AD-A239 708

ARI Research Note 91-78

Examining the Effect of Information Order on Expert Judgment

Leonard Adelman, Martin A. Tolcott, and Terry A. Bresnick
Decision Science Consortium, Inc.

for

Contracting Officer's Representative
Michael Drillings

Office of Basic Research
Michael Kaplan, Director

June 1991

United States Army
Research Institute for the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Research accomplished under contract
for the Department of the Army

Decision Science Consortium, Inc.

Technical review by

Michael Drillings

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS --
2a. SECURITY CLASSIFICATION AUTHORITY -- 2b. DECLASSIFICATION/DOWNGRADING SCHEDULE --	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) 90-6	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Research Note 91-78
6a. NAME OF PERFORMING ORGANIZATION Decision Science Consortium, Inc. 6b. OFFICE SYMBOL (If applicable) --	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute Office of Basic Research
6c. ADDRESS (City, State, and ZIP Code) 1895 Preston White Drive, Suite 300 Reston, VA 22091	7b. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences 8b. OFFICE SYMBOL (If applicable) PERI-BR	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903-89-C-0134
8c. ADDRESS (City, State, and ZIP Code) Office of Basic Research 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102B PROJECT NO. 74F TASK NO. N/A WORK UNIT ACCESSION NO. N/A

11. TITLE (Include Security Classification)
Examining the Effect of Information Order on Expert Judgment

12. PERSONAL AUTHOR(S)
Adelman, Leonard; Tolcott, Martin A.; and Bresnick, Terry A.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim	FROM 89/05 TO 90/05	1991, June	

16. SUPPLEMENTARY NOTATION
Michael Drillings, Contracting Officer's Representative

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Expert judgment; Judgment heuristics and biases; Order effects; Army air defense; Hypothesis testing

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

For this report, researchers performed an experiment in which trained Army air defense personnel created a paper-and-pencil representation of their task to test the predictions of the Einhorn-Hogarth model for belief updating. For a simple task with a short series of information, the model predicted a significant information order by response mode interaction. The experiment supported the model's predictions. When information was presented sequentially and a probability estimate was obtained after each piece of information, the order in which the same information was presented significantly affected the final mean probability estimates. In contrast, when all the information was presented at once and a probability estimate was obtained at that time, information order had no effect. There were, however, significant individual differences. Moreover, order and response mode cumulatively accounted for less than 10 percent of the total variation in the participants' probability estimates. These findings, and their implications, are discussed in the report.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Michael Drillings	22b. TELEPHONE (Include Area Code) (703) 274-8722 22c. OFFICE SYMBOL PERI-BR

DD Form 1473, JUN 86 — Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

ACKNOWLEDGMENTS

The research described herein was supported by U.S. Army Research Institute for the Behavioral and Social Sciences (ARI), Contract No. MDA903-89-C-0134 to Decision Science Consortium, Inc. (DSC). We would like to acknowledge the effort of the many people who contributed to this endeavor. In particular, we would like to thank Jim Chinnis, Julia Doherty, Diane Laaksonen, Kathryn Laskey, John Leddo, Freeman Marvin, and Lynn Merchant-Geuder at DSC; Laurel Allender, Michael Drillings, Mark Leavitt, Michelle Saks, and Michael Strub at ARI; and the Army air defense personnel at Fort Bliss, Texas, who supported or participated in the study. The views, opinions, and findings contained in this report are those of the authors and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation.

EXAMINING THE EFFECT OF INFORMATION ORDER ON EXPERT JUDGMENT

EXECUTIVE SUMMARY

Requirement:

Substantial research indicates that humans use heuristics to make inferences and that, depending on task characteristics, these heuristics can lead to inconsistencies and errors in judgment—that is, cognitive biases. Most of this research has been performed with university students performing tasks requiring logical thinking but not expertise in a particular substantive area. Our concern is in determining whether heuristics can lead to cognitive biases among experienced personnel performing their substantive task. In particular, we examined whether information order and response mode could affect the judgments of Army air defense operators.

Procedure:

A within-subject factorial experiment was performed in December, 1989, with 63 Army air defense operators stationed at Fort Bliss, Texas. Five factors were varied using a paper-and-pencil format: (1) whether the information about an incoming aircraft of unknown identity was presented sequentially or all at once, (2) the order with which the same information was presented, (3) whether the first piece of information supported the conclusion that the aircraft was friendly or hostile, (4) whether the first piece of information was strong or weak, and (5) whether subsequent information negating the first piece of information was strong or weak. The dependent variable was a probability estimate for the identity of the aircraft (friend or hostile) after receiving all information.

Findings:

Information order and response mode interacted to affect the Army air defense operators' judgments. When information was presented sequentially and a probability estimate was obtained after each piece of information, participants gave different probability estimates of whether an unknown aircraft was friendly or hostile, depending on the order with which the same information was presented. In contrast, there was no order effect when the information was presented all at once. These results support the predictions of the Hogarth-Einhorn belief updating model. There were, however, large individual differences.

Utilization of Findings:

Discussion with senior-level Army air defense personnel indicates that the results may have implications for (1) training air defense operators, and (2) designing information displays for future air defense systems. The case would be stronger if the results were obtained using (1)

actual simulators, and (ii) if we had investigated whether changes in identification judgments also affected engagement behavior. Preparation of such a study is currently underway.

EXAMINING THE EFFECT OF INFORMATION ORDER ON EXPERT JUDGMENT

CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXAMINING THE EFFECT OF INFORMATION ORDER ON EXPERT JUDGMENT

INTRODUCTION

Substantial research reviewed in Hogarth (1987), Kahneman et al. (1982), Nisbett and Ross (1980), etc. indicates that humans use heuristics to make inferences and that, depending on task characteristics, these heuristics can lead to inconsistencies and errors in judgment; that is, cognitive biases. Although there are exceptions (e.g., see Fischhoff et al., 1978; Tversky and Kahneman, 1971), most of this research has been performed with university students performing tasks requiring logical thinking but not expertise in a particular substantive area. Our concern is in determining whether heuristics can lead to cognitive biases among experienced personnel performing their substantive task. In particular, the experiment reported herein examined whether information order and response mode could affect the judgments of Army air defense operators.

Einhorn and Hogarth (1987) and Hogarth and Einhorn (1989) developed a belief updating model attempting to explain the diverse findings in the literature regarding the effect of information order. This model assumes an anchoring and adjustment process that depends on four task characteristics: the amount of information presented, whether the information is simple or complex, the order in which the information is presented, and whether a probability estimate is obtained after presenting each piece or all the information. Since the experiment reported herein involved only a short series of information and was simple for trained personnel, we consider only their model's prediction for information order and response mode.

