Psychological Bulletin
1991, Vol. 109, No. 3, 472-489

0033-2909/91/$3.00

Heuristics and Biases: Expertise and Task Realism in Auditing

James F. Smith and Thomas Kida
Department of Accounting and Information Systems
University of Massachusetts at Amherst

Important issues concerning generalizations drawn from heuristics research that depict humans as highly prone to a number of judgmental biases have been raised in recent years. Some researchers have argued that the tasks and subjects examined do not adequately represent the contexts to which generalizations are made, citing evidence of improved performance for expert judges working realistic, familiar tasks. This article presents additional evidence on these issues by reviewing experimental research on certain heuristics and biases in professional auditor judgments. In general, the significance of generalizability issues is underscored. Although certain biases seemed to persist in auditor judgments, they were often mitigated or modified in studies that used expert subjects performing familiar tasks.

An extensive body of research in psychology and related disciplines has examined a number of heuristics and biases in human decision making. Although this line of research has provided many insights into human information-processing abilities and limitations, important issues concerning the generalizability of the findings have been raised (Berkeley & Humphreys, 1982; Ebbesen & Konecni, 1980; Edwards, 1983; Einhorn, 1976; Fischoff, 1982, 1987; Funder, 1987; Hogarth, 1981). A central theme is that the often cited impression that individuals readily fall prey to a number of judgmental biases may have been overgeneralized when one considers knowledgeable decision makers performing tasks that are more realistic and familiar to them.

Edwards (1983) questioned the degree to which experimental tasks and subjects are representative of the contexts to which generalizations are made. He argued that the "generalized normal adult human mind" is not the appropriate target for research on human intellectual performance, noting that minds vary, that tools can help, and that expertise (presumably in the subject matter of the task and in probability itself) can also help (p. 511). In essence, to effectively study judgment heuristics and biases, Edwards argued that we must examine judgments of experts, working on tasks that are realistic and familiar to them.

In a review and appraisal of laboratory efforts to debias judgments, Fischoff (1982) noted that responsibility for biased judgments can be attributed to faulty tasks, faulty judges, or a judge–task mismatch. That is, although experimental findings may be due to faulty decision processes of the judge, they may also be caused by tasks that are unfair or misunderstood by the judge, or by a mismatch between the task and judge, where the task is fair and the judge knowledgeable, but the task may not be structured in a way that allows the judge to use existing cognitive skills to their best advantage (p. 427). Therefore, Fischoff pointed to the importance of developing tasks that are readily understood by subjects or finding subjects that readily understand the tasks (i.e., experts; also see Fischoff, 1987).

In addition, Hogarth (1981) noted that although much of the research on judgmental heuristics has focused on discrete instances of judgment that lack any form of meaningful feedback, important decision tasks are often continuous in nature, with feedback that allows the individual to make corrective adjustments, perhaps resulting in more effective judgments or choices (for similar arguments, see Klayman, 1984). Kleinmuntz (1985) and Klayman (1983) also provided evidence suggesting that, in dynamic choice tasks, corrective feedback provides a rich source of additional information. Similar issues concerning the generalizability of laboratory research have also been raised by Berkeley and Humphreys (1982) and Ebbesen and Konecni (1975, 1980); and the use of normative models to define the standard for characterizing optimal judgments has been questioned by L. Cohen (1981) and Berkeley and Humphreys (1982).

Finally, there is evidence indicating that simple judgment heuristics can be efficient and effective. For example, Kleinmuntz (1985) and Thorngate (1980) used computer simulations to compare the judgments/choices of a range of decision strategies. The results suggest that under certain conditions simple heuristics can perform well, and increasingly complex strategies provide little improvement (Kleinmuntz, 1985, p. 696). Paquette and Kida (1988) also found that individuals were more efficient when using a reduced-information choice strategy such as elimination by aspects than when using a more complex strategy such as additive compensatory in certain decision contexts.

Thus, the often-cited conclusion that depicts humans as highly prone to judgmental biases may have been overgeneralized. In fact, an emerging view is that, although individuals may use heuristics that lead to biases in certain situations, they are able to adapt to different decision environments and can be effective decision makers when operating in natural, familiar contexts. Proponents of a more optimistic view of human cognitive abilities note that some studies report "good" performance in certain decision contexts (Berkeley & Humphreys,

Correspondence concerning this article should be addressed to James F. Smith, Department of Accounting and Information Systems, School of Management, University of Massachusetts, Amherst, Massachusetts 01003.

1982; Christensen-Szalanski & Beach, 1984; L. Cohen, 1981; Edwards, 1983). Often cited are studies on the judgments of weather forecasters (Murphy & Winkler, 1974, 1977), livestock judges (Phelps & Shanteau, 1978), and physicians (Christensen-Szalanski, Diehr, Bushyhead, & Wood, 1982; Christensen-Szalanski, Beck, Christensen-Szalanski, & Koepsell, 1983), which suggest that experts, performing reasoning tasks that are familiar to them, are often not subject to the same kinds of cognitive limitations and biases found when students perform abstract, unfamiliar tasks. Furthermore, Christensen-Szalanski and Beach (1984) categorized a sample of studies as to whether they reported good or poor performance, and they found that those reporting poor performance were cited significantly more often than those reporting good performance, and that instances of poor performance most often used students performing abstract tasks.

Accordingly, Kleinmuntz (1985, p. 682) noted that estimates in the literature of the prevalence of biased decisions are based on a nonrepresentative sample of task environments. Therefore, to evaluate the costs and benefits associated with cognitive heuristics, research should examine the individuals whom we wish to generalize, performing tasks that the decision makers perform regularly.

The purpose of this article is to (a) review a substantial body of research that has examined professional auditor judgments for the presence of heuristics and biases, and (b) present evidence on the effects of task realism and subject expertise on heuristics and biases by comparing studies in the audit-judgment literature that differ on these dimensions.

The audit-judgment literature is particularly relevant to examining these issues. Many audit judgments are similar in nature, at least fundamentally, to the judgments examined in other heuristics research. As a matter of "routine," auditors make probability assessments and probabilistic inferences; they design and implement sampling procedures and analyze sample data to assess characteristics of underlying populations; they gather and evaluate information from differentially reliable sources; and they make diagnostic judgments requiring the evaluation of alternative hypotheses by considering complex information sets containing both confirming and disconfirming evidence. Thus, auditors appear susceptible to the same biases found in the heuristics literature. Importantly, although some of the audit research has investigated students' judgments in simplistic tasks, much of it has examined experienced professionals in the performance of familiar, job-related tasks. Consequently, comparisons of these results can be made to examine the potential effects of task realism and subject expertise on the presence of judgment heuristics and biases.

The potential synergistic effects of attending to evidence from both generic and applied judgment research on the progress of both fields is recognized by psychologists and audit-judgment researchers. For example, Ashton, Kleinmuntz, Sullivan, and Tomassini (1988) noted that applied research, such as audit-judgment research, has relied on paradigms developed by cognitive psychologists, Bayesian statisticians, economists, decision theorists, and others concerned with generic human information-processing issues. Conversely, evidence from research in such applied settings may serve to illuminate the robustness of generic research and thereby also aid progress in basic fields.

In general, the evidence strongly underscores the fundamental issues raised by Edwards (1983), Fischoff (1982, 1987), and others. Important differences in the findings of some studies using experts in the performance of familiar, job-related tasks and those using less expert subjects or less realistic tasks are apparent. For example, the evidence indicates that the extent of the biases investigated is often less when experts perform job-related tasks. In addition, a behavior not typically found in other heuristics research has been observed frequently in audit-judgment research. This finding, consistent with Klayman and Ha's (1987) view that judgmental heuristics may be used as an adaptive response to situational variables, suggests that "specialized" heuristics may be used for tasks within an expert's domain, whereas "basic" cognitive heuristics may be used in situations in which the individual lacks the information or expertise to indicate a strategy better suited to a particular task.

It is important to note, however, that biases found in studies using subjects with little or no expertise or artificial, abstract tasks often persist, to some extent, in the judgments of experienced professional auditors, pointing to the importance of basic research in identifying issues of potentially serious concern in professional judgment contexts. Furthermore, although the tasks used in audit-judgment studies are often more representative of those actually performed by the subjects compared with tasks commonly used in heuristics research, the experiments are not identical to the tasks and environments in which auditors actually perform because real-world audit environments are quite complex.

Numerous heuristics and biases associated with various aspects of human-judgment processes have been reported in the psychology literature (for an extensive summary, see Hogarth, 1987, chapter 10). However, we limited the scope of our review to those heuristics and biases that have received sufficient attention in the audit-judgment literature to allow us to reasonably draw inferences concerning the effects of task realism and subject expertise on their use or presence. Accordingly, in the remaining sections we review and compare relevant experimental research concerning the judgment heuristics anchoring and adjustment, representativeness, and confirmatory strategies, along with biases that have been associated with these heuristics, including poor calibration and coherence in probability assessments, base-rate neglect, insensitivity to source reliability and data predictability, insensitivity to sample size, and the confirmatory bias. The extent of task realism and subject expertise in the tasks is indicated, and the importance of these factors to the presence of biases is discussed. Table 1 summarizes the characteristics and findings of the studies reviewed in the order in which they are discussed in this article.

Anchoring and Adjustment

Individuals using the anchoring and adjustment heuristic focus on an initial value (i.e., anchor) in the decision setting and, based on the available information, adjust from that value to arrive at a judgment. The anchor may be suggested by factors such as the task, a preliminary calculation, or experience. Research indicates that adjustments from the anchor, although

(*text continues on page 476*)

Table 1
Summary of Studies Investigating Heuristics and Biases in Audit-Judgment Research

