Journal of Accounting Research
Vol. 19 No. 1 Spring 1981
Printed in U.S.A.

Anchoring and Adjustment in Probabilistic Inference in Auditing

EDWARD J. JOYCE* AND GARY C. BIDDLE†

Auditors are faced with the task of formulating opinions about the fairness of their clients' financial statements. In doing so, they use their professional judgment to determine the type and amount of information to collect, the timing and manner of collecting it, and the implications of the information collected. This information is rarely, if ever, perfectly reliable or perfectly predictive of the "true" state of a client's financial statements. Nevertheless, auditors may be held liable at common law or under the federal securities laws should the audited financial statements prove to be unrepresentative of this true state. Thus, it is important for auditors to have the ability to formulate appropriately judgments based on probabilistic data.

In this paper, we describe the results of experiments designed to assess whether auditors formulate judgments in accordance with normative principles of decision making or whether a particular alternative to the normative model of decision making under uncertainty is employed. In the next section, we discuss several alternatives to normative decision models, focusing on the anchoring and adjustment heuristic which forms the basis for our experiments.

Heuristics in Probabilistic Judgment

The empirical research on human judgment has led psychologists to the following general conclusions (Hogarth [1975, p. 272]): (1) humans

---

* Associate Professor, University of Minnesota; †Assistant Professor, University of Chicago. The experiments reported in this paper were among those funded by a grant from the Peat, Marwick, Mitchell Foundation through its Research Opportunities in Auditing Program. The views expressed herein are our own and do not necessarily reflect the views of the Peat, Marwick, Mitchell Foundation. This paper has benefited from the comments

have limited information-processing capacity (Miller [1956]; Newell and Simon [1972]; Slovic and Lichtenstein [1971]), and (2) the nature of the judgmental task determines to a great extent the decision strategies which they employ (Edwards [1971]; Einhorn and Hogarth [1981], reprinted in part in this issue; Payne [1976]; Simon and Newell [1971]; Slovic, Fischhoff, and Lichtenstein [1977]).

This literature suggests that when coping with complex situations humans resort to cognitively tractable decision strategies known as heuristics. These heuristics are cognitively simpler mechanisms than normative models and often result in decisions which are compatible with the latter (Tversky and Kahneman [1974]). However, because these heuristics imply fundamentally different cognitive activity than normative models, and because they rely on variables not relevant in normative models (or ignore variables which are relevant in normative models), they sometimes lead to systematic decision errors.

Where beliefs are expressed as subjective probabilities and the task facing the decision maker is to revise his beliefs given new information, the appropriate normative model is Bayes' Theorem.[1] Much of the empirical research on human probabilistic judgment has consisted of comparing human responses on probability revision tasks to the Bayesian optimal response. The evidence suggests that humans process information in a manner that is fundamentally different from Bayes' Theorem: "In his evaluation of evidence man . . . is not Bayesian at all" (Kahneman and Tversky [1972*a*]).

Much of the recent research in probabilistic judgment moved toward developing representations of human judgments that have more descriptive power than Bayes' Theorem. If humans are not Bayesian, what, then, are they? In a series of papers, Tversky and Kahneman [1971; 1973; 1974; 1977] and Kahneman and Tversky [1972*a*; 1972*b*; 1973] have formulated and empirically tested three decision heuristics which they believe provide more accurate descriptions of human judgment in many situations than Bayes' Theorem. These heuristics are: (1) representativeness, (2) availability, and (3) anchoring and adjustment. Tests of the representativeness heuristic in the accounting literature include Swieringa et al. [1976], Uecker and Kinney [1977], Gibbins [1977], and Joyce and Biddle [forthcoming]. In general, these studies provide qualified support for the employment of the representativeness heuristic in probabilistic inference tasks. No tests for the existence of availability or anchoring and adjustment have yet appeared in the accounting literature. This paper reports the results of a series of tests for the existence of the anchoring and adjustment heuristic in auditing-related inference tasks.

---

of Nick Dopuch, Mike Gibbins, Bob Libby, Ed Nathan, and the accounting workshop of the University of Texas at Austin. [Accepted for publication October 1980.]

[1] There is consensus, though not unanimity, among researchers that Bayes' Theorem is the normative model for belief revision. Alternative normative theories of inference include Shafer [1976] and Krantz and Miyamoto [1980].

THE ANCHORING AND ADJUSTMENT HEURISTIC

Judgments made in conformance with anchoring and adjustment are hypothesized to result from the following process: (1) An initial value or starting point (the anchor) which seems natural to the decision maker is selected. (Just how an initial value is chosen is not clear.) (2) The anchor obtained in (1) is adjusted to take into account the additional information received. This adjustment is typically in the normatively appropriate direction, but generally not of sufficient magnitude.

To illustrate, consider the problem of determining the magnitude and timing of substantive tests to perform in the audit of accounts receivable. A natural starting point would be the substantive tests performed in the previous engagement, adjusted for changes in circumstances between the two engagements, for example, a weaker internal control system. If the auditor were employing the anchoring and adjustment heuristic, he would modify his substantive tests in the normatively appropriate direction (i.e., he would increase the tests and/or change their timing to make them more diagnostic), but he would not move far enough from the anchor. This could result in insufficient evidential matter to support the audit opinion.

Although behavior consistent with the anchoring and adjustment heuristic has been documented, the results have been obtained almost exclusively using student subjects in rather abstract, general knowledge settings. (For example, "What is the percentage of African countries in the United Nations?" Tversky and Kahneman [1974].) It is conceivable (perhaps likely) that trained professionals would use fundamentally different cognitive strategies in working problems related to their expertise—strategies much more in tune with normative principles than anchoring and adjustment. In the following sections, we present results from a series of experiments in which a total of 182 practicing auditors participated which were designed to test for the existence of the anchoring and adjustment heuristic in auditing-related inference tasks. The tasks varied in the extent to which auditors would find them familiar in order to assess whether use of anchoring and adjustment was inversely related to problem familiarity.

Experiments

A total of six experiments were conducted, all of which were pilot-tested on experienced auditors and revised where necessary. The first four of these experiments were designed to test for the bias of insufficient adjustment from an anchor. In these experiments, the subjects were given, or asked to provide, a starting point and required to make an adjustment from this starting point. If employed, the anchoring and adjustment heuristic would result in judgments insufficiently far from the starting point. Thus, the final judgment would be contingent on the starting point, even where the latter is normatively irrelevant.

The last two experiments, 3*A* and 3*B*, were designed to test for anchoring and adjustment in conjunctive and disjunctive events. An example of a conjuctive event is getting "heads" in n successive flips of a coin where n is an integer greater than one. An example of a disjunctive event is getting *at least* one "heads" in n flips of a coin. The anchor in both of these cases would be the probability of an elementary event (i.e., the probability of "heads" in one toss of the coin). In a conjunctive event, the appropriate direction of the revision from this anchor would be downward, of course. Use of the anchoring and adjustment heuristic would lead to an adjustment in this direction, but by an insufficient amount. Thus, the probability of conjunctive events would be overestimated. The appropriate direction of the revision from the elementary event probability in a disjunctive event would be upward, but by an insufficient amount, if the anchoring and adjustment heuristic were employed. Thus, the probability of disjunctive events would be underestimated. Table 1 summarizes the experiments.

