Why Study Expert Decision Making? Some Historical Perspectives and Comments

JAMES SHANTEAU
Kansas State University

AND

THOMAS R. STEWART
University of Albany

This special issue is concerned with two topics of increasing importance in judgment and decision-making (J/DM) research, "Experts and Expert Systems." Although these topics have a shared interest in expertise, their origins are different. The study of domain experts is intertwined with the development of the J/DM field. In contrast, the analysis of expert systems is a recent undertaking that has just begun to affect decision-making research.

The seven papers and two commentaries in this special issue offer varying perspectives on the question of how experts make decisions. All reflect the rapid development of method and theory in this field in recent years. And all take note of the increasing interest in using knowledge about experts to improve expert systems.

Before continuing, it is necessary to define expertise. According to Webster (1979), an expert is someone "displaying special skill or knowledge derived from training or experience." This is consistent with our view that expert judgment applies in situations where there are grounds for saying that some judgments are better than others. That rules out, for instance, the study of experts in astrology or graphology. The domains of expertise represented here are consistent with this definition.

The remainder of this introductory paper is divided into three sections. The first summarizes previous research on experts in the J/DM field.
second contains observations about why this line of research is important. The final section comments on the nine papers in this special issue.

BACKGROUND

The complete history of the J/DM field has yet to be written. When it is, however, there will be an emphasis on the analysis of expertise. At nearly every point in the development of the field, the study of how experts make decisions and judgments has played an important role. Consider the following examples:

Psychometric Research

Systematic research on decision-making experts began with psychometric analyses of validity (accuracy) and reliability (repeatability). In the first known study of experts, Hughes (1917) reported that grain rated highest by corn judges did not always produce the highest crop yields. He also observed considerable variability between corn judges. More recently, Trumbo, Adams, Milner, and Schipper (1962) found that grain judges misgraded nearly 1/3 of wheat samples, with over 1/3 of the samples graded differently when judged a second time; they also found more experienced judges to be more confident, but not necessarily more accurate. Similar results have been observed for other types of agricultural judges (Foss, Wright, & Coles, 1975; Gaeth & Shanteau, 1979; Shanteau & Gaeth, 1981).

The best-known psychometric analyses of experts involved research in the 1950s and 1960s on clinical psychologists. For instance, Goldberg (1959) had 22 subjects (4 practicing clinical psychologists, 10 psychology interns, and 8 naive subjects) examine 30 Bender–Gestalt protocols (a test for cortical brain damage); the accuracy for all groups was between 65 and 70%, where 50% is chance. In an analysis of reliability, Oskamp (1962) found a correlation of .44 between repeated judgments of the same cases by clinicians.

Similar validity and reliability results have been obtained in a variety of other expert domains, such as medical doctors (Einhorn, 1974), clinical psychologists (Oskamp, 1962), parole officers (Carroll & Payne, 1976), and court judges (Ebbesen & Konecni, 1975). Such research also showed that experience is not related to judgment ability (Meehl, 1954).

In related research, Oskamp (1965) reported that confidence ratings of clinical psychologists increased as more information was supplied; the ratings across four successive stages were 33, 39, 46, and 53%. However, the accuracy was unchanged, 26, 23, 28, and 28%, where chance was 20%. Further, there were no differences between expert clinicians and novices (students). Similar findings of misplaced confidence have been observed in a variety of other fields (Phillips, 1982).

Altogether, the experts were lacking in valid confidence but not in accurate confidence.

Linear Model Analyses

Extensive use of linear model analyses began in the 1950s and 1960s. Findings emerged from analyses of expert judgments. For example, they argued that one could estimate the parameters of a linear model and that any simple model failed to account for the richness of the data.

The clinicians' argument was that: (a) the judgments were not linear functions of the data, (b) linear models were not appropriate for the data, (c) linear models were not appropriate for the data, and (d) the clinical model simply added the linear model to the data. This work inspired a great deal of research into the nature of human judgment.

Common sense is reflected in their judgments. For instance, Ebbesen and Khanna (1973) found that only one of three factors was relevant information in the judgment of medical radiologists (also see Einhorn, 1974). Information about the patient and his symptoms was available.

The evidence shows that judgments contain a small number of factors but are not linear. In some cases, the evidence suggests that judgments are made on the basis of a small number of factors, but these factors are not linearly related to the data. For example, in the case of medical radiologists, information about the patient and his symptoms was available. However, the evidence suggests that judgments are made on the basis of a small number of factors, but these factors are not linearly related to the data.
observed in a variety of judgment domains (Lichtenstein, Fischhoff, & Phillips, 1982).