Study	Subjects	Bias or heuristic	Task description	Summary
Joyce & Biddle (1981a)	Exp. 1A: 50 auditors; Exp. 1B: 132 auditors; Exp. 2A: 50 auditors; Exp. 2B: 132 auditors	Direct test of anchoring	Task used in Exp. 1A and 1B was markedly less representative of a typical audit task than that in Exp. 2A and 2B	Anchoring bias was strongly supported in Exp. 1A and 1B (less realistic task), but was not evident in Exp. 2A and 2B (more realistic task); in addition, results of Exp. 2B were consistent with conservatism
Kinney & Uecker (1982; analytical review experiment)	154 senior auditors	Direct test of anchoring	Task was simplistic but should have been familiar to subjects	Anchoring bias was supported; evidence also indicated behavior consistent with conservatism
Biggs & Wild (1985)	121 auditors (range of experience levels)	Direct test of anchoring	Kinney & Uecker (1982) task was modified (more data provided) to be more representative of typical audit task	Although anchoring bias was supported, extent of bias was mitigated by expanded, more realistic task; evidence regarding conservatism was largely inconclusive
Tomassini, Solomon, Romney, & Krogstad (1982)	58 auditors (average experience about 4 years)	Calibration of probability assessments	Task was job related, realistic, and of a type that should have been familiar to subjects	Although some miscalibration was apparent, results show relatively better calibration than studies using student subjects; auditors exhibited a tendency toward underconfidence, which is consistent with conservatism
Butler (1986)	107 auditors (average experience 5 years) and 58 undergraduate students	Anchoring in probability assessments	Task was somewhat simplified, but should have been familiar to auditors and not student subjects	Differences existed between judgments of auditors and students; auditor judgments were consistent with using anchor provided in the task in only one of three treatment conditions; auditor responses were apparently affected by experiential factors
Kinney & Uecker (1982; compliance sampling experiment)	169 senior auditors	Anchoring in probability assessments	Task was simplistic, but job related and moderately familiar to subjects	Support for anchoring was found for auditors using fractile assessment method, but findings were not conclusive for those using direct risk-assessment method
Joyce & Biddle (1981a)	Exp. 3A: 50 auditors; Exp. 3B: 120 auditors	Anchoring and coherence	Task was simplistic; although it represented an important auditing decision, it is typically made by auditors with more experience than those used in the study	Experiments examined coherence of probabilistic judgments involving conjunctive (Exp. 3A) and disjunctive (Exp. 3B) elementary events; although results were not always consistent with normative principles, there is some evidence of coherence in the auditors' judgments
Frederick & Libby (1986)	64 auditors and 119 accounting students	Features of similarity model and coherence	Tasks were highly simplified problems on internal-control assessment	Most subjects overestimated the probability of conjunctions relative to their constituent events; experienced auditors' knowledge of specific internal-control prototypes made them more prone to conjunction errors than students on certain problems
Swieringa, Gibbins, Larsson, & Sweeney (1976)	Exp. 1A and 1B: 120 master of business administration students (60 for each experiment)	Base-rate neglect	For both experiments, tasks were abstract and not familiar to subjects	Exp. 1A used Kahneman & Tversky's (1973) lawyer/engineer problem; Exp. 1B was designed similarly, but used brief descriptions of "effective vs. ineffective" operating controls for business activities in lieu of personality sketches; results of both experiments provided strong support for the neglect of base rates
Johnson (1983)	66 accounting students	Direct test of representatives	Task was less abstract than Kahneman & Tversky's (1973) lawyer/engineer problem, but not familiar to subjects	Results strongly supported presence of representativeness; base rates did not significantly affect subjects' probability judgments, and there was a high correspondence between probability judgments and similarity assessments

Table 1 (*continued*)

Study	Subjects	Bias or heuristic	Task description	Summary
Joyce & Biddle (1981b)	Exp. 1: 132 auditors	Base-rate neglect	Topic is of interest to auditors, but task was not familiar to subjects	Extent of attention to base rates was normatively insufficient, but was markedly greater than has typically been found for naive student subjects in other heuristic research
Kida (1984a)	73 audit partners and managers	Base-rate neglect	Task was designed to represent type of judgment within professional responsibilities of this subject group and should be familiar to them	Effect of base-rate specificity and causality on attention to base rates was examined; results across experimental conditions indicate that most auditors attended to both base rates and indicant data; also, attention to base rates was enhanced when given a causal interpretation
Joyce & Biddle (1981b)	Exp. 2A–C: 182 auditors	Base-rate neglect	Context was job related, but the task, adapted from Kahneman's (1973) lawyer/engineer problem, was highly abstract and not familiar to subjects	Results generally indicated that attention to base rates was normatively inadequate; in experiment most closely conforming to Kahneman & Tversky's (1973) lawyer/engineer study, findings suggest somewhat increased attention to base rates by auditors, but, in other two experiments, level of attention was no greater than for student subjects in other heuristic research
Joyce & Biddle (1981b)	Exp. 3A and 3B: 182 auditors	Sensitivity to source reliability	Task represents a typical audit judgment for subjects	Differences in source reliability significantly affected auditors' judgments in Exp. 3A but not Exp. 3B
Bamber (1983)	35 audit managers	Sensitivity to source reliability and data predictability	Task was highly familiar to subjects	Results strongly supported auditor sensitivity to source reliability and data predictability and suggested somewhat excessive discounting for less than perfect source reliability; conservatism may have contributed to the apparent enhanced sensitivity of auditors
J. Cohen & Kida (1989)	96 audit managers and seniors	Sensitivity to source reliability	Task was highly familiar to subjects	Examined sensitivity to source reliability in a nonprobabilistic judgment context; results strongly indicated that auditors were sensitive to differences in reliability of data
Swieringa, Gibbins, Larsson, & Sweeney (1976)	Exp. 3: 120 master of business administration students	Sensitivity to sample size	Tasks were abstract; used Kahneman & Tversky's (1972) large- versus small-hospital problem and a similar task in business scenario	Majority of subjects were not sufficiently sensitive to sample size in a normative sense, but they performed markedly better than Kahneman & Tversky's (1972) student subjects; more than 42% of the subjects responded correctly compared with 20% of Kahneman & Tversky's undergraduates
Gibbins (1977)	71 auditors	Sensitivity to sample size	Highly simplified exercises using job-related scenario (adapted from Kahneman & Tversky's, 1972, hospital problem)	The results suggest greater sensitivity to the role of sample size in probabilistic inference on the part of auditors compared with those typically reported for student subjects
Biddle & Joyce (1979)	182 auditors	Sensitivity to sample size	Task context was representative of audit decision settings, but was highly simplified	Results indicated a greater appreciation for the role of sample size by auditors than has typically been found for student subjects in other heuristic research; auditors were found to generally understand inverse relationship of sample size and sampling error

(*table continues*)

Table 1 (*continued*)

Study	Subjects	Bias or heuristic	Task description	Summary
Uecker & Kinney (1977)	112 certified public accountants	Sensitivity to sample size	Task was greatly simplified, but should have been familiar to subjects	Overall approximately 69% of responses were normatively correct, indicating a greater appreciation by auditors for the role of sample size in probabilistic inference than by student subjects in related heuristic studies
Kida (1984b)	66 audit partners and managers	Confirmatory bias in information search	Task was designed to represent judgment for which auditors at this level of experience are typically responsible	Little support for the presence of confirmatory strategies was evident; more predominant evidence, consistent with conservatism, indicated pervasive attention to more failure than viability items, implying a skepticism toward positive hypotheses (or outcomes) and acceptance of negative hypotheses
Trotman & Sng (1989)	80 auditors (average experience 5.8 years)	Confirmatory bias in information search for sequential judgment processes and in the presence of cue-diagnosticity information	Task was adapted from Kida (1984b) and should have been familiar to subjects	Results provided only weak support for confirmatory bias; consistent with conservatism, principal bias observed was that auditors chose more failure than viable items; bias seemed to be mitigated somewhat by presence of cue-diagnosticity information
Anderson & Kida (1989)	135 audit seniors	Effect of source credibility (i.e., bias and expertise) of inherited hypotheses on information-search strategies of auditors	Task was designed to be highly representative of a judgment for auditors at this level of experience	Results did not support presence of confirmatory strategies; neither source expertise nor bias had a significant effect on information search, but auditors pervasively listed more negative than positive items, again suggesting presence of conservatism
Butt & Campbell (1989)	123 auditors (range of experience: 3–9 years)	Effects of hypothesis-testing strategies on audit judgments	Task was somewhat simplified, but should have been familiar to auditors; however, instructing auditors to look for evidence that either confirmed or disconfirmed prior beliefs was artificial	Results indicated that auditors did not generally use confirmatory strategies unless explicitly instructed to do so
Anderson (1988)	60 audit seniors	Effect of source credibility (i.e., bias and expertise) of inherited hypotheses on hypotheses-testing strategies using information stored in memory	Task was designed to be highly representative of a judgment for auditors at this level of experience	Results were not consistent with confirmatory strategies; auditors did not recall significantly more confirming items, nor did they consider confirming information to be more relevant when testing hypotheses; however, there was greater attendance to negative information irrespective of hypothesis direction or time of hypothesis introduction, again suggesting the use of conservatism

directionally appropriate, typically are insufficient (see Slovic & Lichtenstein, 1971). Therefore, different starting points result in different estimates, which are biased toward the starting points. In addition to directly examining the basic anchoring phenomenon, researchers have used anchoring to explain decision behaviors observed in related areas of research, such as investigations of judgment *coherence* (i.e., the extent to which individuals' subjective probability estimates conform to normative probability axioms) and *calibration* (i.e., the extent of agreement between individuals' subjective probability distributions and the true underlying distributions).

In the typical audit setting, there are numerous instances in which auditors make probabilistic inferences. Considering the nature of these judgments and the information on which they are based, they seem readily susceptible to the anchoring phenomenon. A number of studies have directly tested for the anchoring bias, whereas others have examined its effect on various probabilistic audit judgments.

Direct Tests of Anchoring

Joyce and Biddle (1981a) conducted a series of experiments to examine the probability assessments of experienced auditors (see Joyce & Biddle, 1981a, Experiments 1A, 1B, 2A, and 2B in Table 1). Their Experiment 1A used a highly simplistic task, not the type of decision context that auditors typically encounter. Auditors were first asked to judge whether the incidence of "significant executive-level management fraud" was more or less than 10 in 1,000 clients (or 200 in 1,000 in a second treatment condition). They then estimated the number of clients per 1,000 that have "significant executive-level management fraud." The results strongly supported the presence of an anchoring bias. That is, the initial value suggested to the auditors, although normatively irrelevant, significantly affected their judgments. These findings were replicated with a different group of auditors in Experiment 1B using a wider range of anchors and a specified dollar threshold to define "significant" management fraud.

However, when the experimental tasks were more analogous to typical audit judgments, and hence more familiar to the subjects, anchoring results were often not found or were significantly mitigated or modified. For example, in Joyce and Biddle's (1981a) Experiments 2A and 2B (see Table 1) auditors completed an exercise analogous to a common audit judgment. They were given a list of accounting controls for one part of a client's operation, and were asked to rate the extensiveness of the audit tests they would perform to prevent or detect certain types of errors (the hypothesized anchor). Two of the controls were then changed, and the auditors provided a second extensiveness rating. In Experiment 2A, the changes weakened accounting controls and hence warranted more extensive testing (i.e., a higher rating). These ratings were compared with a control group that saw only the "weaker controls" case. In Experiment 2B, one group started with stronger accounting controls and then evaluated weaker controls, and the order was reversed for another group.

Results in both experiments indicated behavior that was not consistent with anchoring. The adjustments from initial anchors were typically directionally appropriate and extensive (there was only slight evidence of anchoring in Experiment 2B when controls changed from weak to strong; all the other cells provided no support for anchoring and adjustment). In fact, auditors indicated that the most extensive testing should be performed on the weaker system that followed an initial stronger system anchor. This result is actually contrary to anchoring and adjustment.

