement/effort by varying the degree of direct control that subjects' could exercise over a device—a device which provided random success feedback. Subjects that directly manipulated the device expressed higher confidence in their performance. Langer's (1975) results suggest that people who gain the impression that they are manipulating aspects of the task and exerting more effort form inflated performance beliefs (see also Paese & Sniezek, 1991; Yates & Kulick, 1977). Since what-if analysis allows decision makers to actively manipulate a model of a decision problem, we expected subjects using what-if analysis to expend more effort (which was measured by time spent on the task) and to form inflated confidence beliefs. When manipulated by the user, what-if analysis tools calculate and display simulated outcomes. This additional information can also contribute to inflated confidence beliefs (see Oskamp, 1965; Paese & Sniezek, 1991). Involvement/effort and information are factors associated with performance on skilled tasks, and, consistent with the illusion of control principle, they can inflate confidence beliefs. Given that these same factors are intrinsic to what-if analysis, we hypothesize that inflated confidence beliefs will result from its use.

The present experiment manipulates (between-subjects) the availability of what-if analysis using a production scheduling task. Because pretesting revealed that what-if analysis does not improve performance for this task (see also Kottemann & Remus, 1987, 1991), we did not expect an improvement in actual performance. However, we did expect what-if analysis to inflate confidence beliefs because it allows users to manipulate variables in a decision model—not only decision variables (such as the amount of a product to produce) but also uncontrollable variables (such as future environmental conditions)—and presents users with information in the form of simulated outcomes. Specifically, we hypothesize that what-if analysis will lead subjects to (a) believe they can better predict uncontrollable variables and (b) express higher confidence in their performance.

METHOD

The Production Scheduling Task

The experimental task is based on the Holt, Modigliani, Muth, and Simon (1956 and 1960) formulation of the production scheduling problem. In this formulation, the decision maker faces *uncertain demand* over a series of decision periods and must determine production levels and workforce levels with the objective of minimizing cumulative total costs. The cost function for a given new period is the sum of three quadratic cost components:

Workforce level change cost =
$64.3\ ((\text{current workforce} - \text{new workforce})^2)$,
Worker overtime/idletime cost =
$0.8\ (((\text{new workforce} * 5.67) - \text{new production})^2)$
Cost for Nonoptimal inventory =
$0.02\ ((\text{current inventory} + \text{new production} - \text{new demand} - 320)^2)$.

Note that the cost components are highly interrelated and that one cost component can be avoided only at the expense of another. For example, if the workforce is set artificially low relative to production, then higher worker overtime cost is incurred. If, in turn, production is set artificially low relative to demand, then inventory outage cost is driven higher. In effect, then, the decision maker's task is to determine how to allocate costs among the three cost components. Our pretests with this task found that what-if users do not outperform unaided decision makers.

After 4 practice periods, subjects made production and workforce decisions for each of 24 periods. All subjects received the following information for each period: the current workforce level, the current inventory level, and demand forecasts for each of the next 3 periods. The demand streams used in the experiment were generated as follows. Demand began at 2500 units and increased at a rate of 100 units per period. This "unadjusted" demand was then randomized ±200 units to generate the *actual demand*. Subjects' social security numbers were used to seed the random number generator (and so each subject received a different demand stream). Subjects were not presented with the actual future demands, but were given uncertain demand forecasts for the next three upcoming periods. Forecasts were generated by randomizing the actual demands ±100 units.

The production scheduling task is representative of the large class of recurring decisions under uncertainty. While recognizing that differences do exist, the basic nature of the task is present in other domains, such as maintaining an investment portfolio over time, periodic planning, budgeting, and medical diagnosis (see Hogarth, 1981; Kleinmuntz, 1985; Kleinmuntz & Thomas, 1987).

The Decision Aid

The experiment was a between-subjects design with two treatments: one group had computer-assisted what-if analysis available while the other did not (termed the what-if and non-what-if groups, respectively). Figure 1 shows the computer screen for the non-what-if treatment. The lower lines on the screen display the worker productivity index and the optimal level at which to maintain inventory. Both these factors remained constant throughout the task. The upper right corner displays the current workforce and inventory levels. The upper left displays the average total