EXPERIMENT 1A

The subjects in condition *A* of experiment 1*A* responded to the following problem:

It is well known that many cases of management fraud go undetected even when competent annual audits are performed. The reason, of course, is that Generally Accepted Auditing Standards are not designed specifically to detect executive-level management fraud. We are interested in obtaining an estimate from practicing auditors of the prevalence of executive-level management fraud as a first step in ascertaining the scope of the problem.

1. Based on your audit experience, is the incidence of significant executive-level management fraud more than 10 in each 1,000 firms (i.e., 1%) audited by Big Eight accounting firms? (Circle one number.)
   Yes, more than 10 in each 1,000 Big Eight clients have significant executive-level management fraud 1
   No, fewer than 10 in each 1,000 Big Eight clients have significant executive-level management fraud 2
2. What is your estimate of the number of Big Eight clients per 1,000 that have significant executive-level management fraud?
   (Fill in the blank below with the appropriate number.)
   ____ in each 1,000 Big Eight clients have significant executive-level management fraud.

Condition *B* differed only in that the subjects were asked whether the incidence of executive-level management fraud among Big Eight clients was more or less than 200 in each 1,000 audited. They were then instructed as in part (2) to provide a point estimate of the number of Big Eight clients per 1,000 where such frauds have been committed.

The purpose of part (1) in each condition of this experiment was to provide a starting point or anchor for the subjects' estimates in part (2). That starting point was 10 per 1,000 in condition *A* and 200 per 1,000 in condition *B*. These starting points should be irrelevant for purposes of making a response in part (2), and thus the responses to part (2) should not differ between the two experimental conditions. If the anchoring and adjustment heuristic were employed by the subjects, however, the mean

TABLE 1

Summary of Experiments

Experiment	Bias Tested	Task	Number of Subjects	Number of Experimental Conditions	Factor(s) Manipulated	Remarks
1*A*	Insufficient adjustment	Fraud estimation	50	2	Anchor extremity	Findings consistent with anchoring
1*B*	Insufficient adjustment	Fraud estimation	132	4	Anchor extremity	Findings consistent with anchoring
2*A*	Insufficient adjustment	Audit planning	50	3	(1) Existence of anchor (2) Internal control strength	Findings inconclusive
2*B*	Insufficient adjustment	Audit planning	132	4	(1) Internal control strength (2) Order of system presentation	Order effect
3*A*	Bias in evaluation of conjunctive events	(1) Critical event prediction (2) Audit report	50	2	Number of conjunctive elements	Probability assessments—OK Opinions—?
3*B*	Bias in evaluation of conjunctive and disjunctive events	(1) Critical event prediction (2) Audit report	132	4	(1) Number of elements (2) Conjunctive or disjunctive problem formulation	Probability assessments—OK Opinions—?

response in condition *A* would be lower than the mean response in condition *B*.

The subjects were 50 practicing auditors from the Chicago offices of two Big Eight public accounting firms. Twenty-seven auditors from firm 1 were tested first in a single group at the firm's Chicago office. Shortly thereafter 23 auditors from firm 2 were tested at that firm's Chicago office. These subjects varied widely in their experience, with a range of 0 to 27 years and a mean experience of 4.00 years. The mean experience level varied considerably between the firms: 2.44 years for firm 1 and 5.33 years for firm 2. The present problem was one of eleven problems on probabilistic judgment the subjects were asked to complete. The problems were self-paced and completed in 35–65 minutes. All subjects were randomly assigned to the experimental conditions.

Results. The results of experiment 1*A* are reported in figure 1. The mean response in part (2) differed considerably between conditions ($A = 16.52$ per 1,000 versus $B = 43.11$ per 1,000). This difference is statistically significant ($p < .015$) and consistent with the use of the anchoring and adjustment heuristic. The starting point provided the auditors, although normatively irrelevant, apparently affected their judgments of the incidence of fraud.

The anchors apparently also had a marked effect on the within-cell variability of the subjects' responses. The standard deviation of the subjects' responses was greater in the extreme anchor condition (condition *B*) than in condition *A* (36.97 versus 22.41). One possible explanation is that the condition *A* subjects who answered "no" to part (1) (and thus thought that the frequency of fraud was less than 10 in 1,000) could exhibit little variability in their responses because of the floor effect of the response scale. However, in figure 1, the condition *A* subjects who answered "yes" on part (1) (and thus were not subject to the floor effect)

Experimental Condition

	A(1%)		B(20%)	
	Subjects responding "yes" to part (1) (n=14)	Subjects responding "no" to part (1) (n=11)	Subjects responding "yes" to part (1) (n=24)	Subjects responding "no" to part (1) (n=1)
Mean	33.27	3.36	43.86	25.0
Std. dev.	24.31	1.63	36.76	0
Condition mean	16.52		43.11	
Condition std. dev.	22.41		36.97	

FIG. 1.—Experiment 1*A*: means and standard deviations for part (2) as a function of experimental condition and part (1) response.

exhibited less variability than the subjects in condition B (standard deviation = 24.31 versus 36.97). Thus, the farther the anchor from the judged stimulus (the base rate of fraud), the greater the variance in judgments.

While the results of this experiment are consistent with the use of the anchoring and adjustment heuristic, they are also consistent with many other hypotheses which would make the same prediction in this situation. While it is impossible to rule out all of these alternative hypotheses, it is possible to rule out one of the most likely. It is conceivable that the auditors' perceptions of what the experimenters meant by "significant" executive-level management fraud was contingent on the anchor provided in part (1). Thus, the subjects in condition A given an anchor of 10 in 1,000 might have thought that "significant" was a higher dollar amount given the relatively low base rate provided, while the subjects in condition B with an anchor of 200 in 1,000 might have thought that the dollar threshold for a significant fraud was much lower given the high anchor provided. This would mean that the significant difference between the group means on part (2) was due to a systematic difference in the perceived threshold for significant executive-level management fraud between the experimental conditions, rather than the result of anchoring and adjustment. Experiment 1B was conducted to rule out this alternative explanation of the results reported above.

EXPERIMENT 1B

The only differences between experiment 1A and experiment 1B were: (1) The threshold for significant executive-level management fraud was fixed across all conditions at $250,000. If the rival hypothesis were true, then no systematic differences among the group means on part (2) of experiment 1B should occur. However, the anchoring and adjustment hypothesis would again be confirmed if significant differences among the group means are observed. (2) Four experimental conditions (versus two) were employed. By introducing both a very low (1 in 1,000) and an extremely high (300 in 1,000) anchor, we were able to test the hypothesis that the anchoring effect is a monotone increasing function of the magnitude of the anchor. In other words, this hypothesis suggests that subjects may ignore an anchor or starting point which they view as extreme.