Altogether, the conclusion from psychometric research is that experts are lacking in validity and reliability and that more information increases confidence but not accuracy (Goldberg, 1968).

**Linear Model Analysis**

Extensive use of multiple regression methods for modeling judgment began in the 1950s (Hammond, 1955; Hoffman, 1960). Several notable findings emerged from research on the use of the MMPI by clinical psychologists for diagnosis of mental illness (Meehl, 1954). Clinicians had argued that one could not make proper use of the 11 MMPI scales without extensive training and experience. They discounted the possibility that any simple model, such as a sum of weighted scale scores, could capture the richness of the diagnostic process (Goldberg, 1969).

The clinicians' arguments were challenged by a series of studies showing that: (a) the judgments of experienced clinicians were no better than those of graduate students (Oskamp, 1967), (b) statistical diagnoses based on linear regression models outperformed the diagnoses made by the clinicians (Goldberg, 1969), (c) diagnoses derived from regression models of the clinicians' own judgments outperformed the clinicians (Goldberg, 1970), and (d) the clinicians' performance was matched by a model which simply added the 11 equally weighted scores (Dawes & Corrigan, 1974). This work inspired a large body of research and contributed to the generally negative view of experts among decision-making researchers.

Common sense suggests the greater knowledge of experts should be reflected in their judgments. However, the research has shown that models of expert judgments contain less information than expected. For instance, Ebbesen and Konone (1975) found experienced court judges used only one to three factors when setting bail, despite the presence of other relevant information. Hoffman, Slowic, and Rorer (1968) reported that medical radiologists had two to six dimensions relevant in their judgments (also see Einhorn, 1974). Stockbrokers had six to seven significant dimensions (Slovic, 1969). In each case, more relevant information was available.

The evidence shows that linear models of experts and nonexperts alike contain a small number of significant factors. "A robust finding in research on human judgment is that relatively few cues account for virtually

---

1 In what appears to be the first regression analysis of expert judgment, Wallace (1923) reanalyzed the data of Hughes (1917) on corn judges using a precursor of path analysis. Wallace observed that experts relied primarily on ear length, whereas kernel weight was most highly related to yields.
all of the systematic variance" (Reilly & Doherty, 1989, p. 123). Since the
effect of greater information use by experts was not supported, the
conclusion was that experts are limited in the same manner as novices.

**Biases and Heuristics**

In the decision-making literature, a common explanation for the low
level of performance by experts is that they rely on heuristics in making
judgments. Although these heuristics are often functional, they can lead
to systematic biases or errors. Similar biases have been observed for both
expert and novice decision makers.

In a study of representativeness, Tversky and Kahneman (1971) ob-
served use of the "law of small numbers" by experts. They found that
sophisticated psychologists made inappropriate extensions from large
sample logic to small sample sizes in evaluating hypothetical research.
Parallel results were observed in judgments of undergraduate students.
The representativeness heuristic was offered as an explanation of the
biases in both groups.

In research on pricing decisions, Norcraft and Neale (1987) had 21
real estate agents estimate the value of various properties. The agents
were given starting list prices manipulated to serve as initial anchors. The
realtors’ estimates were biased toward the initial prices, as predicted by
an anchoring-and-adjustment heuristic. The authors conclude (p. 96),
"decision biases and heuristics are more than just parlor tricks and (they)
should play an important role in our understanding of everyday decision
behavior."

Because of heuristics, experts apparently are limited in the same way as
naive subjects. "Numerous studies show that intelligent people have
great difficulty judging probabilities, making predictions, and, otherwise,
attempting to cope with uncertainty. Frequently these difficulties can be
traced to the use of judgmental heuristics" (Slovic, Fischhoff, & Lich-
tenstein, 1985).

**Probability Assessments**

The study of experts also has played a role in research on decisions
under uncertainty. Suboptimal performance in Bayesian probability revi-
sion tasks led to numerous efforts to explain conservatism (Edwards,
Lindman, & Phillips, 1965). A key question concerned generality: Does
conservatism apply to real decision makers faced with realistic judg-
ments? The answer (1968) was that experts do.

Many studies of assessed probabilities have used general knowledge
undergraduates (e.g., Christensen) and experts (Lichtenstein) miscalibrated in similar
manner.