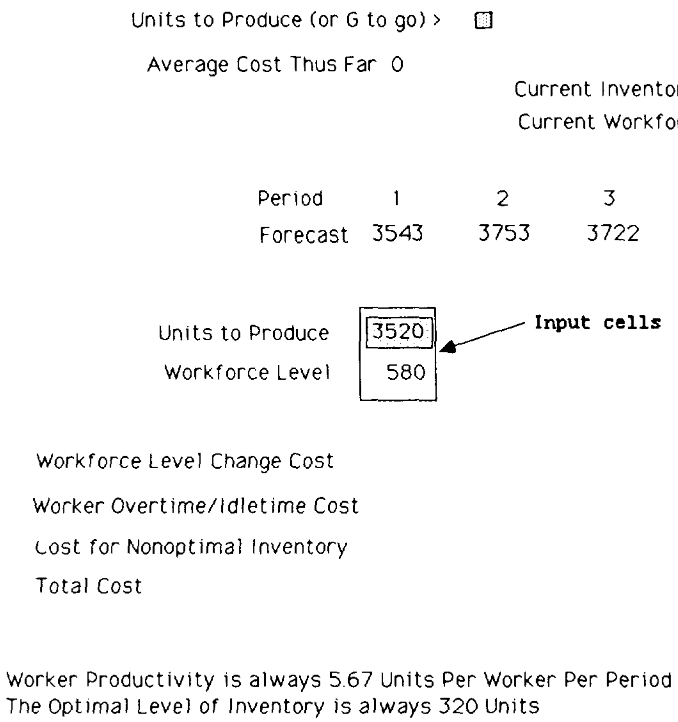

FIG. 1. The non-what-if treatment's input screen.

cost thus far incurred. The upper middle of the screen displays the forecasts generated for the next three upcoming periods. The middle of the screen was used to accept subjects' production and workforce level decisions.

The what-if treatment group received, in addition to the above information, a what-if analysis capability. Figure 2 shows the what-if treatment computer screen. With the what-if analysis capability, subjects were able to simulate any number of scenarios per period to assess cost ramifications. They could enter levels for anticipated sales demand, workforce, and production levels and have the system calculate the resulting costs. It is important to note that the what-if analysis capability did not provide improved demand forecasts and therefore should not be expected to improve subjects' predictive abilities. But, since it does allow subjects to manipulate anticipated demand, we expect what-if subjects to express the belief that they can better predict future demand.

Subjects entered a "go" command whenever they were ready to process the decision they had entered. The system then performed checks for gross numeric range errors (e.g., producing 20,000 units with a workforce

Anticipated Sales (or G to go, S to Simulate) >S

Average Cost Thus Far 0

Current Inventory 300
Current Workforce 500

	1	2	3	
Period	1	2	3	
Forecast	3335	3251	4063	
Anticipated Sales	3300	3250	4000	Input cells
Units to Produce	3320	3300	4100	
Workforce Level	510	515	525	
Ending Inventory	320	370	470	
Workforce Level Change Cost	6430	1608	6430	
Worker Overtime/Idletime Cost	36688	28872	252338	Simulation Result cells
Cost for Nonoptimal Inventory	0	206	1856	
Total Cost	43118	30686	260624	

Worker Productivity is always 5.67 Units Per Worker Per Period
The Optimal Level of Inventory is always 320 Units

FIG. 2. The what-if treatment's screen.

of 0) and displayed error messages as appropriate. In any event, users were asked to confirm the go command before the decision was processed. Upon confirmation, the results were computed and displayed. As shown in Fig. 3, the display of results included the actual demand for the period, the new inventory and workforce levels, workforce change cost, worker over/idle time cost, inventory over/outage cost, total cost for the period, and the updated average total cost thus far incurred. During execution of the task, the system automatically recorded subjects' decisions, resulting costs, and the elapsed time spent on the task.

Subjects and Procedure

The 26 subjects were MBA student volunteers from an introductory course in Operations Research. They were randomly assigned to treatments. All subjects had end-user computing experience including the use of a spread-sheet package. A week prior to the experiment, one of the experimenters introduced the entire group of subjects to the production scheduling task. They were told that their objective was to minimize

ENTER RETURN TO PROCEED TO THE NEXT DECISION PERIOD

Average Cost Thus Far 85846

Current Inventory 245

Current Workforce 510

Actual Sales 3575

Period	1	2	3
Forecast	3543	3753	3722

Anticipated Sales	$\boxed{3600}$
Units to Produce	3520
Workforce Level	510

Workforce Level Change Cost	6430
Worker Overtime/Idletime Cost	78952
Cost for Nonoptimal Inventory	464
Total Cost	85846

Worker Productivity is always 5.67 Units Per Worker Per Period
The Optimal Level of Inventory is always 320 Units

FIG. 3. After the decision has been made and processed (screen for both treatments).

long-run total cost. None of the lecture material pertained to what-if analysis.