The subjects were 132 practicing auditors from a single Big Eight public accounting firm. The subjects were tested in two groups. The first group of 55 subjects was tested at their firm's Chicago-Region senior-level training school. The second group of 77 subjects was tested at their firm's Chicago-Region pre-senior or in-charge training school. Ninety-five percent of the subjects had one or two years of experience. The subjects were randomly assigned to the experimental conditions. The present problem was one of ten on probabilistic judgment the subjects were asked to complete. The subjects worked at their own pace and completed the problems in 30–60 minutes.

Results. The mean responses per experimental condition are reported in figure 2. Again, since the anchors provided the subjects in part (1) of the experiment are normatively irrelevant for purposes of responding to part (2), the four cell means should be identical. The null hypothesis of no differences among the cell means for the part (2) responses was tested against two different alternative hypotheses.

First, the Kruskal-Wallis one-way analysis of variance procedure was employed to test the null hypothesis against the alternative hypothesis that the cell means were not all equal. This procedure was chosen in lieu of a parametric one-way analysis of variance (*ANOVA*) because the within-cell variances were so disparate (see fig. 2). While *ANOVA* is fairly robust for type I errors with respect to violations of the homogeneity of variance assumption if cell sizes are equal (Glass and Stanley [1970, p. 371]), the present case had a ratio of variances as large as 17:1 (cells A and D). This could result in a serious underestimation of a type I error. Since a total of four subjects failed to provide a response in this experiment, a total of eight observations were discarded at random in order to equalize the cell sizes ($n_k = 30$). The cell sizes were equalized so that an exact distribution-free multiple comparison procedure based on the Kruskal-Wallis rank sums could be employed (see below). The null hypothesis of no differences among the four cell means for part (2) was rejected ($H = 41.26, p < .00001$).

Next, Jonckheere's distribution-free test for ordered alternatives (Jonckheere [1954]) was used to test the null hypothesis against the alternative hypothesis:

$$H_a: \tau_1 \leq \tau_2 \leq \tau_3 \leq \tau_4$$

where τ_k = the average rank of the kth cell and where at least one of the inequalities is strict. The hypothesis of interest here is whether the subjects' estimates of the prevalence of fraud were a monotonic function of the size of the anchors they were provided. In other words, do larger anchors result in larger estimates of fraud? Jonckheere's test is preferred to the Kruskal-Wallis test "when the treatments are ordered and the experimenter expects a deviation from H_0 to be in a particular direction" (Hollander and Wolfe [1973, p. 122]). Two orders were tested: *ABCD* and *ABDC*. The null hypothesis was rejected for both orders; *ABCD*: $J = 2150, p < .0000001$; *ABDC*: $J = 2268, p < .0000000$.

A multiple comparison procedure based on Kruskal-Wallis rank sums was employed to isolate the cell means responsible for rejection of the null hypothesis above. Figure 3 reports the results of these tests. Note that all contrasts but *AB* and *CD* are significant.

The results reported for this experiment are roughly consistent with the anchoring and adjustment heuristic. Moreover, the results are incon-

Experimental Condition

	A(1%)		B(1%)		C(20%)		D(30%)	
	Subjects responding "yes" to part (1) (n=13)	Subjects responding "no" to part (1) (n=17)	Subjects responding "yes" to part (1) (n=7)	Subjects responding "no" to part (1) (n=23)	Subjects responding "yes" to part (1) (n=0)	Subjects responding "no" to part (1) (n=30)	Subjects responding "yes" to part (1) (n=0)	Subjects responding "no" to part (1) (n=30)
Mean	12.62	.76	45.00	3.22	--	40.47	--	35.33
Std. dev.	17.62	.66	31.49	1.73	--	39.47	--	34.41
Condition mean	5.90		12.97		40.47		35.33	
Condition std. dev.	12.82		23.03		39.47		34.41	

FIG. 2.—Experiment 1*B*: means and standard deviations for part (2) as a function of experimental condition and part (1) response.

Contrast	$\lvert R_{.u} - R_{.v} \rvert$	p
AB	18.33	NS
AC	48.43	<.0001
AD	47.00	<.0001
BC	30.10	<.005
BD	28.67	<.01
CD	1.43	NS

FIG. 3.—Experiment 1*B*: multiple comparisons based on Kruskal-Wallis rank sums.

sistent with the rival hypothesis that the differences in experiment 1*A* were due to differences in perceived materiality thresholds.

In addition, it is interesting to note that the magnitude of the anchoring effect did not seem to be monotone increasing with the size of the anchor. While on average the larger anchors resulted in higher frequency estimates, figure 3 indicates that there was no significant difference between the mean response of the subjects as the anchor increased from 200 in 1,000 to 300 in 1,000. This finding is consistent with research in psychophysics by Helson and Masters [1966] and Sarris [1967] who found nonmonotone anchoring effects.

Not only did the cell means for the subjects' part (2) responses increase with increasing anchors from condition *A* through condition *C*, but so did the standard deviations of the subjects' responses. However, both cell mean and standard deviation decreased between conditions *C* and *D*, suggesting that the effect of the anchor on subject response variability is a nonmonotonic function of the extremity of the anchor.

EXPERIMENT 2A

Experiments 2*A* and 2*B* asked the subjects to provide their own starting points or anchors in a perhaps more common auditing judgment involving the extensiveness of substantive testing. Condition *A* of experiment 2*A* follows:

You are conducting a routine year-end audit of a large closely-held tire wholesaler. Your firm has conducted its annual audit for the last three years. During that time no significant errors were discovered by the audit tests and unqualified opinions were issued. The management and employees of the client seem both competent and trustworthy. You are about to plan your substantive tests of the sales and collection cycle.

You review the client's system of internal control to identify the types of errors that could occur in the system and whether specific controls exist which would prevent or detect such errors. The first error you consider is "sales recorded for goods not shipped." Your review of the relevant controls for this error reveals the following:

(1) There is adequate control over back orders and partial shipments.
(2) There is adequate control of access to the shipping area.
(3) Prenumbered shipping documents are used.
(4) Sales invoices are matched to shipping documents.

(5) Shipping is segregated from the billing function.
(6) Regular shipping reports are prepared and reviewed monthly.
(7) Overdue accounts receivable are investigated.
(8) Unmatched sales invoices are independently reviewed and followed up.

Your compliance tests and observations of the system confirm the system is operating as just described.

1. Based on this information, indicate below the extensiveness of the substantive (detailed) tests you would perform in this engagement to test for the specific error "sales recorded for goods not shipped." (Circle one number.)