It appears that task familiarity contributed to the differences in the studies. The experimental tasks in the two pairs of experiments were distinctly different in the extent to which they represented familiar, recurring judgments for the subjects. Joyce and Biddle (1981a) described the tasks for Experiments 1A and 1B as being "easily the least familiar of those presented to the subjects" (p. 141). Although the incidence of management fraud was undoubtedly of interest to the auditors, the immediate judgment was surely not within the scope of their responsibilities and seems best characterized simply as a question of "general knowledge." Conversely, the task in the second pair of experiments (2A and 2B) is described as a "more common audit judgment" (p. 129) and seems to better represent a typical judgment for this subject group.

Two results of the study point to task familiarity as a plausible explanatory variable for differences between the two pairs of experiments. Specifically, the difference in mean ratings across treatments was significantly larger for the "stronger to weaker" treatment than for the "weaker to stronger." Also the extensiveness rating for the weaker system was actually higher when the weaker system followed a stronger system than when it was given first, a result contrary to anchoring. Thus, the context of the initial anchor and subsequent information can result in an overadjustment instead of an insufficient adjustment when certain audit judgments are examined. In this case, a change from a stronger to weaker system seems to signal a "red flag" to auditors, suggesting potential problems that require considerable additional audit testing. Given the auditors' knowledge of the relationships between internal controls and other important audit concerns, the sequence of occurrence (i.e., the direction of change in the strength of the internal-control system) may have provided additional information to the auditor beyond the present level of control strength. For example, the realization that management has allowed once strong controls to become weak may raise concerns for auditors beyond those associated with the presence of a weak control system.

This behavior appears to be a by-product of the utilities associated with the job. That is, auditors typically associate greater risks with overreliance on a weak control system than underreliance on strong controls. Hence, they are less willing to reduce testing when the control system seems to improve than they are to increase testing when the controls seem to weaken. In effect, because the task was more realistic, additional factors relating to the audit environment influenced auditor behavior. This type of behavior (i.e., a tendency to give more attention to, and to be more influenced by, negative information or outcomes) has been observed in numerous studies involving audit professionals, across a range of tasks and auditor-experience levels, and has been labeled *conservatism*.[1] The factors that contribute to the presence of this phenomenon have not specifically been identified and examined, but it probably stems from the potential serious consequences of audit judgments. For example, auditors may be held legally liable by third parties (e.g., lenders or investors) for misstatements in financial reports, especially when such reports overstate the profitability or economic viability of a company. Such consequences could result in increased reliance on negative data.

In another direct test of anchoring, Kinney and Uecker (1982) asked senior auditors to complete an analytical review exercise in which the reasonableness of an unverified amount for a specific item on the current-year financial report was evaluated by comparing it with prior years' verified amounts. In so doing, auditors were asked to specify a "noninvestigation inter-

[1] *Conservatism*, as used in audit-judgment research, must be distinguished from the heuristic of the same name used in the psychological literature to describe the finding that individuals revise probability estimates much less than required by the optimal Bayesian rule (see Edwards, 1968; Slovic & Lichtenstein, 1971). Conservatism in audit judgments is discussed more extensively in a subsequent section of the article.

val" for the current-year amount. Because reasonable bounds for a particular item should be determined without reference to the unverified amount, use of that amount (anchor) is normatively inappropriate. The task was of a type that should have been familiar to the subjects but was highly simplified. The results were consistent with anchoring (i.e., the noninvestigation regions set by groups given high vs. low unverified amounts [anchors] were significantly different and biased in the direction of the anchor). However, the study also found that 81% of the noninvestigation intervals of auditors in the high-anchor group indicated more investigation was needed compared with 61% for the low-anchor group. Thus, auditors were less willing to accept (without investigation) amounts suggesting economic improvement (i.e., the unverified amount in the high-anchor condition) than amounts suggesting economic decline. These findings are, once again, consistent with conservatism, and suggest that the auditors' behavior was affected both by the anchor and by conservative tendencies.[2]

Biggs and Wild (1985) replicated Kinney and Uecker (1982), and to make the experimental task more representative of the actual audit task, they added two treatment conditions (one for each anchor) in which data for the prior 5 years (instead of 2 years) was provided. The results corroborated the major finding of Kinney and Uecker (1982) in that the judgments were biased in the direction of the unaudited values. However, the anchoring effects were mitigated by providing a larger and more realistic data set. Findings regarding the relative influence of the two anchors, indicating conservatism, were largely inconclusive.

Anchoring Effects in Related Research

Audit professionals routinely use statistical sampling procedures to assess underlying populations. Methods such as fractile assessment, which requires estimates for specific percentiles of a population value (i.e., a confidence interval for that value), and direct risk assessment, which requires estimates of the risk (probability) that the population value exceeds some specified value, are described in authoritative pronouncements regarding standards of audit conduct. Findings in psychological studies indicate that distributions elicited using the fractile assessment method are often too narrow, and that fractile and direct risk assessments, although formally equivalent, often yield different distributions, findings that have been attributed, in part at least, to anchoring (e.g., Alpert & Raiffa, 1982; Slovic & Lichtenstein, 1971; Tversky & Kahneman, 1974). The evidence on auditor calibration indicates less overconfidence than typically found in psychological studies, and suggests the influence of factors other than anchoring. In addition, other studies examining distributions assessed by auditors suggest that although anchoring is present, it is not necessarily consistent with the behaviors predicted by Tversky and Kahneman (1974), and that factors other than anchoring again seem to play an important role.

Tomassini, Solomon, Romney, and Krogstad (1982) examined experienced auditors' calibration by assessing their probability distributions in testing client-account balances using fractile assessments. The task was realistic and analogous to judgments within the normal professional responsibilities of the subject group. The results differ from the typical findings reported in the calibration literature. Although some miscalibration was observed, the results indicate relatively better calibration than studies using student subjects. Furthermore, consistent with conservatism, the findings indicated a tendency toward underconfidence among auditors. As Tomassini and co-workers noted (p. 400), task realism and expertise may have played an important role in the relatively improved calibration of auditors, and the tendency toward underconfidence may have been due to auditor risk aversion in response to characteristics of the typical audit environment, process, and training.

Butler (1986) investigated the effects of three different anchor amounts in a scenario in which experienced auditors used direct risk assessments to evaluate the "reasonableness" of a total dollar amount said to be due from customers. Auditors were given an allowable risk for incorrectly accepting the balance as reasonable, and were asked to determine the true risk based on sample data. Although somewhat simplistic, the task should have been familiar to the subjects. The results provided only limited support that the anchor specified in the case served as the anchor for the auditors' judgments, and indicated that some other factor strongly influenced those judgments. If a strict anchoring and adjustment process was used, the mean response for each treatment group should have fallen between the anchor and the normative (or control-group) response. However, this occurred in only one of the three treatment conditions. Only the low-anchor response was consistent with a strict anchoring process. In the high-anchor condition, the auditors' mean response indicated an overadjustment, and in the middle-anchor condition the auditors' response was in the "wrong" direction. Although the evidence is inconclusive, it is feasible that auditors are accustomed to risk levels that were significantly lower than the middle and high anchors presented in the case. Butler (p. 107) suggested that the auditors may have brought an "internal, experiential" anchor to the task. The experiment yielded different results when replicated using undergraduate students. The mean responses for all three treatment groups were between the anchor and the normative response.

In a similar experiment, Kinney and Uecker (1982) asked experienced auditors to use sample data to make either a fractile or a direct risk assessment of a population error rate. The findings for the fractile assessment group were consistent with anchoring (i.e., the assessments appeared to be too narrow), but were not clear for the direct risk-assessment group.

Certain audit judgments require combining conjunctive or disjunctive events (at least implicitly) to arrive at a probabilistic inference. Coherence research in psychology investigating whether such judgments conform to probability axioms may therefore have important implications for professional auditors. Results from a number of psychological studies suggest that individuals' probability judgments are often not coherent. The findings indicate, generally, that individuals tend to overes-

[2] Biggs and Wild (1985) noted, however, that the Kinney and Uecker (1982) results must be interpreted cautiously, because the problem materials indicated a 2-year downward trend that reversed in the high-anchor condition but continued in the low-anchor condition. This may have contributed to the different frequencies of investigation indicated for the two groups.

timate the probability of conjunctive events or underestimate the probability of disjunctive events (e.g., Bar-Hillel, 1973; Beyth-Marom, 1981; J. Cohen & Chesnick, 1970; J. Cohen, Chesnick, & Haran, 1972; Morier & Borgida, 1984; Tversky & Kahneman, 1983). The results are often explained either within the framework of anchoring, in which subjects are said to anchor on the probability of such events at some stage in the decision process (Tversky & Kahneman, 1974), or features of similarity, in which subjects are said to judge the probability of a conjunctive event based on its similarity to a prototype (Tversky, 1977; Tversky & Kahneman, 1983). The findings on audit-judgment coherence indicate that experienced auditors are susceptible to overestimating conjunctive events, but also provide some evidence of coherence in their judgments.

Joyce and Biddle (1981a, Experiments 3A and 3B) asked auditors (average of 4 years experience) to make a "going-concern" judgment (i.e., assess the probability that an economically troubled client could remain solvent). Although this judgment is clearly job related for auditors, it occurs relatively infrequently and at a high organizational level. Thus, it is questionable whether this subject group would have been familiar with the experimental task. In addition, the exercise seems highly simplistic in terms of the number of factors that may be considered in an actual judgment of this type. Experiment 3A examined conjunctive events where client success was dependent on the company's completing all of a set of events, whereas Experiment 3B examined disjunctive-event conditions in which the occurrence of any one of the events would prevent success. The results regarding the presence of an anchoring bias were inconclusive. Although the mean adjustment from the initial anchor was insufficient assuming independence of the elementary events, the authors noted that such independence is unlikely in the context examined. Thus, we cannot draw conclusions regarding the sufficiency of adjustments from the anchor.

However, there is limited evidence suggesting coherent judgment processes. For example, mean probability estimates for conjunctive (disjunctive) events were greater (less) in a three-event condition than in a five-event condition, there was no significant main effect for problem formulation (conjunctive vs. disjunctive), which is normatively appropriate because the two formulations were simply different representations of the same problem, and the effect of the number of elementary events on the auditors' probability estimates was not contingent on the problem formulation. Considering our characterization of the experimental task, the apparent coherence may indicate an increased familiarity with fundamental axioms of probabilities on the part of auditors relative to more naive student subjects.

Frederick and Libby (1986) examined auditors' probability assessments of conjunctive events within the framework of Tversky's (1977) feature-matching model. They asked experienced auditors and students to complete a set of audit problems concerning internal controls. Subjects were given a brief description of internal controls indicative of a system prone to certain types of financial statement errors and a list of seven items (errors) that included certain constituent events and conjunctions of interest. They were asked to rank order the items by the probability that they would be among the consequences of the described internal-control weaknesses. The results were consistent with the predictions of Tversky's (1977) feature-matching model and with prior findings in psychology in that most subjects overestimated the probability of conjunctions relative to their constituent events. Frederick and Libby hypothesized that, because of the expert auditors' knowledge of the internal-control prototype used, their judgments would differ from those of novices and would, in fact, be more prone to conjunction errors in certain circumstances. The results generally supported their expectations.