Specifically, when information is presented sequentially and a probability estimate is obtained after each piece of information, the Hogarth-Einhorn model predicts that people will anchor on the current position and adjust it on the basis of the strength and direction (positive or negative) of each new piece of information. Since each new piece of information creates a new anchor, recent information is weighted more than prior information, thereby resulting in a recency effect. Moreover, they hypothesized that the larger the anchor, the greater the impact of the same piece of negative evidence. Conversely, the smaller the anchor, the greater the impact of the same piece of positive evidence. Consequently, the order in which the same positive and negative evidence is presented is predicted to result in different final probability estimates.

In contrast, when information is presented all at once and a probability estimate is obtained at that time, the Hogarth-Einhorn model predicts that people will anchor on the piece of information presented first and adjust it on the basis of the aggregate impact of all subsequent information in support of the initial information (i.e., positive evidence) and against it (i.e., negative evidence). Since the model predicts that subsequent information will be considered in the aggregate, the order of positive and negative evidence is predicted to have no effect on the final probability estimates.

Thus far, there is minimal empirical support for this theoretical model using trained personnel performing their substantive task. Although Einhorn and Hogarth (1987) and Hogarth and Einhorn (1989) present empirical results supporting the model, all the experiments were with university students performing general tasks—not trained personnel performing their substantive task. This is also true of most of the studies they reviewed, as well as others (e.g., Hamm, 1987) supporting some of their predictions. Asare (1990), Ashton and Ashton (1988; 1990), and Meisser (1990) do, however, present empirical results consistent with the model's predictions using professional auditors. Serfaty et al. (1989) also present results obtained from Army officers showing an order effect consistent with the model's predictions.

There is also research with trained personnel predicting results different from the Hogarth-Einhorn model. In a series of experiments where information was presented sequentially, Tolcott et al. (1988, 1989a, 1989b) showed that Army intelligence analysts exhibited a "confirmation bias," whereby they gave greater weight to positive evidence confirming their initial hypothesis about the enemy's course of action than to negative evidence disconfirming their early judgment. An anchoring and adjustment model was inferred by the researchers to explain the findings, but one where (1) the initial information was overweighted, and (2) subsequent information was given less weight the later it was received. Consequently, that research would predict a primacy effect when information is presented sequentially, not the recency effect predicted by the Hogarth-Einhorn model.

It should be noted, however, that the procedures used in the studies by Tolcott et al. differed from those used by Hogarth and Einhorn in three important ways. First, the initial information was embedded in complex, ambiguous scenarios in the Tolcott et al. studies; straightforward, one-paragraph scenarios ranging between 68 to 109 words were used in the Hogarth-Einhorn scenarios. Second, in the Tolcott et al. studies, participants received groups of information sequentially. For example, participants in Tolcott et al. (1989b) received three intelligence reports containing fifteen pieces of information in each report. Participants in the Hogarth-Einhorn studies received only one piece of information each time they were provided subsequent information. Third, the subsequent information presented at each time period in the Tolcott et al. studies was ambiguous. For example, in Tolcott et al. (1989b), three of the fifteen pieces of information supported one hypothesis, three the other hypothesis, and nine pieces of information were neutral. In the Hogarth-Einhorn studies, subsequent information varied in strength but it always supported one hypothesis or the other.

Before turning to the Method section, we want to point out that order effects may be an important issue in the Army air defense task. Army air defense operators must identify incoming aircraft as friendly or hostile, and then engage hostiles under conditions of severe stress and time pressure. The identification is based on information of various degrees of diagnosticity. Information can be received sequentially and in various orders. The interfaces for newer air defense systems can display much of the essential initial information at once and continue to display this information as new information is received, but older systems can not and represent more of a sequential information flow. Moreover, some information (e.g., from headquarters) is not entered into the automated system, just passed to the operator via communications. In sum, the Army air defense task was considered both appropriate and important for studying the effect of information order on expert judgment.

METHOD

Participants

Sixty-three Army air defense operators from four battalions (three PATRIOT, one HAWK) stationed at Fort Bliss, Texas, participated in the experiment, which was conducted in mid-December 1989. All participants had completed their training and participated in Army air defense exercises simulating combat conditions in field settings, as well as on training simulators, prior to participating in the study. Examination of the background-information sheets completed by the participants at the time of the study indicated that thirty of the participants were enlisted personnel and thirty-three were officers. The mean air defense experience of the enlisted personnel was 4.1 years, with a range from 0.5 to 14 years. The mean air defense experience of the officers was 2.2 years, with a range from 0.5 to 7.5 years.

Design

There were five independent variables. The first independent variable was the order in which positive and negative information was presented to the participants. Positive information was defined as information consistent with the conclusion one would reach on the basis of the first piece of information; negative information was inconsistent with the conclusion. In Order #1, the second piece of information was positive (P) and the third piece of information was negative (N). In Order #2, the second piece of information was negative and the third piece of information was positive. Only three pieces of information were presented about each case (i.e., unknown aircraft). The positive and negative pieces of information were always the second and third pieces of information. In all cases, two pieces of information would point toward one conclusion (e.g., the aircraft was friendly) and the other piece of information would point toward the other conclusion (e.g., hostile).

The second independent variable was response mode: information was presented either sequentially with a probability estimate being obtained after each piece of information, or all at once with the probability estimate being obtained at that time. The third independent variable was whether the first piece of information supported the conclusion that the unknown aircraft was friendly or hostile. The fourth independent variable was whether the first piece of information was strong or weak. Based on pilot testing, a strong piece of information was defined as that with a probability ≥ 0.8 in support of friend or hostile. A weak piece of information had a probability > 0.5 but ≤ 0.6. The fifth independent variable was whether the negative information was strong or weak.

In sum, our experiment was the following crossed, factorial design: 2 (order) × 2 (response mode) × 2 (conclusion) × 2 (strength of first piece of information) × 2 (strength of negative information). The last three independent variables defined eight hypothetical aircraft. We defined sixteen critical cases by varying the order in which the last two pieces of information were presented about these aircraft. The experiment was a completely within-subject design because all participants evaluated all cases. The dependent variable for all cases was a probability estimate that the unknown aircraft was friendly or hostile.