Minimum Audit Tests									Most Extensive Audit Tests
1	2	3	4	5	6	7	8	9	10

2. Suppose that instead of the preceding system description, you had received one that was identical except for the following (assume your compliance tests and observations of the system confirmed this description as well):
   (3) Prenumbered shipping documents are not used.
   (5) The accounts receivable billing clerk also performs clerical duties for the shipping department.
   Based on this different information, indicate below the extensiveness of the substantive (detailed) tests you would perform in this engagement to test for the specific error "sales recorded for goods not shipped." (Circle one number.)

Minimum Audit Tests									Most Extensive Audit Tests
1	2	3	4	5	6	7	8	9	10

The condition B subjects received the same brief client description as the subjects in condition A. However, instead of getting the eight positive control features, making a work extensiveness rating, having controls (3) and (5) become absent, and then making another work extensiveness rating given the new system, the condition B subjects were asked to make work extensiveness ratings where controls (3) and (5) where absent to begin with. Thus the condition B subjects started (and finished) where the condition A subjects ended up: making a work extensiveness rating in a situation where two control features were absent. The purpose of having the condition A subjects make a judgment in part (1), where the internal control system was strong, was to provide them with an anchor from which they had to adjust in making a judgment in part (2), where the internal control system was weaker. The *direction* such adjustments should take is clear: if the system is weaker, more extensive substantive tests should be performed. But the anchoring and adjustment heuristic predicts these adjustments will be of insufficient amounts.

The purpose of the condition B subjects was to provide a control group against which to compare the part (2) responses of the condition A subjects. Normatively, the responses of the condition B subjects should be the same as the responses of the condition A subjects on part (2). Since the judgments are of work extensiveness ratings given the same client and the same relevant internal control features.

The subjects were the same 50 practicing auditors who participated in experiment 1A above. The experiment was performed at the same sitting as experiment 1A and the ten other problems on probabilistic judgment. Again, the problems were self-paced and completed in 35–65 minutes,

Internal Control		Experimental Condition A	Experimental Condition B
	Stronger	3.08 (1.50)	
	Weaker	6.00 (2.40)	5.16 (2.56)

FIG. 4.—Experiment 2*A*: means as a function of experimental condition and internal control. Standard deviations in parentheses.

with the subjects randomly assigned to the two experimental conditions.

Results. Figure 4 reports the mean work extensiveness ratings by experimental condition and internal control system. Three things are notable from figure 4. First, the condition *A* subjects adjusted their work extensiveness ratings upward as the internal control system became weaker, which is consistent both with anchoring and adjustment and normative principles. Second, instead of being an insufficient adjustment, the average adjustment of the condition *A* subjects, if anything, was somewhat extreme compared with the condition *B* responses—6.00 versus 5.16—although the difference was not statistically significant ($p > .10$).

Of course, the responses of the condition *B* subjects which were used to provide a standard of reference did not constitute a definitive criterion for the "correct" work extensiveness rating in the experiment, since condition *B* subjects might have grossly underestimated the appropriate amount of audit work to perform. Thus, it cannot be definitively concluded that the condition *A* subjects adjusted insufficiently, sufficiently, or too much. The possibility remains, however, that a normatively irrelevant factor—an immediately preceding judgment of a different internal control system—did affect the judgments of the condition *A* subjects. Experiment 2*B* was designed to provide a more definitive test of this possibility.

EXPERIMENT 2B

The subjects randomly assigned to condition *A* of this experiment received the same materials and were asked to perform the same task as the subjects in condition *A* of experiment 2*A*. They first received the stronger internal control system and then received the weaker one. The subjects in condition *B* received the same materials but in the reverse order. That is, the judgment in part (1) of condition *B* was in response to an internal control system where control features (3) and (5) were absent; in part (2) the judgment was in response to a system where features (3) and (5) were present.

The research design employed in this experiment was a 2×2 repeated measures *ANOVA*. The factors were internal control strength (stronger, weaker) and the order in which the internal control systems were presented to the subjects (stronger first or weaker first). A significant main effect was expected for the internal control strength manipulation, since this is consistent with both normative principles and anchoring. However, neither anchoring nor normative principles predict a main effect for order. An order × internal control strength interaction was expected, given the findings in experiment 2*A*.

The subjects who participated in experiment 2*B* were the same 132 practicing auditors who participated in experiment 1*B* working this problem as one of the ten problems on probabilistic judgment. The subjects were randomly assigned to the two conditions, worked at their own pace, and completed the ten experiments in 30–60 minutes.

Results. The cell means and standard deviations for experiment 2*B* are presented in figure 5. Figure 6 is the sources table for the 2 × 2 repeated measures *ANOVA* performed on the data.

A significant main effect was observed for the internal control manipulation ($F = 331.74$, $p < .0001$), indicating that, on average, the subjects chose more extensive audit tests when internal control was weaker than when it was stronger. This finding is consistent with both normative principles and use of anchoring and adjustment heuristic.

No significant main effect was observed for order, which does not provide evidence either for or against the use of anchoring and adjustment.

Finally, a significant order × internal control strength interaction was observed ($F = 15.98$, $p < .0002$). This suggests that the effect of a particular level of internal control depended on whether it was provided first or last to the subjects. As shown in figure 5, the subjects in condition *B* start with weaker internal control and indicate less extensive work than the condition *A* subjects who finish with the weaker internal control (5.44 versus 6.29). Note also that the subjects in condition *A* start with stronger internal control and indicate less extensive work than the

		Internal Condition: A Stronger	Internal Condition: B Weaker	
Order Condition	A (stronger first)	2.78 (1.41)	6.29 (2.25)	4.54
	B (weaker first)	3.19 (2.01)	5.44 (2.20)	4.32
		2.99	5.87	

FIG. 5.—Experiment 2*B*: cell means and standard deviations as a function of internal control strength and order. Standard deviations in parentheses. Order condition factor manipulated between subjects. Internal condition factor manipulated within subjects.

Source of Variation				
Between subjects	131			
Order	1	3.28	0.50	.4792
Subject within groups	130	6.52		
Within subjects	132			
Internal control strength	1	545.72	331.74	.0001
Order x internal control strength	1	26.28	15.98	.0002
Internal control strength x subject within groups	130	1.65		

FIG. 6.—Experiment 2*B*: *ANOVA* sources table.

condition *B* subjects who finish with the stronger internal control (2.78 versus 3.19).

Although experiment 2*B* also does not provide us with a definitive criterion for the correct work extensiveness ratings in each internal control strength condition, the order in which the ratings are made should not have affected the auditors' judgments. The results indicate that they did.

The final two experiments provide results on whether the anchoring and adjustment heuristic biases evaluations of conjunctive and disjunctive events.

EXPERIMENT 3A

The subjects in condition *A* of experiment 3*A* performed the following task:

As part of your regular year-end audit of a publicly-held client, you must make an estimate of the probability of success of its proposed new product line. The client has experienced financial difficulties during the last few years and—in your judgment—a successful introduction of the new product line is necessary for the client to remain a going concern.