Summary

In general, anchoring effects are evident in the judgment processes of experienced auditors. However, the evidence points to instances where it is mitigated or modified, and these appear to be associated with the level of expertise of the subjects used and the extent to which the experimental task is job related and familiar to the subjects (see Table 1 for a summary of the results). For example, it appears that the use of an experimental task that was job related and familiar to auditors explains the absence of an anchoring effect in Joyce and Biddle's (1981a, Experiments 2A and 2B) research and the improved calibration of auditor probability assessments in Tomassini and colleagues' (1982) study. Although knowledge of fundamental probability axioms may have contributed to the coherence evident in auditor assessments of probabilities of conjunctive and disjunctive events in Joyce and Biddle (1981a, Experiments 3A and 3B), knowledge of internal-control prototypes appears to make expert auditors more prone to conjunction errors than novices in certain circumstances (Frederick & Libby, 1986). In addition, experiential factors seem to have contributed to the different anchoring effects between student judgments and auditor judgments on the same common audit task (Butler, 1986). In addition, the findings in the studies of Joyce and Biddle (1981a, Experiment 2B), Kinney and Uecker (1982), and Tomassini and associates (1982) are consistent with the presence of conservatism, a phenomenon that has been observed frequently in studies of audit judgments.

It appears that these results are due to the subjects' response to characteristics of the audit process and environment evoked by experimental tasks that are analogues of typical audit judgments. Although anchoring and adjustment are found in some audit decisions, characteristics inherent in other audit-decision contexts result in behavior that is contrary to anchoring or indicate mitigated anchoring effects. Such findings can often be explained by conservatism or a combination of anchoring and conservative tendencies. The specific factors (task, environment, and individual) that contribute to differing experimental results across audit settings, or more generally across generic versus applied settings, remain largely unidentified. This issue is more fully discussed in the Conclusion.

Representativeness

When the representativeness heuristic is used, the individual is said to assess the probability that an item belongs to a population based on the extent to which the item is similar in essential properties to that population. Events that are more representative tend to be judged to have a higher probability of occurrence than less representative events. This heuristic has been used to

explain several biases (Tversky & Kahneman, 1974), including (a) neglect of base rates, (b) disregard for the impact of sample size, and (c) insensitivity to the predictive ability and source reliability of data. Although Bayesian revision specifies that proper inference from fallible evidence requires attending to each of these factors, because they have no effect on the degree to which an observed sample is perceived as representative of an underlying population, they will not affect probability judgments made using representativeness (Kahneman & Tversky, 1973).

Early experiments on base rates indicated that if individuating data were available, base rates would typically be ignored (Hammerton, 1973; Kahneman & Tversky, 1972, 1973; Lyon & Slovic, 1976). Replications of these experiments, varying base rates, problem content, information order, and response mode, have provided general support for the early findings, but also showed instances in which base rates are considered. The evidence suggests that base-rate use may be increased by enhancing the relevance or salience of the base rate (Ajzen, 1977; Bar-Hillel, 1980; Carroll & Siegler, 1977; Christensen-Szalanski & Beach, 1982; Fischoff, Slovic, & Lichtenstein, 1979; Manis, Dovalina, Avis, & Cardoze, 1980; Tversky & Kahneman, 1982) or by decreasing the relevance or salience of the individuating data (Bar-Hillel, 1980; Ginosar & Trope, 1980; Hinsz, Tindale, Nagao, Davis, & Robertson, 1988). Factors such as the causal link between base rates and target events, the level of specificity of the base rates, and the extent to which individuals have experience with the relationship between base rates and the individuating data have been identified as apparent underlying contributors to the increased use of base rates.

Studies examining insensitivity to sample size have typically found that student subjects, in laboratory settings, ignore sample-size information when evaluating sample results (e.g., Bar-Hillel, 1979; Fischoff et al., 1979; Kahneman & Tversky, 1972). Finally, insensitivity to predictive ability and source reliability of individuating data has been examined in the heuristics literature (Kahneman & Tversky, 1973). Although some studies in social psychology provide evidence that, in certain contexts (e.g., person perception), source credibility affects the way in which information is processed (Birnbaum, Wong, & Wong, 1976; McGinnies & Ward, 1980; Rosenbaum & Levin, 1968, 1969), results from some heuristics research indicate that subjects are often insensitive to differences in source reliability (Gettys, Kelly, & Peterson, 1973; Kahneman & Tversky, 1973; Youssef & Peterson, 1973). That is, relative to the normative model, subjects often do not sufficiently adjust their use of evidence for variations in source reliability.

There are numerous instances in the audit setting where auditors use fallible evidence, from differentially reliable sources, to assess the probabilities of uncertain events or revise probability judgments. In addition, auditors regularly use sample evidence to assess population characteristics. In that context, they are called on to determine appropriate sample sizes for sampling procedures or to analyze and interpret sample results. As such, the findings on representativeness in psychology may have important implications for auditors. Auditing studies have investigated the heuristic and examined the related biases.

Attention to Base Rates

Studies examining representativeness and the related base-rate phenomenon have been reported in the auditing literature using both student and auditor subjects working a variety of tasks (e.g., Johnson, 1983; Joyce & Biddle, 1981b; Kida, 1984a; Swieringa, Gibbins, Larsson, & Sweeney, 1976). Although on balance the results are mixed, an analysis of individual studies yields interesting findings. In experiments asking students to perform abstract and unfamiliar tasks, the results more strongly support the early findings in the heuristics literature, whereas in experiments asking experienced auditors to complete familiar, job-related tasks, the results are much less clear.

Swieringa and colleagues (1976) conducted two experiments using student subjects in highly abstract tasks. The first experiment (Experiment 1A on Table 1) partially replicated Kahneman and Tversky's (1973) classic lawyer/engineer study, in which subjects were presented with brief personality sketches, allegedly sampled at random from a group consisting of 70 engineers and 30 lawyers (or 30 engineers and 70 lawyers), and were asked to assess the probability that each description belonged to an engineer rather than a lawyer. The results conform substantially to the Kahneman and Tversky findings. Median responses indicated that the students focused on base rates only when no individuating information was provided or when they perceived the individuating data as "useless," but they ignored base rates in the presence of "useful" individuating data. The experiment was repeated (Experiment 1B on Table 1) using brief descriptions of companies' internal-control systems in lieu of personality sketches, and the results were very similar.

In a direct test of representativeness, Johnson (1983) asked accounting undergraduates to make judgments as to the probability that each of 12 sample firms were bankrupt. The task seems considerably less abstract than those used in early experiments in psychology, but it was certainly not familiar to the undergraduate subjects. They were given the incidence of bankruptcy for the industry to which the sample firms belonged (i.e., a base rate) and a five-item financial profile (i.e., individuating data) for each of 12 firms. Consistent with prior studies of this phenomenon in psychology, base rates did not have a significant effect on the subjects' probability judgments, and there was a high correspondence between the subjects' probability judgments and their similarity assessments.

A more central issue is whether experienced professionals, performing familiar tasks, will fall prey to this bias. The results indicate generally that the judgment behaviors of experienced auditors are not normative, but some suggest that auditors' behaviors correspond more closely to normative principles than the behaviors of naive student subjects. This increased attention to base rates by auditors seems to be present primarily for tasks that appear to be familiar to them.

Joyce and Biddle (1981b) conducted a series of experiments in which they asked experienced auditors to make probability judgments regarding the involvement of client managers in fraudulent activities (see Joyce & Biddle, 1981b, Experiments 1, 2A, 2B, and 2C). Although the topic is of compelling interest, the experimental tasks were not familiar to the auditors.

Experiment 1 used an adapted version of the medical-diag-

nosis problem of Hammerton (1973). Auditors were told that a team of accountants and psychologists had developed a procedure to test for management fraud by comparing a personality profile for key managers to a master profile compiled by clinical psychologists. Based on the similarity of the two profiles, the test was said to signal "fraud" or "no fraud." Auditors were provided with the positive hit rate and the false-positive rate for the test, purportedly based on extensive field testing, along with a base rate for manager involvement in material fraud. They were then asked to assess the probability that a key manager who received a "fraud" test signal was actually involved in fraudulent activity.

The results indicated an insufficient response to base rates by the auditors, but also suggested that they did not ignore base rates. That is, median and modal responses (a) were not consistent with ignoring base rates, and (b) differed significantly across base-rate conditions in a manner consistent with normative principles. As an example of differences in base-rate attention, we compared the responses of Joyce and Biddle's (1981b, Experiment 1) auditors with student subjects from selected heuristics experiments investigating related issues (Bar-Hillel, 1980).

Bar-Hillel (1980) is of particular interest because it allows a comparison of Joyce and Biddle's (1981b) results to findings that typify "ignore base-rate" results (i.e., Problems 1 and 2) as well as results indicating "increased" attention to base rates (i.e., Problems 7 and 8). In Problems 1 (cab problem) and 2 (suicide problem), which prior research indicated would result in the general base rate being perceived as less relevant than the indicant data, a majority of the subjects attended only to the indicant data. Conversely, for Problems 7 (intercom problem) and 8 (motor problem), which were designed to achieve equal relevance of base-rate and indicant data either by reducing the relevance of the indicant data (Problem 7) or enhancing the relevance of the base rate (Problem 8), the subjects attended both to base-rate and indicant data.

A comparison of the results of Joyce and Biddle's (1981b) Experiment 1 and Bar-Hillel's (1980) Problems 1 and 2 points to greater attention to base rates by professional auditors. For example, the percentage of subjects in Bar-Hillel's study whose responses were consistent with attending only to the indicant information was approximately 40% compared with only about 20% of Joyce and Biddle's subjects. Similarly, the percentage of Bar-Hillel subjects whose responses approximated the normative response was around 13% compared with nearly 32% of the Joyce and Biddle auditors.[3] Both the median and modal responses of Bar-Hillel's subjects, but not Joyce and Biddle's subjects, were consistent with using only the indicant information, and only 47% of Bar-Hillel's students regressed down from the indicant data compared with 72% of the auditors. Bar-Hillel (1980; see pp. 220 and 222) also observed that the same pattern of results was obtained across numerous variations of Problem 1; the modal response was virtually always consistent with totally ignoring base rates. In addition, Joyce and Biddle's results correlate much more closely to Bar-Hillel's Problems 7 and 8, where the relevance of indicant data was reduced or base-rate data enhanced. For example, considering the responses for Problems 7 and 8 in aggregate, about 66% were consistent with using both base-rate and indicant data, 32% approximated the normative response, and only 12% indicated sole use of the indicant data. In effect, auditor responses were comparable to these and were achieved without experimentally enhancing the relevance of the base rate or reducing that of the indicant data.