The subjects arrived at the computing lab every 10 min. They were directed to a terminal, and one of the experimenters reintroduced them individually to the production scheduling task and instructed them in the use of the specific software to be used. These orientation sessions were based upon pretested scripts. Subjects were given four practice periods and then performed the 24 "real" periods. Subjects were told that they could take as long as they wanted to complete the task; they took from 45 to 150 min. They were not allowed to interact. Also, given the staggered arrival times and seating arrangements, it was difficult for subjects to compare how long they were spending on the task relative to others. The subjects appeared highly motivated (e.g., subjects were eager to find out how they did).

After completing the task, subjects were asked to fill out a brief questionnaire. One item on the questionnaire pertained to subjects' beliefs about their ability to accurately predict future demand: "I find that I can predict actual demand very accurately." The response was on a 5-point Likert-type scale: Strongly Agree, Agree, Uncertain, Disagree, Strongly Disagree. A second item pertained to subjects' beliefs about their performance: "What percentage of other decision makers, who are also undertaking the production scheduling exercise, do you feel are performing *poorer than you.*" This question was worded in an attempt to encourage subjects to think about how hard the task is for others, which Allwood and Montgomery (1987) suggest might improve the veridicality of confidence judgments. Also note that subjects were not made aware of the differing treatments, so presumably their responses were made relative to the other subjects in the same treatment. To reduce hypothesis guessing, filler items were included asking subjects to judge how challenging and how difficult they found the production planning task to be (no significant differences were found between groups for these filler items).

RESULTS

Recall that use of what-if analysis was not expected to improve subjects' cost performance, but it was expected to be associated with inflated beliefs regarding predictive ability and performance. As shown in Table 1, there is no significant difference in cost performance between the what-if and non-what-if groups ($t(24) = 0.506$, ns)—although the non-what-if group does have a lower mean cost. However, as expected, what-if subjects expressed the belief that they were better able to predict future demand ($t(24) = 2.091$, $p < .05$) and that they performed better than a majority of other subjects ($t(24) = 2.210$, $p < .05$). Consistent with an

TABLE 1
THE EFFECTS OF WHAT-IF ANALYSIS ON ACTUAL PERFORMANCE AND PERFORMANCE BELIEFS

Unaided	Aided	Significance
	Actual (cost) performance	
117,453	153,295	ns
	Expressed beliefs	
I can predict future demand very accurately (1–5 scale; 1 is strongly agree, 3 is neutral)		
3.4	2.5	$p < .05$
	Percentage of others performing poorer than I	
45	58	$p < .05$

increase in active involvement and effort, what-if subjects spent significantly longer performing the task ($t(24) = 2.295$, $p < .05$).

To get a sense of how well-calibrated subjects' performance beliefs were, we analyzed the correlations of actual performance with the two types of beliefs (the variables were coded such that positive correlation coefficients reflect performance beliefs that are in the right direction). For the non-what-if group the correlations of actual performance with perceived predictive ability and confidence were .46 ($p = .06$) and .60 ($p = .015$), respectively. This suggests that the confidence beliefs of non-what-if subjects were veridical. Correlations for the what-if group were .04 ($p = .44$) and .11 ($p = .36$), which are nonsignificant. While this seems to suggest that non-what-if subjects were better calibrated, the between-group differences in these correlation coefficients are not significant ($z = 1.01$, ns; $z = 1.30$, ns) due to the small sample size (13 per cell).

While the results are consistent with an illusion of control argument, it is possible that the difference in expressed beliefs may have occurred because what-if subjects might have performed significantly better in the later decision periods—the periods just prior to belief measurement. To explore this possibility, a post hoc analysis was performed to compare performance in the final six decision periods. There was a performance difference in this case, but surprisingly, the what-if subjects actually performed significantly worse than the non-what-if subjects in the final six decision periods ($p < .001$).

DISCUSSION

Langer's (1975) early work has shown that the illusion of control is operative in a wide variety of task contexts when factors associated with skilled performance, such as active involvement and effort, are introduced into situations at least partly governed by chance. The present experiment demonstrates that modern, popular decision technologies may increase decision makers' effort and active involvement, engendering the illusion of control and leading to inflated performance beliefs. One obvious, and disturbing, implication is that popular technologies such as what-if analysis may be popular independent of actual task performance effects.