There are five steps *all* of which are necessary for successful introduction of the product: (1) successful labor negotiations between the construction firms contracted to build the necessary addition to the present plant and the building-trades unions, (2) successful defense of patent rights, (3) product approval by the FDA, (4) successful negotiation of a long-term raw materials contract with a foreign supplier, and (5) successful conclusion of distribution contract talks with a large national retail distributor.

In view of the circumstances, you contact outside experts who have provided your audit firm with reliable estimates in the past. The labor relations expert estimates that there is an 80% chance of successfully concluding labor negotiations before the strike deadline. Legal counsel advises that there is a 90% chance of successfully defending patent rights. The experts in the remaining two areas estimate the probability of successfully resolving (a) the raw materials contract and (b) the distribution contract talks to be 90% in each case. Assume these estimates are reliable.

1. What is your assessment of the probability of successful product introduction? (Circle one number.)

.50 or less	.55	.60	.65	.70	.75	.80	.85	.90	.95	1.00

2. Assume that you have completed your year-end work and are about to issue an opinion on this client's financial statements. If the new product line is successfully introduced you will be satisfied that the statements are in accordance with GAAP applied on a basis consistent with the previous year. If the new product line is not successfully

introduced, however, there is some question about the firm's ability to survive beyond the next period. This is of importance because the client has many special purpose assets whose liquidation values are considerably lower than their present book values. What opinion would you express on their financial statement? (Circle one number.)

Unqualified 1
Qualified (subject to uncertainties) 2
Disclaimer 3
Other (please specify) 4

Notice that the subjects in condition A were given the probabilities of five events which were necessary and sufficient for the successful introduction of the new product line.

The subjects in condition B were asked to make the same two judgments as the subjects in condition A, but were told that there were only three events necessary and sufficient for successful introduction of the new product line. These three events and their estimated probabilities of occurrence were identical to the first three events in condition A.

Note that the probability assessment the auditors provided in part (1) of both experimental conditions is the probability of a *conjunctive* event. All five (three) of the events in condition A (B) must obtain for the product to be successfully introduced. Normatively, the mean probability assessment provided in part (1) should be lower for the condition A subjects than for the condition B subjects, since the addition of two necessary conditions each with probability of occurrence less than 1.0 must reduce the probability of successful product introduction. An exception would be where both additional events are perfectly correlated with one or more linear combinations of the other three factors. This exception appears unlikely in this experiment.

If the five (three) events are assumed independent of one another, then the overall probability of successful product introduction is simply the product of the individual events: $.80 \times .90 \times .95 \times .90 \times .90 = .554$ ($.80 \times .90 \times .95 = .684$). However, it is unlikely that the events are independent. For example, the successful conclusion of distribution contract talks would seem to be contingent on FDA product approval. Nevertheless, .554 (.684) should set a floor for the subjects' assessments in condition A (B).

The subjects were the same 50 practicing auditors who participated in experiments 1A and 2A and, as before, were randomly assigned to the experimental conditions.

Results. The results of part (1) of this experiment are reported in figure 7. The null hypothesis of no difference between cell means was rejected ($F = 4.052, p < .007$). This finding is consistent with normative principles: as elementary necessary events with probability of occurrence less than one are added, the probability of the conjunctive event decreases as long as the elementary events are not completely redundant with one or some linear combination of the preceding events.

The results in figure 7 indicate that the mean responses for both conditions exceeded the floors for each—.644 versus .544 for A ($t = 3.00$,

$p < .005$) and .752 versus .684 for B ($t = 2.12$, $p < .025$). This does not mean, however, that the adjustments were insufficient, since the floor probabilities for each condition were computed assuming independence of the elementary events. If the elementary events are not independent, the normative response would be higher than the floor for each condition. Since the "true" relationship among the elementary events is unknown, it cannot be concluded from these part (1) results that the adjustments were insufficient.

The results of part (2) of this experiment appear in figure 8. This portion of the task asked the auditors for a designation of the opinion they would express on the client's financial statements. The subjects were told that they should be satisfied that the statements would be in accordance with *GAAP* if the new product line is successfully introduced. However, if the new product line is not successfully introduced, then the firm's ability to survive beyond the next period would be in doubt.

Normatively, the selection of an opinion should be related to the probability of successful introduction of the new product line (the conjunctive event). The higher the probability of a successful introduction, the more appropriate an unqualified opinion would be in each experimental condition. A lower probability of successful introduction should elicit a qualified opinion or a disclaimer.

An examination of figure 8 does not confirm this relationship. If anything, it appears that the opinions expressed by the condition A subjects (five necessary events) were a little more favorable than those

Experiment Conditions	
A (5 events)	B (3 events)
.644	.752

FIG. 7.—Experiment 3A: mean response to part (1) as a function of experimental condition.

		Experimental Condition	
		A* (5 events)	B (3 events)
Opinion	Unqualified	3	1
	Qualified	17	23
	Disclaimer	1	1
	Other	3	0

FIG. 8.—Experiment 3A: frequency of response in part (2) as a function of experimental condition. * One subject in condition A did not respond to this part of the experiment.

expressed by the subjects in condition B (three necessary events). This finding should be interpreted with care for at least two reasons. First, the number of observations is small ($n = 49$). A Kolmogorov-Smirnov test could not reject the null hypothesis of no difference between conditions ($Z = .427$, $p > .99$). Second, since probabilities are measured on a continuous scale and audit opinions are discrete, not every change in an auditor's probability assessments will (or should) lead to a change in the audit opinion he would render. Thus, audit opinions can be expected to be only somewhat sensitive to changes in probability assessments.

An examination of individual responses revealed that three condition A subjects who selected unqualified opinions in part (2) assessed probabilities of success in part (1) of .50, .75, and .80, respectively. A condition B subject choosing an unqualified opinion assessed a probability of success of .70. Further investigation revealed quite large individual differences in probability thresholds for each level of opinion. For example, one subject in condition A selected an unqualified opinion for an assessed probability of success of .50, while another condition A subject chose a disclaimer for an assessed probability of success of .80.

Experiment 3B was designed to replicate and extend the findings of Experiment 3A.

EXPERIMENT 3B

Conditions A and B of experiment 3B were virtually identical to conditions A and B of experiment 3A. Conditions D and C changed conditions A and B into disjunctive forms. For example, subjects in condition D received the following:

As part of your regular year-end audit of a publicly-held client, you must make an estimate of the probability of success of its proposed new product line. The client has experienced financial difficulties during the last few years and—in your judgment—a successful introduction of the new product line is necessary for the client to remain a going concern.