Additional data regarding attention to base rates are found in Kida (1984a). Audit partners were given (a) a base-rate probability of failure for firms in general, (b) a prediction (fail vs. not fail) regarding a specific firm, purportedly based on a mathematical model developed by finance researchers, and (c) the positive hit rate and false-positive rate for the model. They were then asked to predict the likelihood of failure for that firm. This type of judgment falls within the professional responsibilities of this subject group and should be familiar to them. In fact, the author points out that mathematical prediction models are used by auditors in actual failure decisions. Kida manipulated causality and specificity such that, in certain conditions, the base-rate data fit a causal schema of company failure or was specific to the company under consideration (both of which may increase attention to base rates). The distribution of responses across all experimental conditions indicates that most auditors attended to base rates to some degree, that is, 79% of the responses regressed toward the base rate, whereas 21% used only the indicant data. In addition, Kida found that attention to base rates was enhanced when given a causal interpretation.

Finally, Joyce and Biddle (1981b, Experiments 2A–2C) asked auditors to complete a probabilistic inference task adapted from Kahneman and Tversky's (1973) lawyer/engineer problem. The task appears to be highly abstract and not of a type familiar to the subjects. Auditors were given 10 brief management descriptions, some of which described managers found to be involved in fraudulent activities. The experimenters varied base rates and individuating data, and the auditors were asked to assess the probability that the description selected was one of those engaged in fraudulent activity. The results generally indicated that attention to base rates was normatively inadequate. Furthermore, evidence regarding the relative extent of attention was inconclusive. In the experiment most closely conforming to Kahneman and Tversky's lawyer/engineer study, the findings suggest somewhat "increased" attention to base rates by auditors. However, in the other two experiments, the level of attention was no greater than for student subjects.[4]

---

[3] *Approximated* is defined as a range of approximately 10 percentage points to either side of the normative response. A very similar measure was used by Bar-Hillel (1980; e.g., pp. 220 and 222).

[4] Holt (1987) replicated certain parts of Joyce and Biddle's (1981b) Experiments 1 and 2, and also reported no difference in sensitivity to base rates of auditors and students. However, a close examination of Holt's experimental results indicates important differences in the response distributions of students and auditors. For example, responses indicating exclusive attention to either the indicant probability or the base-rate probability in the "management fraud" case (see Holt's Table 1) were found much more often for students (61% of the responses) than auditors (32%), whereas responses consistent with combining the two variables were found more often for auditors than students (i.e., 66% vs. 37%).

Attention to Source Reliability

The audit setting provides numerous instances in which auditors make judgments using information from differentially reliable sources. Joyce and Biddle (1981b) conducted two experiments in which auditors (mean of 4 years experience) were asked to assess the collectibility of a large, past-due amount owed to their client by a customer (see Table 1; Joyce & Biddle, 1981b, Experiments 3A and 3B). The experimental task seems to represent a typical, recurring audit judgment for the subjects. The two experiments differ primarily in the form of experimental manipulation used (i.e., between vs. within subjects). The source of the information regarding the delinquent customer was either less reliable (client's credit manager) or more reliable (independent credit agency) depending on treatment condition. Differences in source reliability significantly affected the auditors' judgments in the within-subjects manipulation (Experiment 3B) but not in the between-subjects case. Significant results in one, but not both, experiments may be due to the increased salience of the manipulated variable in the repeated measures design, but in this instance an inadequate manipulation of source reliability in the between-subjects case is equally plausible. The manipulation consisted simply of changing a single phrase in the middle of a rather extended paragraph, which could have easily gone unnoticed when a comparison between conditions could not be made.

For certain decisions in the audit setting, more generalized judgments are made at successively higher levels in the audit team hierarchy. Using this type of scenario, Bamber (1983) examined how subjects responded to different levels of source reliability and data predictability. He asked audit managers to make a series of internal-control judgments that should have been highly familiar to the subject group. The judgments were based on results of sampling procedures conducted by a subordinate, along with the subordinate's recommendation. Auditors responded to eight situations in a repeated measures design in which four distinct levels of source (subordinate) reliability and two levels of data predictability were considered. The results strongly indicated a sensitivity of the auditors' judgments to source reliability and to data predictability. In fact, a comparison of the auditors' posterior judgments to normative probability estimates suggested somewhat excessive discounting for less than perfect source reliability. Although the findings may be due, in part, to the use of a repeated measures design, in this judgment context that design well represents a common situation for audit managers.

Finally, J. Cohen and Kida (1989) examined source reliability in a nonprobabilistic judgment context. They asked audit seniors and managers to determine the amount of work to be conducted on a particular audit assignment where the strength of the client's internal-control system (strong vs. weak) was manipulated in a between-subjects design. In general, stronger (weaker) systems are more (less) reliable as an information source and should warrant relatively less (more) audit work. The results strongly indicated that auditors were sensitive to differences in the reliability of the data.

Insensitivity to Sample Size

When assessing the characteristics of an audit population, auditors frequently rely on sample evidence. They are regularly called on to determine appropriate sample sizes for sampling procedures or to analyze and interpret sample results. As such, insensitivity to sample size could have serious consequences for professional auditors. Limited evidence on this bias can be found in the audit-judgment literature using both student subjects (Swieringa et al., 1976) and experienced professionals (Biddle & Joyce, 1979; Gibbins, 1977; Uecker & Kinney, 1977). The results suggest that although the bias is present, it is not as pervasive among graduate business students or experienced auditors as it is among naive student subjects in other heuristics studies.

Swieringa and associates (1976) asked master of business administration students to complete a number of exercises, including Kahneman and Tversky's (1972) large- versus small-hospital problem and a similar task using a business scenario, to assess their perceptions of the effect of sample size on sample variance. A majority of the subjects were insensitive to sample size, that is, 49% (hospital problem) and 42% (business problem) of the subjects gave the normative response. However, their performance was markedly better than Kahneman and Tversky's undergraduate subjects, of whom only 20% gave the normative response. Furthermore, one third of the subjects were asked why they selected their response, and in virtually every case "correct" respondents explicitly identified the sample-size issue and "incorrect" respondents did not. These results point to an increased familiarity with (or knowledge of) fundamental sampling issues on the part of the graduate business students. This explanation is corroborated by the fact that Kahneman and Tversky's (1972) undergraduate subjects are described specifically as having no background in probability or statistics.

Gibbins (1977) and Biddle and Joyce (1979) asked experienced auditors to complete a number of different sampling exercises involving business or audit scenarios adapted from Kahneman and Tversky's (1972) large- versus small-hospital problem. In these exercises, typically more than 50% of the auditors gave the normative response specific to that exercise. Similarly, in an audit-sampling exercise, Uecker and Kinney (1977) specified a finite population, a sample size, the number of errors, and a sample error rate, and asked certified public accountants to choose (from each of five pairs of sample outcomes) the sample that provided the greatest assurance that the population error rate was 5% or less. Overall, 69% of the responses were correct. These results indicate that many auditors were insensitive to sample size, but also suggest a greater sensitivity to the role of sample size in probabilistic inference on the part of auditors relative to students in heuristics research.

Although not conclusive, the evidence points to the subjects' knowledge of fundamental statistical relationships as a potentially important explanatory variable for the varied experimental results described previously here. Swieringa and co-workers' (1976) graduate students outperformed their undergraduate counterparts on Kahneman and Tversky's (1973) large- and small-hospital problem. Increased emphasis of statistical relationships at the graduate level may explain the difference. Furthermore, although the auditors' performance was also notably better than Kahneman and Tversky's (1973) subjects, it does not appear to differ substantially from Swieringa and associates' graduate students. In addition, because sampling judgments

are rather common in audit settings, it seems reasonable to assume that auditors too are more knowledgeable of fundamental sampling issues than the undergraduate student subjects used in many heuristics studies. Finally, task characteristics do not appear to contribute to the differing results. In the studies examining auditors' judgments, the tasks were simplistic and abstract and do not seem to differ markedly from those used by Kahneman and Tversky.

Summary

In general, the findings indicate that the biases attendant to representativeness found readily in the judgments of student subjects in many heuristics studies are also present in the judgments of experienced professional auditors (see Table 1). However, the extent or severity of those biases is sometimes reduced. A comparison of the findings of Joyce and Biddle (1981b, Experiment 1) and Kida (1984a) with those of Bar-Hillel (1980, Problems 1, 2, 7, and 8) suggests that the extent of base-rate neglect is less for experienced auditors than for naive student subjects, except when the salience of base rates is experimentally enhanced (or individuating data reduced) for the student subjects. It appears that the auditors' familiarity with the task, and possibly a somewhat greater knowledge of the fundamentals of probabilities, are potentially important explanatory variables. In addition, the findings of Bamber (1983) and Joyce and Biddle (1981b) indicated a high level of attention to source reliability by experienced auditors. Finally, in the studies of Swieringa and colleagues (1976), Gibbins (1977), Biddle and Joyce (1979), and Uecker and Kinney (1977), experienced auditors demonstrated an increased sensitivity to sample size relative to findings in related heuristics research, and the evidence points to knowledge of fundamental statistical relationships as a principal explanatory variable. Specific task, environment, and individual factors that may contribute to such differing results across experimental settings are more fully discussed in the Conclusion.

Confirmatory Strategies

When completing judgment tasks that are diagnostic in nature, individuals often formulate and evaluate alternative hypotheses. Snyder and Swann (1978) identified at least three hypothesis-testing strategies that might be used. In one, the individual preferentially attends to evidence that confirms the hypothesis being tested (i.e., a confirmatory strategy). Alternatively, preferential attention to information that disconfirms a hypothesis is indicative of a disconfirming strategy, whereas the third strategy, an equal-opportunity strategy, is exhibited when an individual invests equal amounts of time or effort to uncover both confirming and disconfirming evidence. Numerous studies have reported that individuals pervasively engage in confirmatory strategies when searching for information (e.g., Skov & Sherman, 1986; Snyder, 1981; Snyder & Campbell, 1980; Snyder & Cantor, 1979; Snyder & Gangestad, 1980; Snyder & Swann, 1978; Snyder & White, 1981) and when retrieving information from memory (e.g., Berman, Read, & Kenny, 1983; C. Cohen, 1981; Rothbart, Evans, & Fulero, 1979; Snyder & Cantor, 1979; Snyder & Uranowitz, 1978; Zadny & Gerard, 1974). Similarly, individuals have also been found to regard confirming evidence as more informative (Schustack & Sternberg, 1981; Wason, 1960; Wason & Johnson-Laird, 1972).