The present study is the first we know of that explores the potential influence of illusion of control mechanisms in the context of decision assistance technologies (especially what-if analysis), and so the results must be interpreted with caution. For example, the present study did not orthogonally manipulate effort and information. Future research is needed to determine the relative mediating effects of these two factors. On the other hand, they are so interwoven in practice—processing more information takes more effort—that it may be difficult or unrepresentative to manipulate them orthogonally, particularly in the context of what-if analysis (since both factors are intrinsic to what-if analysis). In a different context, Paese and Sniezek (1991, Expt 1) did attempt to manipulate information and effort orthogonally using a baseball player performance prediction task. However, "manipulation checks indicated that the effort manipulation did not have the desired effect (nor did the effort manipulation have any significant effects on the dependent measures)" (p. 107). Nevertheless, future research should attempt to determine the relative effects of these two factors in general as well as in the context of what-if analysis.

While the present results are consistent with the hypothesis that effort and information are responsible for the inflated confidence beliefs of what-if analysis users, an alternative interpretation is that people believe that computerized what-if analysis is beneficial because they have a predisposition to believe that computer-based tools are helpful. Although the present study is unable to rule out this interpretation, evidence from a recent study (Davis & Kottemann, 1991) casts doubt on it. In that study, instead of a computer-aided vs unaided comparison, we contrasted two different computer aids within-subjects: a what-if tool of the kind studied in the current paper and a noninteractive quantitative model.[1] While both tools were computerized, and subjects had the opportunity to use either tool at any time they chose, subjects significantly overestimated the value of the what-if tool relative to the noninteractive model even though use of the noninteractive model would have led to significantly better decision performance (and less effort). This was true of subjects who were told

[1] The noninteractive quantitative model was based on a simple exponential-smoothing heuristic. It determines the amount to produce by (.5)forecast^{t} + (.3)forecast^{t+1} + (.2)forecast^{t+2} − inventory + safety stock, and then determines the required workforce level by dividing the production level by the number of units produced by each worker (5.67). Subjects were thoroughly introduced to the rationale behind the heuristic.

that "the noninteractive model provides good but rarely optimal decisions" as well as those who were told "the noninteractive model has historically outperformed 90% of your peers" (c.f. Powell, 1991). Since both aids were computer-based, the conjecture that the inflated confidence beliefs associated with what-if analysis was due to the assumption that computers are supposed to help people tends to be ruled out. On the other hand, these results are consistent with an illusion of control argument: Whereas supplying subjects with a what-if analysis tool may engender an illusion of control, asking subjects to use the advice of a noninteractive model may well engender an "impression of *non*control."

One implication of the effect of increased effort and involvement on user acceptance of decision assistance technologies is that decision makers may unjustly avoid decision technologies that force them to follow predefined decision procedures. For example, over the last two decades, computer-assisted decision technologies have been developed and refined for multicriteria decision making (MCDM) based in the disciplines of multiattribute utility theory and mathematical programming. The objective of such systems is to help decision makers determine their most preferred alternative from among a set of alternatives. Most of the available MCDM methods require users to make a series of trade-off judgments between individual decision criteria, after which the system selects the "most preferred" alternative. Other methods generate sets of nondominated alternatives for the decision maker and then ask the decision maker to *choose* from the set. While the former type of approach is typically deemed more thorough by professional decision analysts—indeed, Zionts (1985) points out that the latter type of approach may actually remove the most preferred alternative from consideration—the latter type is typically preferred by subjects (Kottemann & Davis, 1991). One explanation for subjects' preferences is that the latter, choice-oriented methods give subjects a sense of control while the trade-off elicitation methods do not. This is wholly consistent with Langer's (1975) experiment 2, in which allowing subjects to choose their lottery tickets engendered an illusion of control (also see Ronis & Yates, 1987; Sneizek, Paese, & Switzer, 1990). Strikingly similar results can be found in the long history of decision rules and their lack of acceptance (for a thorough review see Kleinmuntz, 1990), and more recently, with techniques that ask decision makers to follow fixed procedures for problem formulation (Abualsamh, Carlin, & McDaniel, 1990; Cats-Baril & Huber, 1987).

Results of the present study are consistent with the hypothesis that what-if analysis leads to inflated confidence judgments as well as inflated beliefs about how accurately uncertain future events can be predicted. In light of other research linking user acceptance of computer-based tools to users' perceptions of how much the tool improves performance (e.g., Davis, 1989; Davis, *et al.*, 1989), the present results imply that decision makers may willingly, but unknowingly, continue using what-if analysis despite lack of actual performance advantages. Future research is clearly needed to examine the robustness of these results and to determine the relative effects of mediating factors. In general, identifying the psychological mechanisms that underlie the formation of beliefs regarding the effectiveness of computer-assisted decision technologies, and determining the conditions under which such beliefs are and are not veridical, is an exciting area for future research.