Participants were also asked to write the reason for their final probability estimate when the information was presented sequentially. This was done in an attempt to understand their decision process. Pilot testing indicated that it would be too onerous to obtain written rationale every time the participants made their probability estimates. We would have preferred to use talk-aloud protocols, but they were not feasible because the participants had to be run in groups. As it turned out, the participants seldom provided written rational for their judgments and when it was provided, it was often too cryptic to be useful. Consequently, no analysis was performed on the written rationale.

Procedure

Information about all the cases defined by our design was presented in a booklet format. One book contained all the cases when the information was presented all at once. Two books were used to present all the cases when the information was presented sequentially. Two books were used to lessen the visual impact of the participants' task. The order in which participants completed the books was randomly determined.

The "all-at-once" book and the "sequential" books contained 36 cases. The same cases were presented, and in the same order, in each book. The 36 cases were composed of the 16 cases comprising the above design (not considering response mode) and 20 other cases. The latter were included to minimize the probability that participants would remember their answers to the 16 design cases. The 16 cases were randomly distributed throughout the books.

Participants completed the books individually. The books were completed in four group sittings—one for each of the four air defense battalions participating in the study. The groups varied in size, ranging from about ten to twenty participants. Participants completed the background information form prior to completing the books.

Hypotheses

Consistent with the Hogarth-Einhorn theoretical model, we predicted (1) a recency effect when the information was presented sequentially and a probability estimate was received after each piece of information, and (2) a no-order effect when the information was presented all at once and a probability estimate was obtained at that time. For the recency effect, negative information should be weighted more when the anchor is large; positive information should be weighted more when the anchor is small. This was hypothesized to result in a final probability estimate that was higher when negative information was followed by positive information (i.e., the NP order) than when positive information was followed by negative information (i.e., the PN order). This is represented pictorially in Figure 1 by what Einhorn and Hogarth (1987) called a "fishtail."

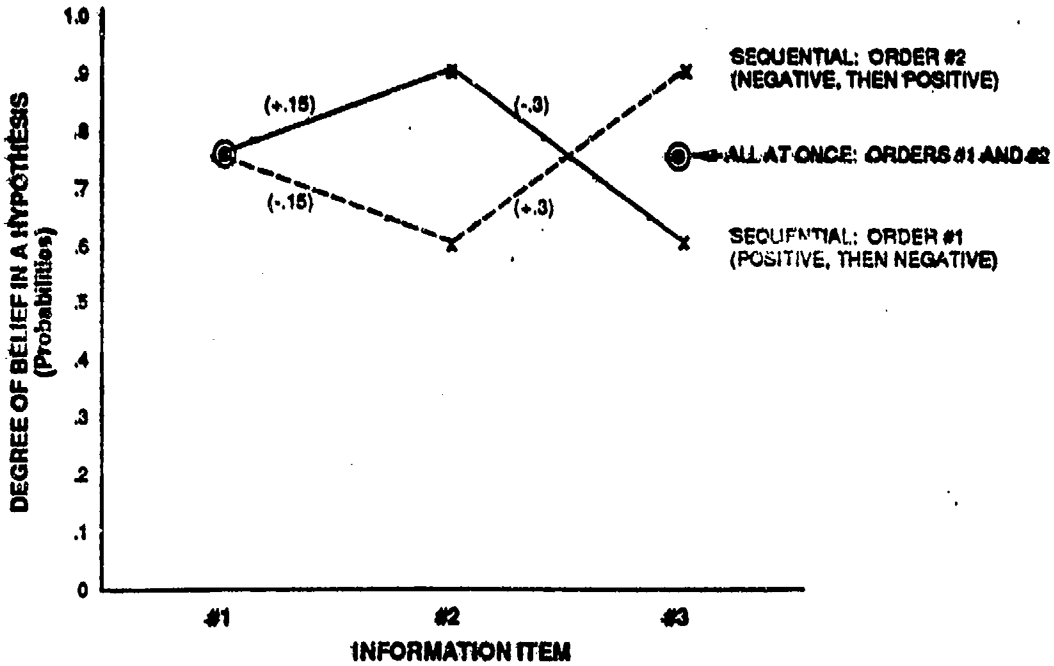

Figure 1. Notional representation: Predictions based on Hogarth-Einhorn model.

When all the information is presented at once, the Hogarth-Einhorn model predicts that people will anchor on the first piece of information in the series and adjust it on the basis of the aggregate impact of the remaining information. For our task in which participants receive only three pieces of information, there should be no differences in the final probability estimates due to order of the last two pieces of information. A notional representation of this prediction is also presented in Figure 1.

There were two alternative hypotheses. The first one, based on the research by Tolcott et al. demonstrating a "confirmation bias," was that there would be a primacy—not recency—effect when information was presented sequentially. This hypothesis was also based theoretically on an anchoring and adjustment model, but one where the initial information (i.e., anchor) was overweighted and subsequent information, particularly if it was negative, was given less weight the later it was received. This was predicted to result in a higher final probability estimate when positive information was followed by negative information (the PN order) than when negative information was followed by positive information (the NP order). The notional representation of this hypothesis is presented in Figure 2. The "all at once" condition did not exist in the studies by Tolcott et al.; consequently, no predictions were made.

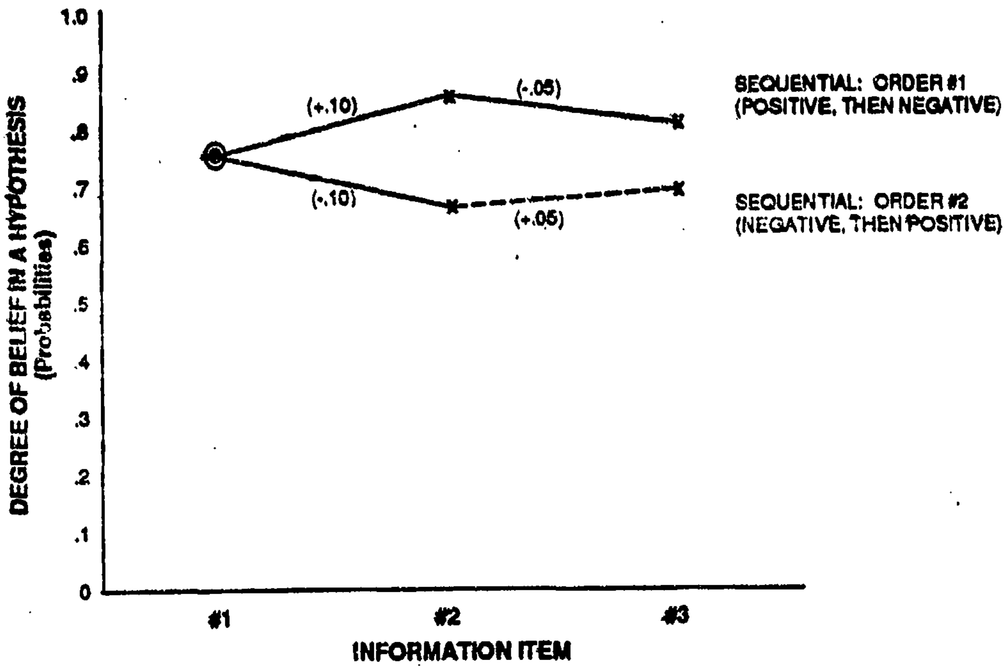

Figure 2. Notional representation: Predictions based on Tolcott et al.'s research.