Snyder and Gangestad (1981) noted that the use of confirmatory strategies is not affected by the origin of the hypothesis, whether the hypothesis is likely to be accurate or inaccurate or whether incentives for accurate hypothesis testing are offered. It should be noted, however, that although considerable support for the presence of confirmatory strategies exists, some evidence indicates that individuals also use search strategies in which attention to information is affected by cue diagnosticity (Skov & Sherman, 1986; Trope & Bassok, 1982), and that individuals sometimes use equal-opportunity strategies (Strohmer & Newman, 1983). Furthermore, Klayman and Ha (1987, 1989) viewed the selection of an information-search strategy as an adaptive process. They argued that confirmatory behaviors may represent a general positive test strategy that is used unless information identifying other strategies as preferable is present.

The findings in audit-judgment research, using professional auditors performing familiar, job-related tasks, provide very little support for the use of confirmatory strategies. Rather the evidence suggests the use of an information-search strategy not typically found in other contexts that is explained readily by conservatism. These findings seem consistent, to some extent, with Klayman and Ha's (1987) adaptive view of the information-search process.

Libby (1981) noted that auditors often formulate hypotheses in their judgment tasks and then search for data to test those hypotheses. Such tasks require auditors to sift through numerous information items, some of which may provide confirming evidence and some disconfirming evidence. Furthermore, audit evidence is acquired at various times and in various forms (e.g., written documents, discussion, and observation) during an audit engagement. In addition, because of the cost of continually referencing information in external storage, audit evidence may be stored in memory for later use in decision processes (Birnberg & Shields, 1984). Thus, auditors' judgments are susceptible to the same information search and recall biases that have been observed in psychology.

Information-Search Biases

Kida (1984b) examined whether auditors would attend to more confirmatory evidence, disconfirmatory evidence, or equal amounts of both when testing a hypothesis. He asked audit partners and managers to assess whether a particular firm would go bankrupt or continue to be economically viable. The subjects were selected because they are typically responsible for, or provide the greatest input into, such decisions. A firm description was provided that contained 20 information items that auditors consider important for these decisions. Ten of the items pointed to failure and 10 to continued viability. The auditors were assigned to either a failure- or a viability-hypothesis condition and were asked to assess the probability that a firm would fail within 2 years or remain viable for at least 2 more years depending on the treatment condition. In addition, the subjects listed the items that they considered relevant and ranked them in order of importance.

The results did not provide strong support for a confirmatory

bias. A strong bias would lead subjects in the failure group to list more failure items as relevant and those in the viability group to list more viable items. However, this was not the case. Although very limited evidence of hypothesis-framing effects were present (i.e., auditors in the viability group listed more viable items than auditors in the failure group), the principal finding was that auditors in both the failure and viability conditions attended to significantly more failure items. An analysis of the auditors' probability judgments corroborated these findings. The difference in the judgment means across the treatment groups was not significant, and correlation analysis indicated a significant correspondence only between probability judgments and the number of failure items.

These findings indicate that confirmatory strategies were not nearly as evident as they had been in prior psychological studies. Furthermore, the more predominant evidence—that auditors attended to more failure items than viable items across hypothesis-framing conditions—suggests the use of conservatism. That is, the pervasive attention to more failure than viability items by auditors implied a skepticism toward the positive hypothesis (or outcome) and acceptance of the negative hypothesis.

Trotman and Sng (1989) examined auditor information-search strategies in sequential judgment processes and in the presence of cue-diagnosticity information. The results, consistent with Kida (1984b), provided very little support for the confirmatory bias. When prior information indicated viability, hypothesis framing had a significant effect on the relative number of failure and viable cues selected, but when prior information indicated failure, hypothesis framing had no significant effect. Consistent with conservatism, the principal bias observed was that auditors chose more failure than viable items. This bias seemed to be mitigated somewhat by providing information on cue diagnosticity.

Anderson and Kida (1989) argued that differences in the loss functions evoked in prior heuristics studies versus in Kida's (1984b) research may serve to explain the differing results. That is, heuristics experiments often use tasks that have no immediate or long-term implications to the subjects. However, the implications for auditors of poor audit judgment are potentially very substantial (e.g., legal liability), resulting in the conservatism effect found by Kida. They suggested that, because of auditors' loss functions, the simple differential framing of a hypothesis may not have been sufficient to reveal evidence of confirmatory strategies among auditors. They proposed that the inheritance of hypotheses, which occurs frequently in the audit context, represents a potentially stronger manipulation because it may result in greater commitment to the hypothesis, especially when the credibility of the hypothesis source is high.

They asked senior auditors to evaluate the level of control risk associated with a client's internal-control system, a judgment made frequently by senior auditors. The auditors were given a description of an internal-control system, comprised of an equal number of items confirming a positive (low-risk) hypothesis and a negative (high-risk) hypothesis. The materials included a statement (hypothesis) from one of four differentially credible sources, concerning the level of control risk (high or low) associated with the internal-control system.

The results again provided little support for the presence of confirmatory strategies. Neither the level of source expertise nor the level of bias had a significant effect on the number of positive or negative items listed. However, auditors in both hypothesis conditions listed more negative than positive items, again suggesting the presence of conservatism.

Butt and Campbell (1989) examined the effects of hypothesis-testing strategies on audit judgments. Auditors started with either high or low initial beliefs regarding the quality of an internal-control system, were presented with mixed evidence (i.e., both positive and negative data) relating to system quality, and were then asked to assess the probability that the internal controls would prevent or detect material errors. Some auditors were either explicitly instructed to look for evidence that confirmed or disconfirmed their initial belief; others were not given explicit instructions. Although the task itself should have been familiar to the subjects, the use of instructions to look for confirming (disconfirming) evidence was highly artificial. The results indicated that auditors did not generally use confirmatory strategies unless explicitly instructed to do so.[5]

Biased Recall

Anderson (1988) investigated auditors' hypothesis-testing strategies when using information stored in memory. Auditors again inherited either a positive (low internal-control risk) or negative (high internal-control risk) hypothesis, and were assigned to one of two hypothesis timing conditions. They inherited the hypothesis either at the same time that they were given the control-system description or immediately before making the control judgment. Judgments were made after the descriptive information was made inaccessible. The results were again not consistent with the presence of confirmatory strategies. The auditors did not recall significantly more confirming (vs. disconfirming) items, and they did not consider confirming information to be more relevant than disconfirming when testing hypotheses from information stored in memory. However, there was again a greater attendance to negative information irrespective of hypothesis direction or time of hypothesis introduction, again suggesting the use of conservatism.

Summary

The findings in the accounting literature provide very little support for the presence of confirmatory strategies in the information search and recall processes of professional auditors in the performance of familiar, job-related tasks. Furthermore, the results strongly suggest the presence of a pervasive, overriding concern by auditors for negative outcomes (i.e., conservatism), which may have nullified or precluded the use of confirmatory strategies. Alternatively, conservatism may manifest itself by causing auditors to habitually frame hypotheses in

[5] Kaplan and Reckers (1989) also examined issues related to the information-search strategies of auditors. However, because of differences in methodologies used, the results are not readily comparable to the studies reported here. The findings suggest the use of more narrowly focused search strategies by inexperienced auditors relative to experienced auditors.

negative terms. Although the exact nature of conservatism effects remains at issue, specific task and subject characteristics (e.g., auditor loss functions in audit contexts) are no less important to explaining this phenomenon.

Recall that this phenomenon was also evident in various other studies (examining auditors performing familiar, job-related tasks) in which the findings failed to strongly support the presence of investigated biases (e.g., Anderson, 1988; Bamber, 1983; Joyce & Biddle, 1981a, Experiment 2B; Kida, 1984b; Kinney & Uecker, 1982; Tomassini et al., 1982). Furthermore, similar behaviors have been observed in audit studies not specifically examining issues of "heuristics and biases." For example, Ashton and Ashton (1988) examined auditors' sequential belief revisions when receiving information concerning the appropriateness of an account balance. The results indicate that auditor judgments were influenced more by negative data than by positive data. In one experiment, the mean belief revision made by auditors receiving negative data was more than twice the revision made by auditors receiving positive data (see also Ashton & Ashton, 1990). Evidence of conservatism has also been found in protocol research to descriptively model auditor-decision processes (Biggs, Mock, & Watkins, 1988) and in a direct test of one aspect of conservatism in a specific audit context (J. Cohen & Kida, 1989). Although the specific factors that contribute to the presence of this phenomenon have not been identified and examined, it appears that characteristics of the audit environment, such as the serious risks and consequences of audit judgments, influence auditor behaviors, resulting in conservatism. Accordingly, conservatism as manifested in auditors' apparent overriding concern for negative information, may be viewed as a functional heuristic that serves to minimize economic losses.

Interestingly, this explanation seems consistent with Klayman and Ha's (1987) arguments concerning confirmatory strategies. They proposed that people use a positive test strategy (i.e., tendency to examine cases that are known or expected to have the property of interest) as a general default heuristic. That is, this strategy is used in the absence of specific information identifying other strategies as preferable or when cognitive demands preclude the use of other strategies. A factor identified by Klayman and Ha as potentially important to such a preference is the type of judgment error of greatest concern to the decision maker. In an audit-judgment context, erroneously identifying a negative outcome (e.g., bankruptcy or high internal-control risk) as positive (e.g., economic viability or low internal-control risk) is generally perceived by auditors as being much more costly than vice versa. Hence, auditors give significantly greater attention to data consistent with negative outcomes than to data consistent with positive outcomes, even in response to positively framed hypotheses.

Importantly, the influence of conservatism on the behavior of auditors in audit settings suggests that certain behaviors may be unique (or at least particular) to the interaction of individuals and contexts. Furthermore, the empirical evidence suggests that the strength of such a heuristic may be sufficient to preclude, mitigate, or modify other phenomena, such as the biases observed readily in the behaviors of student subjects in many heuristics studies. This, of course, points to the importance of considering these factors when generalizing to professional judgment contexts.

Conclusion

There is a considerable body of heuristics research that shows that individuals readily fall prey to a number of judgmental biases. However, important issues concerning the generalizability of these findings have been raised (Berkeley & Humphreys, 1982; Ebbesen & Konecni, 1980; Edwards, 1983; Einhorn, 1976; Fischoff, 1982, 1987; Funder, 1987; Hogarth, 1981). One of the principal arguments is that this generalization is based on studies that, in large part, use tasks and subjects that are not representative of the contexts and individuals for which these findings seem most relevant. In essence, critics argue that judgment behaviors should be evaluated by examining the populations to which we wish to generalize, with individuals performing tasks that are readily familiar to them. In this article we have discussed evidence from a body of research concerning the judgments of professional auditors, in which many of the studies use audit professionals performing tasks that are analogues to those typically performed in their "workplace."

In general, the evidence reinforces the arguments made by Edwards (1983), Fischoff (1982, 1987), and others. Important differences in the findings of studies using experts in the performance of familiar, job-related tasks and those using student subjects or less realistic tasks are apparent. Although the evidence indicates that the heuristics and biases common to many experiments using student subjects and generic tasks are also present in the judgments of professional auditors performing familiar, job-related tasks, the nature of these heuristics or the extent of their presence is often notably different. For example, the extent of the biases attendant to the representativeness heuristic—base-rate neglect, inattention to source reliability, and insensitivity to sample size—are typically less for auditors than for student subjects (e.g., Bamber, 1983; Biddle & Joyce, 1979; Gibbins, 1977; Joyce & Biddle, 1981b, Experiment 1; Kida, 1984a; Swieringa et al., 1976; Uecker & Kinney, 1977). There is evidence of anchoring in some, but not all, auditor judgments, and the anchoring processes of auditors seem to differ (perhaps fundamentally) from those of students.