Together, these studies of experts. As Campbell (1968) noted in the
literature, "The expert judgments in medical diagnosis have been worse than the
judgments of laypeople in the same domains that have

**Good-Performance**

Until recently, research on heuristics and decision makers. But
this theme began to emerge from research on the validity of
probability assessments. For example, typical miscalibration
landmark study demonstrated that expert judgments in typical
situations, of excellent improvement in context: the

There also have been upgrades in the field of probability assessments.
For example, a study by Johnson (1974) reported improvements in the
mean reliabilities of judgments in this field, with many researchers

The question of how these results can be applied to new fields of
expertise has become a continuing interest. For example, by Murphy,
WHY STUDY EXPERTS?


ments? The answer supplied by Edwards, Phillips, Hays, and Goodman (1968) was that experienced decision makers did show the same trends as students.

Many studies of calibration have shown "overconfidence"—the assessed probabilities are greater than the proportion of correct answers to general knowledge questions. Similar findings have been observed for undergraduates (e.g., Fischhoff, Slovic, & Lichtenstein, 1977), physicians (Christensen-Szalanski & Bushyhead, 1981), and other types of experts (Lichtenstein et al., 1982). Thus, both experts and novices are miscalibrated in similar ways.

Together, these studies of experts paint a dismal picture of the abilities of experts. As Camerer and Johnson (1991, p. 203) state after reviewing the literature, "The depressing conclusion from these studies is that expert judgments in most clinical and medical domains are no more accurate than those of lightly trained novices... And expert judgments have been worse than those of the simplest statistical models in virtually all domains that have been studied."

Good-Performance Experts

Until recently, research was consistent in depicting experts as flawed decision makers. Beginning in the late 1970s, however, some exceptions to this theme began to emerge. The first notable counterexample came from research on weather forecasters. In 1951, Williams reported that probability assessments exceeded the proportion of precipitation days—a typical miscalibration result. But, in later research by Murphy and Winkler (1977), calibration improved to fall almost on the identity line. This landmark study demonstrated that experts are capable, in at least some situations, of excellent performance. According to Murphy (1980), the improvement in calibration can be attributed to the increased experience of weather forecasters in making probability assessments.3

There also have been exceptions to the general trend of poor psychometric properties in experts. In a study of 63 auditors, for instance, Ashton (1974) reported a mean correlation of .81 between two administrations of a set of 32 cases. Similarly, Joyce (1976) and Messier (1983) observed mean reliabilities of .86 and .90, respectively, for auditors. These values are much higher than the values typically reported for experts in other fields.

The question of how much information an expert uses has been of continuing interest. Phelps and Shanteau (1978) asked livestock judges to evaluate gilts (female pigs) described either by photographs or by verbal

listings of 11 attributes. Using multiple regression analyses, the evaluations of photos were found to be based on 0 to 3 significant cues. Using Information Integration Theory analyses (Anderson, 1981), however, the ratings of verbal stimuli were based on 9 to 11 significant attributes. The results indicated that experts can integrate many dimensions, but that stimulus intercorrelations in real stimuli tend to reduce the number of cues significant in a regression analysis.

In an extensive review of the auditing literature, Smith and Kida (1991) examined the prevalence of heuristics and biases in auditor judgment. Based on their analysis of 25 studies, they report that the “biases found readily in other research are not evident in judgments of professional auditors.” They feel the difference is that experts in these studies performed familiar job-related tasks, as opposed to the artificial tasks used with inexperienced students.

There is an emerging view that some experts can make competent decisions in at least some situations. This view is reflected in recent books on expertise, such as those of Mumpower, Phillips, Renn, and Uppuluri (1987); Schwartz and Griffin (1986); Dowie and Elstein (1988); Chi, Glaser, and Farr (1988); and Ericsson and Smith (1991). And it appears more strongly in the latest books, such as those of Wright and Bolger (1992); Hoffman, Shanteau, Burton, Shadbolt, and Akerstrom (in press); and Shanteau (in preparation). Together, these books present a new, more positive view of the performance of experts. It is that view which is reflected in most of the papers in this special issue.

WHY STUDY EXPERTS?

There are at least three reasons why research on experts is important and will continue to be of interest to judgment and decision researchers. The first involves the generalizability of results from naive subjects to experts. The second reflects the increasing importance of expert systems. The third reason is the recognition that experts are interesting in their own right. Each of these reasons is discussed below.

Generality

There has been much debate in the J/DM field about whether commonly accepted findings extend from student subjects to experts. Most investigators want to establish that their results apply to a variety of subjects, not just students. Analyses of experts provide one important means of establishing generality. When similar results are found, a strong case can be made for the universality of the behavior. As Kahneman (1991) argues in support of the generalizability of heuristics, “there is much evidence that expert judgments affect other people.”

If experts differ from novices in the way they make and new, more general research has shown considerable variation in, the results. However, the excellent judgments of two of the most important counter to the claim that judgments are better for good and poor performance.