The second alternative hypothesis was that there would be no order effects regardless of how information was presented. This was not only the proverbial null hypothesis, it was the first author's actual hypothesis. This hypothesis was based on the fact that most of the previous research was performed with university students. The first author did not think that highly trained personnel would give different final probability estimates to the same information when it was presented sequentially simply because one had switched the order of the last two (of only three) pieces of information.

Analysis

A within-subject 2 (order) × 2 (response mode) × 2 (conclusion) × 2 (strength of first piece of information) × 2 (strength of negative information) Analysis of Variance (ANOVA) was performed using (1) the final probability estimates after all three pieces of information were presented sequentially, and (2) the sole probability estimates when the information was presented all at once. The probability estimates for each case were coded in the direction specified by the "conclusion" independent variable. For example, a 0.80 probability estimate for either a friendly or a hostile aircraft was coded as 0.80. Probabilities were multiplied by 100 when the data were coded to avoid potential errors due to forgetting decimal points. (Note: This resulted in large "sums of squares" and "mean square" terms in the ANOVA.)

RESULTS

The ANOVA showed main effects for each of the five independent variables. First, there was a main effect for the first piece of information: $F(1,62) = 48.4$, $MS_e = 203.78$, $p < .001$. The final probabilities were, on the average, higher when the first piece of information was strong ($\bar{x}_{FP:S} = .69$) than weak ($\bar{x}_{FP:W} = .64$). Second, there was a main effect for strength of negative information: $F(1,62) = 129.69$, $MS_e = 537.19$, $p < .001$. The final probabilities were higher when negative information was weak ($\bar{x}_{N:W} = .72$) than strong ($\bar{x}_{N:S} = .60$): $F(1.62) = 129.69$, $MS_e = 537.19$, $p < .001$. These main effects indicate that the manipulation of these independent variables was successful. Third, there was a main effect for conclusion: $F(1,62) = 19.58$, $MS_e = 2481.47$, $p < .001$. Friendly aircraft received a higher final probability ($\bar{x}_F = .71$) than hostile aircraft ($\bar{x}_H = .60$). This was not a manipulation check because all cases were coded in the direction of the "conclusion" independent variable. Instead, the results indicate that the participants had more confidence in the identity of the friendly than hostile aircraft on the basis of the information provided to them.

We also obtained main effects for response mode [$F(1,62) = 5.85$, $MS_e = 857.14$, $p < .02$] and order [$F(1,62) = 29.84$, $MS_e = 1198.91$, $p < .001$]. The final probabilities were higher when information was presented all at once ($\bar{x}_O = .68$) than sequentially ($\bar{x}_S = .65$). And the final probabilities were higher when positive information followed negative information ($\bar{x}_{NP} = .71$) than when it preceded negative information ($\bar{x}_{PN} = .60$).

There were a number of significant interactions. Of particular interest, given our hypotheses, was that we obtained a significant order by response mode interaction: $F(1,62) = 20.47$, $MS_e = 1124.6$; $p < .001$. As shown in Figure 3, the mean final probability estimates were essentially identical when information was presented all at once. In contrast, the mean final probability estimates were significantly higher for the NP order ($\bar{x} = 0.73$) than the PN order ($\bar{x} = 0.58$) when the information was presented sequentially.

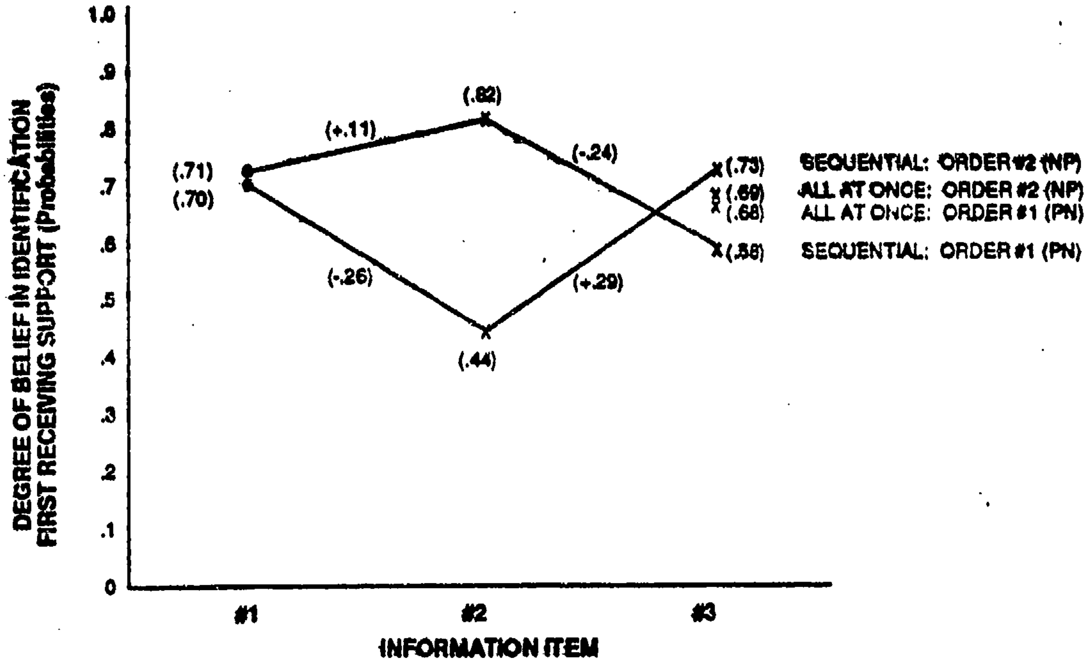

Figure 3. Mean probability estimates showing support for the interaction effect predicted by the Hogarth-Einhorn model.