Furthermore, the results of audit-judgment research indicate instances in which biases found readily in other research are not evident in the judgments of professional auditors. For example, Joyce and Biddle (1981a) found evidence in auditor judgments that was contrary to the anchoring phenomenon. Similarly, although the findings have been quite strong in many psychological studies, there is very little support for the presence of confirmatory strategies in the information search and recall processes of auditors (Anderson, 1988; Anderson & Kida, 1989; Kida, 1984b; Trotman & Sng, 1989).

Consistent with the arguments of Edwards (1983) and Fischoff (1982, 1988), the context, the individual, and the interaction of the two appear to contribute to these differences. For example, expertise in the subject matter and the use of realistic, job-related tasks that are familiar to auditors seem highly plausible as contributing factors to auditors' increased sensitivity to base rates (e.g., Joyce & Biddle, 1981b, Experiment 1; Kida, 1984a) and source reliability (e.g., Bamber, 1983), to the effect contrary to anchoring found by Joyce and Biddle (1981a), and to the difference in the judgments of auditors and students found by Frederick and Libby (1986). Similarly, knowledge of

basic statistical principles seems to readily explain the increased sensitivity of both auditors and graduate business students to sample size (Biddle & Joyce, 1979; Gibbins, 1977; Swieringa et al., 1976; Uecker & Kinney, 1977). Furthermore, when professional auditors perform experimental tasks that are analogues to certain typical audit tasks, conservatism seems to override behaviors commonly found in other heuristics research (Anderson, 1988; Bamber, 1983; Biggs et al., 1988; J. Cohen & Kida, 1989; Joyce & Biddle, 1981a, Experiment 2B; Kida, 1984b; Kinney & Uecker, 1982; Trotman & Sng, 1989).

The apparent use of conservatism in a number of auditing contexts where prior research indicates that other heuristics would be used underscores the fact that the use of a particular heuristic or the presence of a bias depends on the decision maker studied, the task performed, and the match between the two. In these studies, the judgment processes of professional auditors were investigated in a variety of audit-decision contexts, and the results point to factors in the audit environment or audit decision making as potential explanatory variables. For example, auditors have large loss functions with respect to audit judgments as a result of serious immediate and long-run implications of those judgments. In the current litigious environment, auditors may be sued by third parties (e.g., investors, creditors) when those parties believe that a company's financial statements have been misstated. This is especially true when financial reports overstate a firm's profitability or economic viability.

Consequences of this magnitude are often not apparent in heuristics experiments using student subjects. Although many of the biases discussed here have been replicated in experiments using real monetary payoffs of amounts considered significant by the participants, such payoffs do not approach the magnitude of the outcomes associated with many professional judgments. For many audit judgments, the costs associated with certain risks are sufficiently large that they seem to significantly influence the nature of audit training and formalized audit procedures. The use of experimental tasks that are realistic and familiar to the auditors may cause them to bring these loss functions to bear on such tasks. Accordingly, findings from heuristics studies that use students performing generic tasks of limited consequence do not necessarily generalize well to this expert decision-making context.

The specific underlying factors that contribute to the apparent differences in the cognitive processes of experienced auditors and students in certain decision contexts and, perhaps more importantly, to differences in the cognitive processes of experts versus novices in general remain largely unidentified. Basic research in cognitive psychology suggests that a store of substantive knowledge that an expert can bring to bear on particular tasks and the manner in which that knowledge is organized may be important determinants. Libby (1989) noted that a principal factor affecting auditor judgments is the wealth of task-related knowledge that experienced auditors bring to their tasks, and that experience benefits certain, but not all, tasks. That is, training and experience provide specific knowledge, which seems to influence only certain behaviors (see also Bedard, 1989; Bonner, 1990). With respect to heuristics and biases, these findings suggest that through training and experience individuals may develop or acquire "specialized" heuristics that prove highly effective for tasks within their domain of expertise. However, they may use other "general or default" heuristics in situations in which they lack the information or expertise to indicate a strategy better suited to a particular task. For example, among the studies reviewed here, the apparent tendency of experienced auditors to attend preferentially to negative information or outcomes (i.e., conservatism) may be an example of a "specialized" heuristic adopted in response to the substantial risks associated with many audit judgments. Furthermore, anecdotal evidence suggests that experience and professional training may play an important role in developing such behaviors.

Less specialized education or training may also contribute to developing expertise. For example, training in the principles of statistics may increase the likelihood that people will take a statistical approach to everyday problems and improve the quality of the statistical solutions (Nisbett, Krantz, Jepson, & Kunda, 1983). However, the extent to which such knowledge is transferable across contexts appears limited. Perhaps it is only through specialized training and experience that people learn how these principles apply to particular situations.

The findings on such issues are preliminary, but their importance to understanding human-judgment processes is becoming well recognized by investigators interested in both basic and applied research. Using a general framework for individual decision making adapted from Hogarth (1987, chapter 10), Ashton and colleagues (1988) partitioned the field of audit decision making into topic areas (e.g., information acquisition and information processing) and identified specific exemplars of contemporary research questions for each area. Included among the questions identified are many that appear directly relevant to expertise and task-knowledge effects as discussed in this review. Examples include: What structure do expert auditors use to organize knowledge? How well can auditors use their stored knowledge to draw analogies between similar experiences or situations? To what extent can training or decision aids influence the manner in which knowledge is represented and processed? To what extent do experience and task attributes influence how knowledge is represented and processed by auditors? Finding answers to these types of questions is essential to understanding the specific characteristics of expertise or task knowledge that explain differences between the judgmental heuristics of experts and novices.

Much of the evidence reviewed here reinforces the arguments made by Edwards (1968, 1983) and others, and seems consistent with Klayman and Ha's (1987, 1989) propositions suggesting that heuristics may be used as an adaptive response to situational variables. However, without the benefit of a model or framework that predicts the use of different heuristics in response to situational variables, the conclusions and explanations presented regarding differences in the heuristics used by experts versus novices have been made only on the basis of ex-post analysis. Nonetheless, considering that many of the studies reported here examined heuristics and biases using auditing analogues of generic tasks examined previously in the laboratory with students, it seems reasonable that comparisons can provide some evidence on the effect of subject expertise and greater task realism on the use of such heuristics or the presence of biases in analogous task environments.

Finally, it is important to note that the heuristics and biases often observed in studies using subjects with little or no expertise or artificial tasks were also evident to some extent in the judgments of experienced professional auditors. Thus, basic heuristics research can be extremely valuable in identifying biases that can have important implications in professional judgment contexts. However, conclusions drawn regarding specific professional judgments should also be based on investigations using tasks and subjects representative of those contexts.

References

Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology, 55,* 303–314.

Alpert, M., & Raiffa, H. (1982). A progress report on the training of probability assessors. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 294–305). Cambridge, England: Cambridge University Press.

Anderson, B. (1988). *An examination of factors affecting auditors' hypothesis testing strategies.* Unpublished doctoral dissertation, University of Massachusetts.

Anderson, B., & Kida, T. (1989). *An examination of the strategies used by auditors to test inherited hypotheses.* Working paper, Boston University.

Ashton, A., & Ashton, R. (1988). Sequential belief revision in auditing. *The Accounting Review, LXIII* (4), 623–641.

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

Ashton, R., Kleinmuntz, D., Sullivan, J., & Tomassini, L. (1988). Audit decision making. In A. R. Abdel-khalik & I. Solomon (Eds.), *Research opportunities in auditing: The second decade* (pp. 95–132). Sarasota, FL: American Accounting Association, Auditing Section.

Bamber, E. M. (1983). Expert judgment in the audit team: A source reliability approach. *Journal of Accounting Research, 21*(2), 396–412.

Bar-Hillel, M. (1973). On the subjective probability of compound events. *Organizational Behavior and Human Performance, 9,* 396–406.

Bar-Hillel, M. (1979). The role of sample size in sample evaluation. *Organizational Behavior and Human Performance, 24,* 245–257.

Bar-Hillel, M. (1980). The base rate fallacy in probability judgments. *Acta Psychologica, 44,* 211–233.

Bedard, J. (1989). Expertise in auditing: Myth or reality. *Accounting, Organizations and Society, 14*(1/2), 113–131.

Berkeley, D., & Humphreys, P. (1982). Structuring decision problems and the 'bias heuristic'. *Acta Psychologica, 50,* 201–252.

Berman, J., Read, S., & Kenny, D. (1983). Processing inconsistent social information. *Journal of Personality and Social Psychology, 45,* 1211–1224.

Beyth-Marom, R. (1981). *The subjective probability of conjunctions.* (Decision Research Report 81-12). Eugene, Oregon: Decision Research.

Biddle, G., & Joyce, E. (1979). *The role of sample size in probabilistic inference in auditing.* Unpublished manuscript, University of Chicago.

Biggs, S., Mock, T., & Watkins, P. (1988). Auditors' use of analytical review in audit program design. *The Accounting Review, LXIII*(1), 148–161.

Biggs, S., & Wild, J. (1985). An investigation of auditor judgment in analytical review. *The Accounting Review, LX*(4), 607–633.

Birnbaum, M., Wong, R., & Wong, L. (1976). Combining information from sources that vary in credibility. *Memory and Cognition, 4,* 330–336.

Birnberg, J., & Shields, M. (1984). The role of attention and memory in accounting decisions. *Accounting, Organizations and Society, 9*(3/4), 365–382.

Bonner, S. (1990). Experience effects in auditing: The role of task-specific knowledge. *The Accounting Review, 65*(1), 72–92.

Butler, S. (1986). Anchoring in the judgmental evaluation of audit samples. *The Accounting Review, LXI*(1), 101–111.

Butt, J., & Campbell, T. (1989). The effects of information order and hypothesis-testing strategies on auditors' judgments. *Accounting, Organizations and Society, 14*(5/6), 471–479.

Carroll, J., & Siegler, R. (1977). Strategies for the use of base rate information. *Organizational Behavior and Human Performance, 19,* 392–402.

Christensen-Szalanski, J. J. J., & Beach, L. R. (1982). Experience and the base-rate fallacy. *Organizational Behavior and Human Performance, 29,* 270–278.

Christensen-Szalanski, J. J. J., & Beach, L. R. (1984). The citation bias: Fad and fashion in the judgment and decision literature. *American Psychologist, 39,* 75–78.

Christensen-Szalanski, J. J. J., Beck, D. E., Christensen-Szalanski, C. M., & Koepsell, T. D. (1983). Effects of expertise and experience on risk judgments. *Journal of Applied Psychology, 68,* 278–284.