Any one of the following five occurrences will prevent successful introduction of the new product line: (1) unsuccessful labor negotiations between the construction firms contracted to build the necessary addition to the present plant and the building-trades unions, (2) unsuccessful defense of patent rights, (3) failure to obtain product approval from the FDA, (4) failure to successfully negotiate a long-term raw materials contract with a foreign supplier, and (5) failure to successfully conclude distribution contract talks with a large national retail distributor.

In view of the circumstances you contact outside experts who have provided your audit firm with reliable estimates in the past. The labor relations expert estimates that there is a 20% chance that labor negotiations will not be successfully resolved before the strike deadline. Legal counsel advises that there is a 10% chance that the patent rights defense will be unsuccessful. The expert on FDA product approvals estimates the probability of failing to obtain approval at 5%. The experts in the two remaining areas estimate the probabilities of failing to resolve (a) the raw materials contract and (b) the distribution contract talks to be 10% in each case.

Assume these estimates are reliable.

1. What is your assessment of the probability that the product introduction will fail? (Circle the number below which is closest to your probability estimate.)
   0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1.0
2. Assume that you have completed your year-end work and are about to issue an opinion

on this client's financial statements. If the new product line is successfully introduced you will be satisfied that the statements are in accordance with GAAP applied on a basis consistent with the previous year. If the new product line is not successfully introduced, however, there is some question about the firm's ability to survive beyond the next period. This is of importance because the client has many special purpose assets whose liquidation values are considerably lower than their present book values. What opinion would you express on their financial statements? (Circle one number.)

Unqualified 1
Qualified (subject to uncertainties) 2
Disclaimer 3
Other (please specify)__________ 4

The disjunctive formulations were obtained both by changing the wording of the experimental materials where necessary to reflect the disjunctive formulation and by taking the complements of the probabilities of the elementary events in the conjunctive case to get the probabilities of the elementary events in the disjunctive case (e.g., P(unsuccessful defense of patent rights) $= 1 - P$(successful defense of patent rights) $= 1 - .90 = .10$).

The only other change between experiment $3A$ and experiment $3B$ was the response scale to part (1) in each experimental condition. The response scales in experiment $3B$ range from 0 to 1.0 in increments of .1, whereas in experiment $3A$ the scale had a floor of ".50 or less," a ceiling of 1.0, and was graduated in increments of .05. The change was made to minimize any possible subject response bias due to an aversion to using the extremes of the response scale. This might have produced an artifactual "floor effect" in experiment $3A$.

The subjects who participated in this experiment were the same 132 practicing auditors who participated in experiments $1B$ and $2B$, again completed at the same sitting as these experiments. The subjects were randomly assigned to the four experimental conditions.

Results. The cell means for part (1) of this experiment are presented in figure 9. The mean probabilities in the disjunctive row have been transformed by subtracting from 1.0. This states both rows in comparable terms (P(Success) $= 1 - P$(Failure)). A two-way *ANOVA* was performed to test for the significance of the conjunctive-disjunctive problem formulation manipulation, the three versus five necessary elementary event

		Number of Necessary Events: 5	3
Problem Formulation	Conjunctive	A .638	B .741
	Disjunctive†	D .625	C .730

FIG. 9.—Experiment $3B$: mean responses to part (1) as a function of experimental condition. Capital letters in cells refer to experimental condition. † Probabilities in this row have been transformed by taking complements of raw scores.

manipulation, and the problem formulation × number of elementary events interaction. The *ANOVA* sources table is presented in figure 10. A significant main effect for the number of elementary events was obtained ($F = 8.749$, $p < .005$). As in the preceding experiment, those subjects in the experimental conditions where there were five necessary elementary events for success assessed (on average) a lower probability of successful product introduction than those in the experimental conditions where there were only three necessary elementary events. This, again, is in conformance with normative principles.

There was no significant main effect for problem formulation, which is also in conformance with normative principles. The conjunctive and disjunctive cases were simply different representations of substantively identical problems. Figure 11 represents a graphic representation of the subject responses in part (1) as a function of experimental condition. Note that the mean responses in both the conjunctive and disjunctive cases are quite close. The subjects in the conjunctive treatment groups assessed a slightly higher (but statistically insignificant) probability of successful product introduction than those in the disjunctive treatment groups.

We also tested whether the effect of the number of elementary events on the auditors' probability assessments was contingent upon whether the problem formulation was conjunctive or disjunctive. This would imply a problem formulation × number of elementary events interaction. The two-way *ANOVA* test for the significance of this effect was not significant ($F = .247$). Normatively, of course, such an interaction should not occur in this experiment.

The results of part (1) of this experiment were encouraging. The conjunctive versus disjunctive formulation of the problem should not have and did not affect the probability assessments. In addition, a significant main effect for number of elementary events should have occurred, and it did. All mean responses to part (1) are thus in conformance with the normative principles.

An examination of figure 12, however, reveals a similar phenomenon to that observed in the preceding experiment. Presented are the responses

Source of Variation	df	MS	F	P	ω^2
Problem Formulation	1	239.25	.839	--	--
Number of Elementary Events	1	2493.87	8.749	<.005	5.7%
Problem Formulation x Number of Elementary Events	1	70.50	.247	--	--
Within Cells	124	285.05			

FIG. 10.—Experiment 3*B*: *ANOVA* sources table.

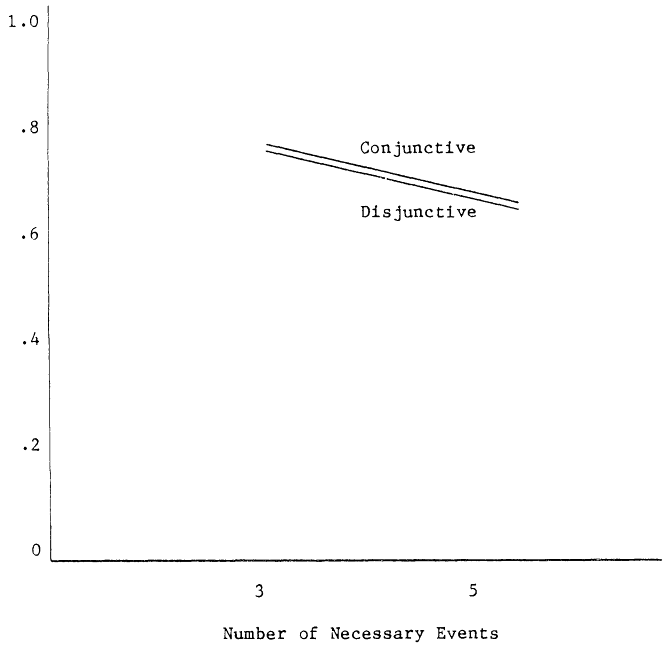

FIG. 11.—Experiment 3*B*: graphic representation of probability assessments in part (1) as a function of problem formulation and number of necessary events.