These results support the predictions of the Hogarth-Einhorn model. No order effect was obtained when the information was presented all at once; an order effect was obtained when the information was presented sequentially. Further examination of Figure 3 shows that when the information was presented sequentially, positive information had a larger impact the smaller the anchor ($\bar{x}$ = +0.29 vs. $\bar{x}$ = +0.11), consistent with the model's predictions. However, in contrast with the model's predictions, negative information did not have a larger impact the larger the anchor ($\bar{x}_{PN}$ = -0.24 vs. $\bar{x}_{NP}$ = -0.26). Although there were minor differences, this finding was obtained, as shown in Table 1, for most of the eight comparisons defined by the 2 (conclusion) × 2 (strength of first piece of information) × 2 (strength of negative information) portion of the design.

Table 1

The Difference in the Mean Probability Estimate Prior to and After Receipt of Negative Information in the Sequential Information Presentation Condition

		NEGATIVE INFORMATION: STRONG PN ORDER	NEGATIVE INFORMATION: STRONG NP ORDER	NEGATIVE INFORMATION: WEAK PN ORDER	NEGATIVE INFORMATION: WEAK NP ORDER
CONCLUSION: FRIEND	INITIAL SUPPORT: STRONG	-.22	-.25	-.21	-.21
	INITIAL SUPPORT: WEAK	-.20	-.21	-.18	-.20
CONCLUSION: HOSTILE	INITIAL SUPPORT: STRONG	-.36	-.36	-.15	-.21
	INITIAL SUPPORT: WEAK	-.44	-.38	-.19	-.21

There were six other interactions that reached traditional significance levels. Five involved either the response mode or order independent variables. We consider them, in turn.

First, there was a conclusion × strength of negative information interaction: $\underline{F}(1,62) = 53.24$, $\underline{MS}_e = 538.54$, $\underline{p} < .001$. This occurred because strong negative information had a much greater effect on the final probabilities when the aircraft was hostile ($\bar{x}_{H,N:S} = .52$, $\bar{x}_{H,N:W} = .71$) than when it was friendly ($\bar{x}_{F,N:S} = .69$, $\bar{x}_{F,N:W} = .74$).

Second, there was a response mode × strength of negative information interaction: $\underline{F}(1,62) = 3.92$, $\underline{MS}_e = 173.28$, $\underline{p} = .05$. This occurred because the effect of strong versus weak negative information was slightly more pronounced when the information was presented all at once ($\bar{x}_{O,N:S} = .61$, $\bar{x}_{O,N:W} = .74$, diff. $= .13$) than sequentially ($\bar{x}_{S,N:S} = .59$, $\bar{x}_{S,N:W} = .70$, diff. $= .11$).

Third, there was an order × strength of negative information interaction: $\underline{F}(1,62) = 6.52$, $\underline{MS}_e = 100.21$, $\underline{p} < .02$. This occurred because the obtained order effect was slightly more pronounced when the negative information was strong ($\bar{x}_{PN,N:S} = .55$, $\bar{x}_{NP,N:S} = .65$, diff. $= .10$) than weak ($\bar{x}_{PN,N:W} = .69$, $\bar{x}_{NP,N:W} = .76$, diff. $= .07$).

Fourth, there was a conclusion × order × strength of negative information interaction: $\underline{F}(1,62) = 8.13$, $\underline{MS}_e = 80.113$, $\underline{p} < .01$. This occurred because the order × strength of negative information interaction was somewhat more pronounced for hostile than friendly aircraft when the negative information was strong, but not when it was weak.

Fifth, there was a response mode × conclusion × order × negative information interaction: $\underline{F}(1,62) = 8.32$, $\underline{MS}_e = 72.94$, $\underline{p} < .01$. This occurred because the conclusion × order × strength of negative information interaction occurred primarily when information was presented sequentially, not all at once.

Lastly, there was a strength of first piece of information × conclusion × order × strength of negative information interaction: $\underline{F}(1,62) = 6.96$, $\underline{MS}_e = 76.49$, $\underline{p} = .01$. This occurred because the conclusion × order ×

strength of negative information interaction occurred when the first piece of information was weak, but not when it was strong.

We examined the amount of variation (R^2) in the final probability estimates accounted for by the statistically significant effects. Four points are important to note here. First, we found that the significant effects due to order and response mode accounted for, in total, a small proportion of the total variation in the participants' final probability estimates; $R^2 = 0.074$. For example, the R^2 for the order by response mode interaction was 0.025; the R^2 for the order main effect was 0.039. Second, the "strength-of-negative-information" main effect ($R^2 = 0.076$) and the "conclusion" (i.e., friend versus hostile) main effect ($R^2 = 0.053$) accounted for more variance than any single "order" or "response-mode" effect. This suggests that these factors had a stronger effect on Army air defense operators' probability estimates than did order or response mode. Third, the R^2 for each of the third- and fourth-order interactions was .001. Although they achieved traditional levels for statistical significance, these effects were small. And, fourth, all significant (i.e., $p \leq 0.05$) effects accounted for, in total, 24.7 percent of the total variation. This indicates that most of the variation in the participants' final probability estimates was due to subjects and not the manipulated independent variables.

Three post-hoc analyses were performed in an attempt to better understand the effect of individual differences in the participants' probability estimates. First, an ANOVA was performed using participants' rank as a sixth independent variable. As the reader will remember, thirty of the participants were enlisted personnel and thirty-three were officers.

The ANOVA resulted in three interesting findings. First, there was a main effect for rank: $F(1,61) = 7.05$, $MS_e = 1445.596$, and $p = 0.01$. Officers had a larger final mean probability estimate (0.69) than enlisted personnel (0.645). Second, there was a significant rank-by-order interaction: $F(1,61) = 5.11$, $MS_e = 1124.325$, and $p = 0.026$. The difference in the mean probability estimate for the two order conditions was smaller for officers (0.06) than enlisted personnel (0.12). And, third, the rank-by-order-by-response-mode interaction approached significance: $F(1,61) = 3.67$, $MS_e = 1078.202$, $p = 0.057$. As Figure 4 illustrates, the difference in the final mean probabilities estimates for the NP and PN orders was larger for enlisted personnel (0.22) than officers (0.10) when information was presented sequentially. In contrast, the difference was small (0.03) and identical for enlisted and officer personnel when information was presented all at once. However, all significant effects incorporating rank accounted for little additional variation in the participants' probability estimates ($R^2 < 0.03$).