REFERENCES

Abualsamh, R. A., Carlin, B., & McDaniel, R. R., Jr. (1990). Problem structuring heuristics in strategic decision making. *Organizational Behavior and Human Decision Processes*, **45**, 159–174.

Allwood, C. M., & Montgomery, H. (1987). Response selection strategies and realism of confidence judgments. *Organizational Behavior and Human Decision Processes*, **39**, 365–383.

Benbasat, I., & Dexter, A. S. (1982). Individual differences in the use of decision support aids. *Journal of Accounting Research*, **20**(1), 1–11.

Cats-Baril, W. L., & Huber, G. P. (1987). Decision support systems for ill-structured problems: An empirical study. *Decision Sciences*, **18**, 350–372.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information Systems Quarterly*, 319–340.

Davis, F. D., Bagozzi, R., & Warshaw, P. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, **35**, 982–1003.

Davis, F. D., & Kottemann, J. E. (1991). Decision support systems and the illusion of control. Division of Research Working Paper, Graduate School of Business Administration, University of Michigan.

Einhorn, H. J. (1980). Learning from experience and suboptimal rules in decision making. In T. S. Wallsten, (Ed.), *Cognitive processes in choice and decision behavior* Hillsdale, NJ: Erlbaum.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, **85**, 395–416.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, **32**, 53–88.

Fripp, J. (1985). How effective are models? *International Journal of Management Science (Omega)*, **13**(1), 19–28.

Goslar, M., Green, G., & Hughes, T. (1986). Decision support systems: An empirical assessment for decision making. *Decision Sciences*, **17**(1), 79–91.

Hogarth, R. M. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin*, **90**, 197–217.

Hogarth, R. M., & Makridakis, S. (1981). Forecasting and planning: An evaluation. *Management Science*, **27**, 115–138.

Holt, C. C., Modigliani, F., & Muth, J. F. (1956). Derivation of a linear decision rule for production and employment. *Management Science*, **2**(2), 159–177.

Holt, C. C., Modigliani, F., Muth, J. F., & Simon, H. A. (1960). *Planning production, inventories, and work force*. Englewood Cliffs, NJ: Prentice–Hall.

Johnson, E. J., & Payne, J. W. (1985). Effort and accuracy in choice. *Management Science*, **31**, 395–414.

Kleinmuntz, B. (1990). Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin*, **107**, 296–310.

Kleinmuntz, D. N. (1985). Cognitive heuristics and feedback in a dynamic decision environment. *Management Science*, **31**, 680–702.

Kleinmuntz, D. N., & Thomas, J. B. (1987). The value of action and inference in dynamic decision making. *Organizational Behavior and Human Decision Processes*, **39**, 341–364.

Kottemann, J. E., & Davis, F. D. (1991). Decisional conflict and user acceptance of multicriteria decision making aids. *Decision Sciences*, **22**, 918–926.

Kottemann, J. E., & Remus, W. E. (1987). Evidence and principles of functional and dysfunctional decision support systems. *International Journal of Management Science (Omega)*, **15**(2), 135–144.

Kottemann, J. E., & Remus, W. E. (1991). The effects of decision support systems on performance. In H. G. Sol & J. Vecsenyi (Eds.), *Environments for supporting decision processes* (pp. 203–214). New York: North-Holland.

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, **32**, 311–328.

Langer, E. J., & Roth, J. (1975). Heads I win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task. *Journal of Personality and Social Psychology*, **32**(6), 951–955.

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology*, **29**, 261–265.

Paese, P. W., & Sniezek, J. A. (1991). Influences on the appropriateness of confidence in judgment: Practice, effort, information, and decision-making. *Organizational Behavior and Human Decision Processes*, **48**, 100–130.

Powell, J. L. (1991). An attempt at increasing decision rule use in a judgment task. *Organizational Behavior and Human Decision Processes*, **48**, 89–99.

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.

Sharda, R., Barr, S. H., & McDonnell, J. C. (1988). Decision support system effectiveness: A review and an empirical test. *Management Science*, **34**(2), 139–159.

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

Yates, J. F., & Kulick, R. M. (1977). Effort control and judgments. *Organizational Behavior and Human Performance*, **20**, 54–65.

Zionts, S. (1985). Multiple criteria mathematical programming: An overview and several approaches. In G. Fandel and J. Spronk (Eds.), *Multiple criteria decision methods and applications* (pp. 85–128). Heidelberg: Springer–Verlag.

RECEIVED: May 21, 1991