		Number of Necessary Events: 5	3
Conjunctive Formulation	Unqualified	8	7
	Qualified	22	21
	Disclaimer	2	3
	Other	0	1
Disjunctive Formulation	Unqualified	8	2
	Qualified	18	23
	Disclaimer	5	4
	Other	1	3

FIG. 12.—Experiment 3*B*: frequency of response in part (2) as a function of experimental condition.

to part (2) of the experiment as a function of experimental condition. Although no differences should be observed between cells in the same column of figure 12, 15 auditors in the conjunctive conditions specified an unqualified opinion versus only 10 in the disjunctive conditions.

Nevertheless, a Kolmogorov-Smirnov two-sample test was unable to reject the null hypothesis of no differences between the conjunctive and disjunctive groups ($Z = .446, p > .98$). Two Kolmogorov-Smirnov tests were also employed to test for conjunctive versus disjunctive group differences in the three and five necessary event conditions. Again, no significant differences were found (with three: $Z = .679, p > .74$; with five: $Z = .463, p > .98$).

Any differences between cells in the same row should show a pattern of more unqualified opinions in the three elementary event condition cells than in the five event cells. Figure 12 again indicates that this is not the case. This discrepancy is most dramatic in the disjunctive row. Only two subjects in the three elementary event condition selected an unqualified opinion, while eight subjects in the five elementary event condition selected an unqualified opinion. This is contrary to the notion that an unqualified opinion should be less likely as the probability of success decreases (or the probability of failure increases). However, these differences were not significant either for tests (Kolmogorov-Smirnov) of no differences in the distribution of the three and five event condition ($Z = .627, p > .82$) or for tests for three versus five necessary event group differences within conjunctive and disjunctive conditions ($Z = .207, p = 1.00$ and $Z = .804, p > .53$).

As in the preceding experiment, individual data were examined for those subjects specifying an unqualified opinion in part (2). Figure 13 reports the probability assessments by experimental conditions for those subjects choosing an unqualified opinion. On average, the mean probabilities of success assessed by these subjects were greater than the mean probabilities of success assessed by all subjects in each experimental condition (see fig. 9). Once again, however, there were large individual

	Number of Necessary Events: 5		3	
Conjunctive Formulation	.50	.80	.80	.80
	.80	.80	.90	.90
	.80	.60	1.00	.50
	.50	.90	.90	
	$\bar{p}_a$ = 71.25		$\bar{p}_b$ = 82.86	
Disjunctive Formulation	.90	.60	.80	.90
	.80	.60		
	.70	.80		
	.80	.50		
	$\bar{p}_d$ = 71.25		$\bar{p}_c$ = .85	

FIG. 13.—Experiment 3*B*: individual probability assessments of subjects selecting unqualified opinions.

differences in probability thresholds for opinions across subjects. Subjects who agreed on the probability of success frequently disagreed on the choice of an opinion. For example, the probabilities .5, .6, .7, and .8 were each associated with unqualified opinions, qualified opinions, and disclaimers. The probability .9 was selected by subjects who chose unqualified and qualified opinions. Whether these findings reflect differences in risk preferences across individuals or difficulties in making qualitative decisions from quantitative data is not clear.

In summary, experiments 3*A* and 3*B* both revealed part (1) results (which involved subjective probability assessments) which were consistent with normative principles, but part (2) results (which involved audit opinion selection and which should be a function of the part (1) responses) which were quite idiosyncratic across subjects.

One possible explanation for those findings lies in the different nature of the responses in part (1) and part (2) and their relationship to the other experiments conducted in both series of experiments. The three experiments on anchoring and adjustment in each of the series reported here were administered with a number of other experiments. Many were clearly statistical in nature which may have sensitized the subjects to the use of normative procedures where they are easily applied, for example, in part (1) of experiment 3. Part (1) asked the auditors to provide a subjective but quantitative probability assessment which was clearly a function of other numbers provided in the problem.

Part (2), however, asked them to select an audit opinion, a judgment which is not so clearly a function of the elementary event probabilities. Normative principles here were perhaps not so easy to apply. This could account for the inconsistency in opinion selection across assessed probabilities of success.

This implication is important, since opinions, not probability assessments, are of primary importance in auditing. A probability assessment is important only insofar as it affects a decision. Certainly more research on the relationship between probability assessments and decisions is warranted.

Discussion

The results of these experiments indicate that auditors sometimes make judgments that are in violation of normative principles of decision making, but that these violations cannot always be accounted for by the anchoring and adjustment heuristic. Some other heuristic(s), as yet unidentified, may also be at work.

The first pair of experiments involved a test of anchoring in estimating the base rate of management fraud. While these were easily the least familiar scenarios of those presented to the subjects in the experiments, they still involved an issue of importance to auditors. Many judgments, like this one, involve consideration of the base rate of low-probability high-consequence events. Yet hard base-rate data for these judgments are rarely available to auditors. (We attempted in vain to get the base rate of management fraud from four different Big Eight firms and the SEC.) The important implication of this pair of experiments was how sensitive the auditors' estimates of fraud prevalence were to meaningless and innocuous anchors.

The second pair of experiments raised the interesting issue of the prevalence and importance of order effects in internal control evaluation and audit program planning. Here, it was found that the extensiveness of audit work planned in a given internal control system was partly a function of the immediately preceding control system.

Finally, the third pair of experiments suggests that some auditors may be selecting audit opinions contrary to normative principles and their own beliefs as incorporated in subjective probability assessments. This suggests that auditors may use fundamentally different decision strategies in making quantitative judgments versus qualitative judgments. Auditors (and others) may have more difficulty using normative principles to make qualitative judgments, which is somewhat disturbing since many important audit judgments are qualitative in nature.

The results presented above are often at odds with the psychological research on the anchoring and adjustment heuristic. While the evidence presented here suggests that auditors do not always act in conformance with normative principles, the anchoring and adjustment heuristic cannot account for a majority of the violations. What are some possible explanations for such a discrepancy between the present research and the psychological research?

First, the anchoring and adjustment heuristic is conceptually very vague. Its definition is purely descriptive and at a very superficial level. A vague conceptual definition can lead to a limitless number of operational definitions with no clear a priori basis for selection. The results of the research reported here are likely to be representative of the typical outcomes: sometimes anchoring happens, and sometimes it does not.

Second, humans approach few things as if they were unique. "Instead, they are processed through preexisting systems of schematized and abstracted knowledge—beliefs, theories, propositions, and schemes" (Nisbett and Ross [1980, p. 7]). These knowledge structures provide humans with an important interpretive framework—a framework which imparts meaning to information, resolves ambiguity, and augments the information given. Knowledge structures permit the individual to infer (probably not consciously) additional properties of objects or events. For example, in experiments $2A$ and $2B$, the auditors might have believed (from experience) that when internal control changed, other things which are important but not mentioned in the experiments changed as well. The order effect observed might, then, be quite adaptive. Well-developed knowledge structures are a major difference between college sophomores working general knowledge tasks and auditors working problems related to their expertise. Research which simultaneously considers heuristics and knowledge structure has only begun. (See Einhorn and Hogarth [1981] for an up-to-date review.)