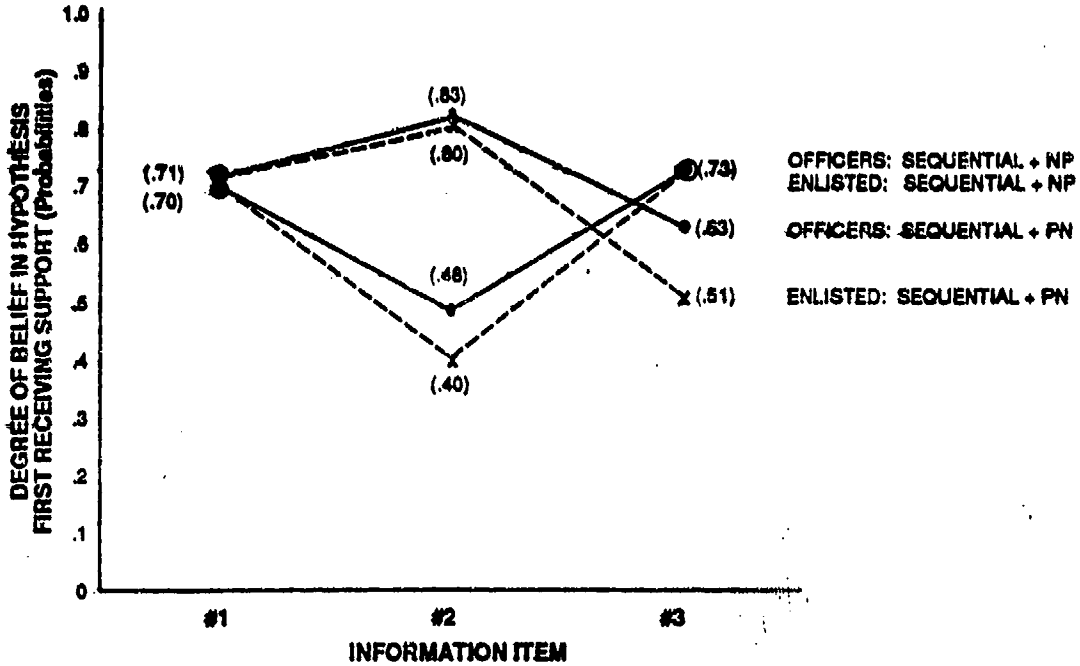

Figure 4. Mean probability estimates showing larger order effects for enlisted personnel than officers when information was presented sequentially.

The second post-hoc analysis was the addition of a seventh factor—experience level—to the ANOVA. This factor had two conditions: (1) less than or equal to one year's air defense experience, and (2) more than one year's experience. There were 17 officers and 9 enlisted personnel with less than one year's experience, and 14 officers and 21 enlisted personnel with more than a year's experience; two of the officers did not indicate their years of experience. Neither the experience main effect, experience-by-rank interaction, nor any of the experience-by-order interactions involving less than five factors were significant at $p < 0.05$ level.

The third post-hoc analysis examined the extent to which individual participants exhibited order effects when information was presented sequentially and all at once, respectively. This was accomplished by computing the difference in the final probability estimates for the NP and PN orders when information was presented sequentially, and the sole probability estimates when information was presented all at once, for the eight cases defined by the 2 (conclusion) × 2 (strength of first piece of information) × 2 (strength of negative information) portion of the design. For the purposes of this analysis, a positive difference was classified as supporting an order effect.

Table 2 shows the differences when the information was presented sequentially. As can be seen, thirty-seven percent (N = 11) of the enlisted personnel and nine percent (N = 3) of the officers showed an order effect for all eight comparisons. The mean effect for the eleven enlisted personnel was 0.45 with a mean standard deviation of 0.118; the mean effect for the three officers was 0.72 with a mean standard deviation of 0.084. These data mean that, on the average, these fourteen participants switched their hypothesis as to whether the unknown aircraft was friendly or hostile, based on the order of the same information.

Table 2

Individual Difference Data Supporting an Order Effect when Information was Presented Sequentially and a Probability Estimate was Obtained after each Piece of Information

# OF COMPARISONS FOR WHICH SUBJECT SHOWED EFFECT	ENLISTED				OFFICERS			
	#	%	AVERAGE EFFECT OVERALL*	AVERAGE STD	#	%	AVERAGE EFFECT OVERALL*	AVERAGE STD
8	11	37	.45	.118	3	9	.72	.084
7	1	3	.34	.239	5	15	.13	.109
6	3	10	.29	.225	1	3	.15	.138
5	4	13	.12	.143	1	3	.03	.147
4	2	7	.03	.098	8	24	.05	.114
3	3	10	.01	.128	6	18	.02	.094
2	5	17	-.03	.103	5	15	-.04	.112
1	0	0	0	---	4	12	-.01	.069
0	1	3	-.06	.067	0	0	0	---
	N = 30	100%	$\bar{X}$ = .22	$\bar{X}$ = .132	N = 33	99%	$\bar{X}$ = .10	$\bar{X}$ = .103

*POSITIVE NUMBERS: DIRECTION SUPPORTING ORDER EFFECT

Further examination of Table 2 indicates the other extreme as well. Specifically, almost one-third of the enlisted personnel and almost one-half of the officers *failed* to show an order effect for at least half (i.e., four or more) of the comparisons. Moreover, the mean difference in their NP and PN orders was essentially zero indicating that, even when they occurred, the order effects were very weak for these participants.

Table 3 shows the differences in the NP and PN orders when the information was presented all at once. As can be seen, some participants showed an order effect more consistently than others. However, most of the differences are close to zero, indicating that the effect, even when it occurred, was quite weak when the information was presented all at once.

Table 3

Individual Difference Data Supporting an Order Effect when Information was Presented All at Once and a Probability Estimate was Obtained at that Time

# OF COMPARISONS FOR WHICH SUBJECT SHOWED EFFECT	ENLISTED				OFFICERS			
	#	%	AVERAGE EFFECT OVERALL*	AVERAGE STD	#	%	AVERAGE EFFECT OVERALL*	AVERAGE STD
8	0	0	----	----	0	0	----	----
7	1	3	.08	.098	0	0	----	----
6	3	10	.10	.110	1	3	.10	.140
5	8	27	.04	.114	3	9	.08	.127
4	7	23	.01	.141	11	33	.03	.126
3	6	20	-.01	.120	8	24	.00	.089
2	2	6	-.02	.120	4	12	.01	.069
1	2	6	-.03	.066	3	9	-.04	.107
0	1	3	-.02	.043	3	9	-.04	.048
	N = 30	98%	$\bar{x}$ = .022	$\bar{x}$ = .115	N = 33	99%	$\bar{x}$ = .011	$\bar{x}$ = .102

*POSITIVE NUMBERS: DIRECTION SUPPORTING ORDER EFFECT

DISCUSSION

The results of the experiment reported herein with trained personnel performing their substantive task support the predictions of the Hogarth-Einhorn model for belief updating. Consistent with their anchoring and adjustment model, (1) an order effect was obtained when information was presented sequentially and a probability estimate was obtained after each piece of information, and (2) a no order effect was obtained when information was presented all at once and the probability estimate was obtained at that time.