There is no doubt that some experts dominate the research literature, but the results from these studies are not enough exceptions to make them irrelevant.

Expert Systems

Using tools from artificial intelligence, a number of tools have been developed to solve problems in various domains. Some well-known systems are based on codified knowledge, such as Mumpower, Phillips, Renn, and Uppuluri (1987) for diagnosing diseases, Duda, Gashnig, & Haralick, and DENDRAL (Friedman & Hoheisel, 1976) for analyzing mass spectrograms. Expert systems are used by researchers in various fields.

The building of an expert system involves converting knowledge into logical propositions. Once reasoning strategies into a rule system, it is necessary first to determine the problem domain and the knowledge representation. It was not long after the first expert systems were developed that the bottlenecks in knowledge acquisition were identified. Discussion of the bottlenecks that are common when building expert systems is necessary. With some exceptions...

---

4 In fact, there is much to suggest that the evaluation of the behavioral aspects of the decision-making process is important. Studies of auditors often did not consider cognitive biases. 
5 This discussion is drawn from Shanteau (in press).
WHY STUDY EXPERTS?

much evidence that experts are not immune to the cognitive illusions that affect other people.\textsuperscript{4}

If experts differ from other subjects, however, then exceptions must be made and new, more general rules must be established. As an example, research has shown consistently that students are poorly calibrated. However, the excellent calibration of weather forecasters provided an important counter to the general rule and led to new insights into reasons for good and poor performance.

There is no doubt that studies of student subjects will continue to dominate the research literature in the future. However, the generality of the results from these studies cannot be taken for granted. There are enough exceptions to make clear the need for separate analyses of experts.

\textit{Expert Systems}\textsuperscript{5}

Using tools from artificial intelligence, computer scientists in the 1960s began developing procedures for simulating expertise. These expert systems are based on codification of the knowledge and decision rules of experts. They are intended either to deal with the kinds of problems that domain experts solve or to operate as "intelligent assistants" to practicing experts. Some well-known examples include MYCIN (Shortliffe, 1976) for diagnosing and treating bacterial infections, PROSPECTOR (Duda, Gashnig, & Hart, 1979) for evaluating geological sites for minerals, and DENDRAL (Fiegenbaum, Buchanan, & Lederberg, 1971) for analyzing mass spectrograms. There are now hundreds of functioning expert systems and thousands more under development (Reddy, 1988).

The building of an expert system involves recasting experts' knowledge into logical propositions (knowledge structures) and translating experts' reasoning strategies into formal rules. But, before programming an expert system, it is necessary first to elicit an expert's knowledge and strategies. It was not long after the first expert systems appeared that the "knowledge acquisition bottleneck" became apparent (Cullen & Bryman, 1988). Discussion of the bottleneck has been a major feature of many review articles and books on expert systems (e.g., Hart, 1989; Kidd, 1987).

With some exceptions (e.g., Mumpower, Phillips, Renn, & Uppuluri,

\textsuperscript{4} In fact, there is much to suggest that heuristics may not generalize to experts. In an evaluation of the behavioral accounting literature, for example, Shanteau (1989) found that studies of auditors often did not replicate the standard findings of the heuristics and biases literature.

\textsuperscript{5} This discussion is drawn from material in Hoffman, Shanteau, Burton, Shadbolt, and Akerstrom (in press).
1987), there has been little connection between decision research on experts and expert system development. The important question is whether decision researchers can contribute to the solution of the knowledge acquisition bottleneck problem. Given nearly a half century of research on experts, it appears that judgment/decision-making research should have something to offer.

The irony, however, is that most J/DM research has focused on the deficiencies of experts (Chan, 1982). What expert systems developers want to know, however, is what experts can do, not what they can’t do. In contrast, research in cognitive psychology has elicited many special ways by which experts think and solve problems (Chi et al., 1988). This difference in emphasis may explain why judgment/decision research has had little effect on expert system development, whereas cognitive psychology has had a major impact.

Nonetheless, the potential for application is considerable—if it leads to improvements in the design and implementation of expert systems. This potential has been sufficient to generate growing interest in studies of decision experts.

*Intrinsic Interest*

The last and perhaps most important reason for an analysis of experts is that they are interesting in their own right. Many researchers are intrigued by domain specialists and want to know more about how they make decisions.

Aside from personal interest, there are broader considerations. The complexity of modern society is forcing people to rely on judgments of experts in numerous ways. We base many of our daily activities on the decisions of physicians, weather forecasters, etc. That means the quality of our lives depends in large part on the decisions of experts, a reliance that is likely to become more pronounced in the future.