In this respect, we might note that anchoring and adjustment could account for the observed behavior only in experiments 1*A* and 1*B*—the two experiments where auditor familiarity was clearly lowest and knowledge structures least well developed. It is apparent that some other process was at work in the other experiments. More research is needed to identify a broader set of auditor heuristics and the decision settings in which they are likely to be employed.

The primary reason auditing researchers are concerned with heuristics is that such studies might improve the quality of audit decision making. The type of research discussed in this paper is merely a first step in this direction. Before decision making can be improved, it is necessary to know in what areas it is deficient. The use of heuristics per se does not provide conclusive evidence of deficiency. Heuristics may work quite well in many tasks. In fact, Thorngate [in press] presents evidence from a simulation showing how heuristics lead to optimal responses across a variety of tasks. Even where heuristics lead to systematic, costly errors, the cost of adopting an alternative procedure which eliminates the error might outweigh the incremental benefit.

Nevertheless, we feel that additional research on the use of heuristics in audit decision making is advisable, especially where this use is likely to result in audit judgment deficiencies. Research on improving the quality of intuitive judgments where they have been found wanting has barely begun. (See Kahneman and Tversky [in press] for an example of a judgment aid.) Efforts in this area are likely to be impeded until the psychological theory of decision making is better formulated. Such a formulation must include a concern for task structure, the cognitive representative of the task, and the information-processing capabilities of the organism (Einhorn and Hogarth [1981]).

A well-known example provided by Tversky and Kahneman [1980] illustrates how cognitive representation can affect decision responses.

Problem 1. Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the consequences of the programs are as follows:
If program *A* is adopted, 200 people will be saved.
If program *B* is adopted, there is a 1/3 probability that 600 people will be saved, and a ⅔ probability that no people will be saved.
Which of the programs do you favor?

Seventy-six percent of the subjects chose program A. A second group of subjects received an equivalent problem, but with a different formulation of the two programs:

Problem 2.
If program *C* is adopted 400 people will die.
If program *D* is adopted there is a ⅓ probability that nobody will die, and a ⅔ probability that 600 people will die.

Eighty-seven percent of the subjects chose program *D*. Note that programs *A* and *C* are identical, as are programs *B* and *D*. Merely changing the descriptions of the outcomes from lives saved to lives lost produced a pronounced change in preference.

This sensitivity of cognitive representations of tasks to normatively irrelevant aspects of tasks might have important implications for auditor education. Firms may wish to structure audit decision problems in such a way that leads auditors to adopt cognitive representatives of the problem conducive to decision making in conformance with normative principles. We are currently investigating this possibility.

REFERENCES

EDWARDS, W. "Bayesian and Regression Models of Human Information Processing—A Myopic Perspective." *Organizational Behavior and Human Performance* (November 1971): 639–48.

EINHORN, H., AND R. HOGARTH. "Behavioral Decision Theory: Processes of Judgment and Choice." *Annual Review of Psychology* (1981): 53–88.

GIBBINS, M. "Human Inference, Heuristics and Auditor's Judgment Process." CICA Audit Research Symposium, 1977.

GLASS, G., AND J. STANLEY. *Statistical Methods in Education and Psychology*. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

HELSON, H., AND H. MASTERS. "A Study of Inflection-Points in the Locus of Adaptation-Levels as a Function of Anchor-Stimuli." *American Journal of Psychology* (1966): 400–408.

HOGARTH, R. "Cognitive Processes and the Assessment of Subjective Probability Distributions." *Journal of the American Statistical Association* (1975): 271–94.

HOLLANDER, M., AND D. A. WOLFE. *Nonparametric Statistical Methods*. New York: Wiley, 1973.

JONCKHEERE, A. R. "A Distribution-Free K-Sample Test against Ordered Alternatives." *Biometrika* (1954): 133–45.

JOYCE, E., AND G. BIDDLE. "Are Auditors' Judgments Sufficiently Regressive?" *Journal of Accounting Research* (forthcoming).

KAHNEMAN, D., AND A. TVERSKY. "Subjective Probability: A Judgment of Representativeness." *Cognitive Psychology* (July 1972*a*): 430–54.

———. "On Prediction and Judgment." *Oregon Research Institute Research Bulletin* (1972*b*).

———. "On the Psychology of Prediction." *Psychological Review* (1973): 237–51.

———. "Intuitive Prediction: Bases and Corrective Procedures." *Management Science* (in press).

KRANTZ, D., AND J. MIYAMOTO. "Non-Bayesian Belief Structures." Working paper, University of Michigan, 1980.

MILLER, G. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review* (March 1956): 81–97.

NEWELL, A., AND H. SIMON. *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

NISBETT, R., AND L. ROSS. *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, N.J.: Prentice-Hall, 1980.

PAYNE, J. "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis." *Organizational Behavior and Human Performance* (1976): 366–87.

SARRIS, V. "Adaptation-Level Theory: Two Critical Experiments on Helson's Weighted-Average Model." *American Psychologist* (1967): 331–44.

SHAFER, G. *A Mathematical Theory of Evidence*. Princeton, N.J.: Princeton University Press, 1976.

SIMON H., AND A. NEWELL. "Human Problem Solving: The State of the Theory in 1970." *American Psychologist* (February 1971): 145-59.

———, AND R. K. SUMNER. "Patterns in Music." In *Formal Representation of Human Judgment*, edited by B. Kleinmuntz, pp. 219-50. New York: Wiley, 1968.

SLOVIC, P., B. FISCHHOFF, AND S. LICHTENSTEIN. "Behavioral Decision Theory." *Annual Review of Psychology* (1977): 1-39.

———, AND S. LICHTENSTEIN. "Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment." *Organizational Behavior and Human Performance* (June 1971): 649-744.

SWIERINGA, R., ET AL. "Experiments in the Heuristics of Human Information Processing." *Studies on Human Information Processing in Accounting: 1976.* Supplement to *Journal of Accounting Research* 14: 159-87.

THORNGATE, W. "Efficient Decision Heuristics." *Behavioral Science* (in press).

TVERSKY, A., AND D. KAHNEMAN. "Belief in the Law of Small Numbers." *Psychological Bulletin* (1971): 105-10.

———. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* (1973): 207-32.

———. "Judgment under Uncertainty: Heuristics and Biases." *Science* (1974): 1124-31.

———. "Causal Schemata in Judgments under Uncertainty." In *Progress in Social Psychology*, Lawrence Erlbaum Associates, 1977.

———. "The Framing of Decisions and the Rationality of Choice." Working paper, 1980.

UECKER, W., AND W. KINNEY, JR. "Judgmental Evaluation of Sample Results: A Study of the Type and Severity of Errors Made by Practicing CPAs." *Accounting, Organization and Society* (1977): 269-75.