The obtained findings are consistent with those of (1) Serfaty et al. (1989), who studied military situation assessment judgments, and (2) Asare (1990), Ashton and Ashton (1988; 1990), and Messier (1990), all of whom studied the judgments of auditors. For clarification purposes, we want to note that the latter studies used the term "presentation mode" to refer to what Hogarth and Einhorn (1989) call "response mode." Actually, presentation mode and response mode are confounded in these studies and ours. Future research should attempt to disentangle these variables by, for example, presenting information sequentially without eliciting a probability estimate for each piece of information. We have chosen to use the term "response mode" instead of "presentation mode" for consistency with Hogarth and Einhorn's (1989) presentation.

The obtained order effect reported herein and elsewhere suggests that, when information is presented sequentially and a probability estimate is obtained each time, people anchor on the current (not initial) position and adjust it on the basis of the strength and direction (positive or negative) of each new piece of information. This results in a recency effect. These findings differ from the predictions based on the research by Tolcott et al. (1988, 1989a, 1989b) demonstrating a confirmation bias—i.e., primacy effect. However, as we noted in the Introduction, task differences may well account for this finding. For the experiment reported herein, the initial information (although it varied in strength) unambiguously pointed toward one hypothesis or another; only one piece of subsequent information was presented at a time; and the subsequent information was unambiguous. None of these conditions existed in the Tolcott et al. studies.

We also found a number of interaction effects involving either response mode or order. For example, we found that the size of the order effect was affected by both the strength of the negative information and whether the aircraft was friendly or hostile. Moreover, this conclusion × order × strength of negative information interaction occurred primarily when information was presented sequentially, not when it was presented all at once. Although these third- and fourth-order interactions accounted for a very small proportion of the variance in the participants' final probability estimates, they do indicate that order effects are affected by more than just the response mode.

Post-hoc analyses indicated that the order effect was subject to individual differences. In particular, enlisted personnel showed a larger order effect than officers when information was presented sequentially; neither group showed an order effect when information was presented all at once. The amount of air defense experience did not affect the results. We consider this finding consistent with that of Messier and Tubbs (1990), although they concluded that there was an experience main effect by accepting an alpha level of p = 0.096 as significant.

We think the effect due to rank is a function of the duties typically performed by Army air defense officers and enlisted personnel. In particular, officers actually make the identification and engagement decisions. Although enlisted personnel monitor the situation, their primary task is to implement the engagement decision. Therefore, due to the requirements of their jobs and on-the-job training, we would hypothesize that officers are less susceptible to order effects when information is presented sequentially and a probability estimate is obtained each time. A second, and, in our opinion, less likely hypothesis is that officers are more likely to attend college and, therefore, obtain more training in analytical thinking than enlisted personnel.

A second finding was that some of the participants exhibited large order effects when information was presented sequentially, while others showed no order effects at all. At one extreme, thirty-seven percent (N = 11) of the enlisted personnel and nine percent (N = 3) of the officers not only showed an order effect for all eight of the possible comparisons, but the differences in their probability estimates was so large that they routinely switched their judgment as to whether the unknown aircraft was friendly or hostile, based on the order with which the same information was presented. In contrast, almost one-third of the enlisted personnel and almost one-half of the officers who participated in the experiment showed essentially no order effects when information was presented sequentially. The latter results detract from the universality of the Hogarth-Einhorn model.

Two other findings also detracted from the support for the Hogarth-Einhorn model. First, in contrast to the model's prediction, negative information did not, on the average, have a larger impact the larger the anchor. Although there were minor differences, this finding held for all eight comparisons defined by the 2 (conclusion) × 2 (strength of first piece of information) × 2 (strength of negative information) portion of the design. This finding is at odds with the empirical results obtained by Asare (1990), Ashton and Ashton (1988), and Einhorn and Hogarth (1987).

We offer two hypotheses to explain this finding. The first hypothesis is that the difference between the two anchors prior to the receipt of negative information was too small in our study to generate the hypothesized effect. Examination of Figure 3 shows that, prior to the receipt of negative information, there was a difference of only 0.12 in the mean probability estimates after the second piece of information for the PN order (0.82) and the first piece of information for the NP order (0.70). This seems small to us. Moreover, the largest difference for any of the eight comparisons was only 0.14.

The second hypothesis, which could co-exist with the first one, is that participants were able to use the positive information in the PN order to generate an explanation for the negative information and, thereby, lessen its impact. For example, if participants were told (1) that the aircraft had an IFF Friendly Response and (2) was identified as friendly by higher headquarters—two strong but not perfect predictors of friend—then they might somewhat discount the third piece of information, which is that the aircraft is jamming friendly radar, by assuming that the aircraft was a "disoriented friendly jammer"—a plausible event. The amount of discounting may not have depended as much on the size of the prior probabilities or number of positive pieces of information in our study, as on the generation and plausibility of the explanation. The larger final probability estimate for the NP than PN order might occur because the positive information not only substantiates but exaggerates the explanation, consistent with the use of an anchoring and adjustment heuristic. Obviously, future research is required to test these hypotheses.

The second finding detracting from the support for the Hogarth-Einhorn model is that, cumulatively, the order effects accounted for a small proportion of the total variation in the participants' probability estimates ($R^2 < 0.10$). Moreover, other factors, most notably the "strength of negative information" and the "conclusion," accounted for more variance than did any single effect due to information order or response mode. Ashton and Ashton (1988; in press) and Messier (1990) also found the strength of negative information to be a strong predictor of participants' probability estimates. And finally, most of the variance was due to subjects and not the independent variables manipulated in this experiment. Together, these data suggest that, although they resulted in statistically significant findings consistent with the Hogarth-Einhorn model's predictions, information order and response mode may not, in total, be the best predictors of air defense operators' mean probability estimates. Or, to put it differently, the effects due to information order and response mode may be statistically significant but small.