While we depend on experts in many ways, the research findings show that there often is room for improvement in expert decisions. This leads to questions such as: “How good is expert judgment?” “Can expertise be improved?” “How can novices learn to become expert more quickly?” “Why do experts disagree?” “Can (or should) such disagreements be reduced?”

The answers to such questions will not only be of interest to J/DM researchers, they also have important societal implications for the selection, development, application, and enhancement of expert decision makers. In short, research on experts matters—both to researchers and to society at large.

*Present Papers*

The papers in this volume address these issues. First, they are characterized by the importance of experts in decision making. Many papers do not always stress the conditions under which experts might be of help, but a few papers exemplify the use of experts in decision making. Second, the papers are directed at understanding decision makers. Third, the authors of the present papers are often aware of the limitations of existing theories, and they attempt to correct these limitations. Finally, the papers illustrate the need for further research in understanding decision making.

*Concluding Comments*

One final facet of professional subjects is the use of experts in other studies using naive compared to expert methods. Many papers have argued that the use of experts is required for the development of expert systems. Let us not turn this argument into a general statement. Although there are a number of cases of expert systems that are not as good as expected, it is clear that readers will have to be cautious in the interpretation of studies of expert systems.

*Although a variety of studies have been conducted, the results are not always consistent. It is notable by their omission, which provide recent research that may help to clarify the role of experts in decision making.*
Present Papers

The papers in this special issue have several properties in common. First, they are concerned with when, and where, experts are superior decision makers. As is well known (e.g., Camerer & Johnson, 1991), experts do not always do well. It is important, therefore, to understand the conditions under which experts do well and do not do well. The present papers contribute to that understanding (see the papers by Edwards; Hammond; Harvey; and Shanteau).

Second, the studies described here illustrate the application of "second-generation" research methods to expert decision makers. First-generation studies used conventional procedures to investigate traditional J/DM issues, such as linear models and biases/heuristics. In contrast, the methods in most of these papers use methods developed for investigating questions directed specifically at experts and expertise (see the papers by Abdolmohammadi & Shanteau; Bradshaw & Shaw; and Stewart, Moninger, Heideman, & Reagan-Cirincione).

Third, the authors are experienced in conducting research on experts. With some exceptions (e.g., Oskamp, 1962, 1965, 1967), few previous investigators engaged in programmatic research on experts. In contrast, the present papers reflect the continuing commitment of these authors to understanding expertise (see Adelman & Bresnick; and Johnson, Grazioi, Jamal, & Zualkernan).

Concluding Comment

One final facet of these papers deserves comment. Any study of professional subjects takes extra care and time to conduct. Compared to studies using naive (undergraduate) subjects, analyses of experts require greater methodological effort. It is not easy to gain the cooperation of professionals and to set up relevant tasks. Nonetheless, the present studies illustrate that ample rewards follow from the extra effort of investigating experts.

Let us now turn to what we consider to be an outstanding set of papers. Although there are no final answers to questions about how experts make decisions or how best to design expert systems, these papers provide valuable insights into how experts think and decide. We have no doubt that readers will have a better understanding of the potential for contribution of studies of experts.

---

6 Although a variety of expert domains are represented in the present papers, two are notable by their omission. For medical decision making, there are several publications which provide recent reviews (e.g., Dowie & Elstein, 1988; Schwartz & Griffin, 1986). For clinical psychology, the review by Camerer and Johnson (1991) raises questions about the level of expertise.
REFERENCES


WHY STUDY EXPERTS?

---


Received: November 14, 1991

---

**Organizational Behavior**

**Effects of Improving the Training of Meteorologists:**

Center for Policy Research, Oregon State University, Corvallis

Phillips Laboratory, OSU

NOAA/ERL

Center for Policy Research, Oregon State University, Corvallis

The relation between meteorologists and people in the community has always been a delicate one. The public often looks to the meteorologist as a source of knowledge and expertise. However, this relationship is often fraught with misunderstanding and confusion. The role of the meteorologist is not always clear to the public, and this can lead to a lack of trust and decreased confidence in their forecasts.

As advances in technology and communication have progressed, it has become increasingly important for meteorologists to be able to communicate effectively with the public. This is especially true in times of severe weather, when the public may be particularly vulnerable. However, the communication between meteorologists and the public is often hindered by a lack of understanding on both sides.

As advances in technology and communication have progressed, it has become increasingly important for meteorologists to be able to communicate effectively with the public. This is especially true in times of severe weather, when the public may be particularly vulnerable. However, the communication between meteorologists and the public is often hindered by a lack of understanding on both sides.

Reprint requests should be addressed to University at Albany, State University of New York.