Christensen-Szalanski, J. J. J., Diehr, P., Bushyhead, J., & Wood, R. (1982). Two examples of good clinical judgment. *Medical Decision Making, 2,* 275–283.

Cohen, C. (1981). Person categories and social perception: Testing some boundaries of the processing effects of prior knowledge. *Journal of Personality and Social Psychology, 40,* 441–452.

Cohen, J., & Chesnick, E. I. (1970). The doctrine of psychological chances. *British Journal of Psychology, 61,* 323–334.

Cohen, J., Chesnick, E. I., & Haran, D. (1972). A confirmation of the inertial effect in sequential choice and decision. *British Journal of Psychology, 63,* 41–46.

Cohen, J., & Kida, T. (1989). The impact of analytical review results, internal control reliability and experience on auditors' use of analytical review. *Journal of Accounting Research, 27*(2), 263–276.

Cohen, L. (1981). Can human irrationality be experimentally demonstrated? (with open peer commentary). *Behavioral and Brain Sciences, 4,* 317–370.

Ebbesen, E., & Konecni, V. (1975). Decision making and information integration in the courts: The setting of bail. *Journal of Personality and Social Psychology, 32,* 805–821.

Ebbesen, E., & Konecni, V. (1980). On the external validity of decision-making research: What do we know about decisions in the real world. In T. S. Wallsten (Ed.), *Cognitive processes in choice and decision behavior* (pp. 21–45). Hillsdale, NJ: Erlbaum.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representations of human judgment* (pp. 17–52). New York: Wiley.

Edwards, W. (1983). Human cognitive capabilities, representativeness, and ground rules for research. In Humphreys, P. C., Svenson, O., & Vari, A. (Eds.), *Advances in psychology: Analyzing and aiding decision processes, 18,* 507–513.

Einhorn, H. (1976). A synthesis: Accounting and behavioral science. *Journal of Accounting Research, 14*(Suppl.), 196–206.

Fischoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444). New York: Cambridge University Press.

Fischoff, B. (1987). Judgment and decision making. In R. Sternberg & E. Smith (Eds.), *The psychology of human thought* (pp. 153–187). New York: Cambridge University Press.

Fischoff, B., Slovic, P., & Lichtenstein, S. (1979). Subjective sensitivity analysis. *Organizational Behavior and Human Performance, 23,* 339–359.

Frederick, D., & Libby, R. (1986). Expertise and auditors' judgments of conjunctive events. *Journal of Accounting Research, 24*(2), 270–290.

Funder, D. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin, 101,* 75–90.

Gettys, C., Kelly, C. W., III, & Peterson, C. (1973). The best guess hypothesis in multi-stage inference. *Organizational Behavior and Human Performance, 10,* 363–373.

Gibbins, M. (1977). *Human inference, heuristics and auditors' judgment processes.* Paper presented at the CICA auditing research symposium. Toronto, Canada: Canadian Institute of Chartered Accountants.

Ginosar, Z., & Trope, Y. (1980). The effects of base rates and individuating information on judgments about another person. *Journal of Experimental Social Psychology, 16,* 228–242.

Hammerton, M. (1973). A case of radical probability estimation. *Journal of Experimental Psychology, 101,* 252–254.

Hinsz, V., Tindale, R., Nagao, D., Davis, J., & Robertson, B. (1988). The influence of the accuracy of the individuating information on the use of base rate information in probability judgment. *Journal of Experimental Social Psychology, 24,* 127–145.

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

Hogarth, R. (1987). *Judgment and choice* (2nd ed.). New York: Wiley.

Holt, D. (1987). Auditors and base rates revisited. *Accounting, Organizations and Society, 12*(6), 571–578.

Johnson, W. B. (1983). "Representativeness" in judgmental predictions of corporate bankruptcy. *The Accounting Review, LVIII*(1), 78–97.

Joyce, E., & Biddle, G. (1981a). Anchoring and adjustment in probabilistic inference in auditing. *Journal of Accounting Research, 19,* 120–145.

Joyce, E., & Biddle, G. (1981b). Are auditors' judgments sufficiently regressive? *Journal of Accounting Research, 19,* 323–349.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3,* 430–454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237–251.

Kaplan, S., & Reckers, P. (1989). An examination of information search during initial audit planning. *Accounting, Organizations and Society, 14*(5/6), 539–550.

Kida, T. (1984a). The effect of causality and specificity on data use. *Journal of Accounting Research, 22*(1), 145–152.

Kida, T. (1984b). The impact of hypothesis-testing strategies on auditors' use of judgment data. *Journal of Accounting Research, 22*(1), 332–340.

Kinney, W. R., Jr., & Uecker, W. (1982). Mitigating the consequences of anchoring in auditor judgments. *The Accounting Review, LVII*(1), 55–69.

Klayman, J. (1983). *Learning in a probabilistic environment.* Paper presented at the Midwestern Psychological Association, Chicago.

Klayman, J. (1984). Learning from feedback in probabilistic environments. *Acta Psychologica, 56,* 81–92.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94,* 211–228.

Klayman, J., & Ha, Y. (1989). Hypothesis testing in rule discovery: Strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 596–604.

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

Libby, R. (1981). *Accounting and human information processing: Theory and applications.* Englewood Cliffs, NJ: Prentice-Hall.

Libby, R. (1989). *Experimental research and the distinctive features of accounting settings.* Working paper, Cornell University.

Lyon, D., & Slovic, P. (1976). Dominance of accuracy information and neglect of base rates in probability estimation. *Acta Psychologica, 40,* 287–298.

Manis, M., Dovalina, I., Avis, N., & Cardoze, S. (1980). Base rates can effect individual predictions. *Journal of Personality and Social Psychology, 38,* 231–248.

McGinnies, E., & Ward, C. (1980). Better liked than right: Trustworthiness and expertise as factors in credibility. *Personality and Social Psychology Bulletin, 6,* 467–472.

Morier, D., & Borgida, E. (1984). The conjunction fallacy: A task specific phenomenon? *Personality and Social Psychology Bulletin, 10,* 243–252.

Murphy, A., & Winkler, R. (1974). Probability forecasts: A survey of National Weather Service forecasters. *Bulletin of the American Meteorological Society, 55,* 1449–1453.

Murphy, A., & Winkler, R. (1977). Reliability of subjective probability forecasts of precipitation and temperature. *Journal of the Royal Statistical Society, Series C (Applied Statistics), 26,* 41–47.

Nisbett, R., Krantz, D., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90,* 339–363.

Paquette, L., & Kida, T. (1988). The effect of decision strategy and task complexity on decision performance. *Organizational Behavior and Human Decision Processes, 41,* 128–142.

Phelps, R., & Shanteau, J. (1978). Livestock judges: How much information can an expert use? *Organizational Behavior and Human Performance, 21,* 201–219.

Rosenbaum, M., & Levin, I. (1968). Impression formation as a function of source credibility and order of presentation of contradictory information. *Journal of Personality and Social Psychology, 10,* 167–174.

Rosenbaum, M., & Levin, I. (1969). Impression formation as a function of source credibility and the polarity of information. *Journal of Personality and Social Psychology, 12,* 34–37.

Rothbart, M., Evans, M., & Fulero, S. (1979). Recall for confirming events: Memory processes and the maintenance of social stereotypes. *Journal of Experimental Social Psychology, 15,* 343–355.

Schustack, M., & Sternberg, R. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General, 110,* 102–120.

Skov, R., & Sherman, S. (1986). Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation. *Journal of Experimental Social Psychology, 22,* 93–121.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance, 6,* 649–744.

Snyder, M. (1981). Seek, and ye shall find: Testing hypotheses about other people. In E. Higgins, C. Herman, & M. Zanna (Eds.), *Social cognition: The Ontario Symposium* (pp. 277–304). Hillsdale, NJ: Erlbaum.

Snyder, M., & Campbell, B. (1980). Testing hypotheses about other people: The role of hypothesis. *Personality and Social Psychology Bulletin, 6,* 421–426.

Snyder, M., & Cantor, N. (1979). Testing hypotheses about other people: The use of historical knowledge. *Journal of Experimental Social Psychology, 15,* 330–342.

Snyder, M., & Gangestad, S. (1981). Hypothesis-testing processes. In J. Harvey, W. Ickes, & R. Kidd (Eds.), *New directions in attribution research* (Vol. 3, pp. 171–196). Hillsdale, NJ: Erlbaum.

Snyder, M., & Swann, W. B., Jr. (1978). Behavioral confirmation in

social interaction: From social perception to social reality. *Journal of Experimental Social Psychology, 14,* 148–162.

Snyder, M., & Uranowitz, S. (1978). Reconstructing the past: Some cognitive consequences of person perception. *Journal of Personality and Social Psychology, 36,* 941–950.

Snyder, M., & White, P. (1981). Testing hypotheses about other people: Strategies of verification and falsification. *Personality and Social Psychology Bulletin, 7*(1), 39–43.

Strohmer, D., & Newman, L. (1983). Counselor hypothesis testing strategies. *Journal of Counseling Psychology, 30,* 557–565.

Swieringa, R., Gibbins, M., Larsson, L., & Sweeney, J. L. (1976). Experiments in the heuristics of human information processing. *Journal of Accounting Research, 4*(Suppl.), 159–187.

Thorngate, W. (1980). Efficient decision heuristics. *Behavioral Science, 25,* 219–225.

Tomassini, L., Solomon, I., Romney, M., & Krogstad, J. (1982). Calibration of auditors, probabilistic judgments: Some empirical evidence. *Organizational Behavior and Human Performance, 30,* 391–406.

Trope, Y., & Bassok, M. (1982). Confirmatory and diagnosing strategies in social information gathering. *Journal of Personality and Social Psychology, 43,* 22–34.

Trotman, K., & Sng, J. (1989). The effect of hypothesis framing, prior expectations and cue diagnosticity on auditors' information choice. *Accounting, Organizations and Society, 14*(5/6), 565–576.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Tversky, A., & Kahneman, D. (1982). Causal schemas in judgment under uncertainty. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 49–72). New York: Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293–315.

Uecker, W., & Kinney, W. (1977). Judgmental evaluation of sample results: A study of the type and severity of errors made by practicing CPAs. *Accounting, Organizations and Society, 2*(3), 269–275.

Wason, P. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12,* 129–140.

Wason, P., & Johnson-Laird, P. (1972). *Psychology of reasoning: Structure and content.* London: D. T. Batsford.

Youssef, I., & Peterson, C. (1973). Intuitive cascaded inferences. *Organizational Behavior and Human Performance, 10,* 349–358.

Zadny, J., & Gerard, H. (1974). Attributed intentions and informational selectivity. *Journal of Experimental Social Psychology, 10,* 34–52.

Received May 8, 1989
Revision received August 19, 1990
Accepted August 29, 1990 ■