This conclusion must be considered with caution for three reasons. First, most of previous research investigating the impact of information order and response mode has not presented R^2 information. Ashton (1988; in press), Einhorn and Hogarth (1987), Hogarth and Einhorn (1989), and Messier (1990), for example, do not present all the data necessary to reconstruct the R^2s. Asare (1990) does, however, present enough data in the ANOVA tables for his two experiments to calculate R^2s. The R^2 for the significant order effect in his first experiment was 0.22; it was 0.067 in the second experiment.

Second, the error terms in the current experiment were no doubt inflated by (1) the artificialities of our paper-and-pencil task, and (2) the fact that air defense operators do not make probability estimates as part of their job. The inflated error terms certainly reduced the amount of systematic variation in the participants' responses that could be accounted for by information presentation and response mode. Indeed, under these circumstances, one might argue that accounting for ten percent of the variance is quite good.

Third, as Christensen-Szalanski and Fobian (in press) point out in their meta-analysis of the hindsight-bias literature, determining whether an effect size is of practical significance or utility is situation-specific. Discussion with senior-level Army air defense personnel indicates that the results may have implications for (1) training air defense operators, and (2) designing information displays for future air defense systems. The case would be considerable stronger, however, if the results were obtained (i) using actual air defense simulators, and (ii) if we had investigated whether changes in the probability estimates also affected behavior, which, in this case, would be the decision to engage aircraft identified as hostile. Each point is considered briefly, in turn.

First, the generalizability of the obtained results is the critical issue driving the desire for simulators. The issue of generalizability is not unique to our study, for all the studies referenced herein have used paper-and-pencil tasks. Whether such representations adequately addressed the issue of generalizability depends on their task domain. Second, there is minimal research investigating the relation between order effects and behavior. Asare (1990) has shown that information order can affect auditors' likelihood estimates and behavior. However, Meisser (1990) found the effect on probabilities—but not behavior—using a different auditing task.

Finally, the current study suggests that care be used in employing knowledge-engineering procedures to develop expert systems. Knowledge engineering with trained personnel can proceed in an interview format that emphasizes the sequential presentation of information and uses paper-and-pencil tasks analogous to those used in the current experiment (e.g., see Forsythe and Buchanan, 1989). As Lehner and Adelman (1987; in press), Hammond (1987), and von Winterfeldt (1988) note, most knowledge engineers are not familiar with the cognitive heuristics and biases literature and may, inadvertently, encode biases into the knowledge bases of expert systems. The results of the current effort indicate that knowledge engineers should use multiple-information orders and response modes to assess the consistency of obtained probability estimates. Ideally, they would then use case data with the correct answers to assess the predictive accuracy of the elicited knowledge.

REFERENCES

Asare, S.K. (1990). *The auditor's going concern opinion decision: Interaction of task variables and the sequential processing of evidence* (Technical Report). Gainesville, FL: University of Florida, Fisher School of Accounting.

Ashton, A.H. and Ashton, R.H. (1988). Sequential belief revision in auditing. *The Accounting Review, 63*, 623-641.

Ashton, R.H. and Ashton, A.H. (1990). Evidence-responsiveness in professional judgment: Effects of positive versus negative evidence and presentation mode. *Organizational Behavior and Human Decision Processes, 46*, 1-19.

Christensen-Szalanski, J.J.J. and Fobian, C.S. (in press). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes.*

Einhorn, H.J. and Hogarth, R.M. (1987). *Adaptation and inertia in belief updating: The Contrast-Inertia Model* (Technical Report). Chicago, IL: University of Chicago, Graduate School of Business.

Fischhoff, B., Slovic, P., and Lichtenstein, S. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 330-344.

Forsythe, D. and Buchanan, B.G. (1989). Knowledge acquisition for expert systems: Some pitfalls and suggestions. *IEEE Transactions on Systems, Man, and Cybernetics, 19*, 435-442.

Hamm, R.M. (1987). *Diagnostic inference: People's use of information in incomplete Bayesian word problems* (Technical Report 87-11). Boulder, CO: University of Colorado, Institute of Cognitive Science.

Hammond, K.R. (1987). Toward a unified approach to the study of expert judgment. In J. Mumpower, L. Phillips, O. Renn, and V. Uppuluri (Eds.), *Expert judgment and expert systems.* NY: Springer-Verlag.

Hogarth, R.M. (1987). *Judgment and choice* (2nd Edition). NY: Wiley-Interscience.

Hogarth, R.M. and Einhorn, H.J. (1989). *Order effects in belief updating: The Belief-Adjustment Model* (Technical Report). Chicago, IL: University of Chicago, Graduate School of Business.

Kahneman, D., Slovic, P., and Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases.* NY: Cambridge University Press.

Lehner, P.E. and Adelman, L. (1987). Procedures for investigating cognitive biases in knowledge engineering. *Proceedings of the 1987 IEEE Conference on Systems, Man, and Cybernetics, 1*, 376-377.

Lehner, P.E. and Adelman, L. (in press). Behavioral decision theory and its implications for knowledge engineering. *Knowledge Engineering Review*.

Messier, W.F., Jr. (1990). *The sequencing of audit evidence: Its impact on the extent of audit testing and report formulation* (Technical Report). Gainesville, FL: University of Florida, Fisher School of Accounting.

Messier, W.F., Jr., and Tubbs, R.M. (1990). *Mitigating recency effects in belief revision: The impact of experience and review* (Technical Report). Gainesville, FL: University of Florida, Fisher School of Accounting.

Nisbett, R.E. and Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Serfaty, D., Entin, E. and Tenney, R. (1989). Planning with uncertain and conflicting information. In S.E. Johnson and A.H. Levis (Eds.), *Science of command and control. Part 2: Coping with complexity*, Fairfax, VA: AFCEA International Press, 91-100.

Tolcott, M.A. and Marvin, F.F. (1988). *Reducing the confirmation bias in an evolving situation* (Technical Report 88-11). Reston, VA: Decision Science Consortium, Inc.

Tolcott, M.A., Marvin, F.F., and Bresnick, T.A. (1989a). The confirmation bias in military situation assessment. *Proceedings of the 57th MORS Symposium*, to appear.

Tolcott, M.A., Marvin, F.F., and Lehner, P.E. (1989b). Expert decision making in evolving situations. *IEEE Transactions on Systems, Man, and Cybernetics, 19*, 606-615.

Tversky, A. and Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105-110.

von Winterfeldt, D. (1988). Expert systems and behavioral decision research. *Decision Support Systems, 4*, 461-471.