What Does It Mean When Experts Disagree?

James Shanteau

Kansas State University

The author can be contacted at:

Department of Psychology

Bluemont Hall 492

Kansas State University

Manhattan, KS 66506

(Ph: 785/532-0618)

(Fax: 785/532-5401)

(E-mail: shanteau@ksu.edu)

Running head: EXPERT DISAGREEMENT

In: Salas, E., & Klein, G. (Eds.). (2001). *Linking expertise and naturalistic decision making.* Mahwah, N.J.: Lawrence Erlbaum Associates.

Abstract

Why is it that experts so often seem to disagree? Although decision researchers often view such disagreements as a reason to mistrust experts, this chapter argues instead that such disagreements reflect the way that experts think and work. Two factors have limited the ability of previous researchers to understand expert decision making. The first is a reliance on economics and statistics for developing theoretical standards. The second is the acceptance of the Generalized Normal Human Adult Mind (*GNHAM*) as a goal for empirical research. Together, these two factors have prevented most decision researchers from understanding expertise.

The purposes of this chapter are (1) to review the evidence on disagreement by domain experts, (2) to outline a commonly-accepted hypothesis labeled *ESC* (Experts Should Converge), (3) to discuss 10 structural and functional factors related to why experts often disagree, (4) to explore domain differences in the degree of agreement between experts, (5) to consider some implications of previous views of experts, (6) to suggest an alternate hypothesis labeled *MSM* (Multiple Solution Model), and (7) to look at future directions for research on expertise.

What Does It Mean When Experts Disagree?

Decision researchers often seem perplexed when they observe sizable and consistent disagreements between subjects. This is particularly true when the subjects are experts. As shown by the following quotations, however, disagreements have been viewed historically as necessary.

"By different methods different men excel" (Churchill, 1764)

"The history of scholarship is a record of disagreements" (Hughes, 1936)

"The tough-minded . . . respect differences" (Benedict, 1940)

The position taken here is that disagreement between experts is not a problem, but rather is a normal part of their job.

The purpose of this chapter is to explore why it is natural for experts to disagree with each other. The chapter is organized as follows. First, there is a review of the literature on disagreement between experts with a discussion of a commonly held hypothesis about expertise. Second, the chapter presents 10 structural and functional factors that help explain why experts often disagree. Third, domain differences in the extent to which experts agree are considered. Fourth, an alternate hypothesis is offered that more closely corresponds to how experts view their tasks. Finally, the chapter concludes with implications for future research directions.

Background

Since the start of systematic analyses of decision making in the 1950's, investigators have expressed surprise and dismay at the extent to which domain experts disagree. For example, if we ask two financial experts to assess an investment, the expectation of most investigators is that they should make the same recommendation. If they arrive at different recommendations, then we wonder whether they are as skilled as they claim.

In a seminal paper, Einhorn (1974) argued that *consensus* or between-expert reliability is a necessary condition for expertise. He reported, however, significant differences in diagnoses by three expert medical pathologists. The average between-expert correlation (r) was .55 (where .0 is chance and 1.0 is perfect).

Similar evidence was presented in a study of four expert livestock judges evaluating overall breeding quality of swine (Phelps, 1977). Despite a high level of internal consistency (average r = .96), the consensus was much lower, r = .50. Apparently, livestock experts have internally consistent strategies, but do not agree with each other about what those strategies should be.

Comparable results have been reported by researchers for other domains. For example, Hoffman, Slovic, and Rorer (1968) and Goldberg and Werts (1966) reported consensus values of less than .40 for judgments by professional stockbrokers and clinical psychologists, respectively.

Several studies of financial experts (auditors) have reported somewhat higher between-expert consensus. For instance, Kida (1980) asked 27 audit partners to evaluate 40 financial profiles based on five accounting ratios; the average between-expert correlation was .76. Ashton (1974) had 63 professional auditors evaluate the strength of internal controls for 32 cases; the mean r between auditors was .70. Finally, Libby, Artman, and Willingham (1985) observed an average correlation of .68 between 12 auditors making control reliance judgments.

Several studies have explored whether agreement increases with experience. Ettenson, Shanteau, and Krogstad (1987) compared 11 accounting students with 10 (mid-level) audit seniors and 10 audit partners; the mean between-judge correlations increased from .66 to .76 to .83, respectively. Messier (1983) reported similar results – audit partners with more than 15 years experience had greater consensus than partners with less experience. In contrast, Hamilton and Wright (1982) found no difference in consensus for internal control assessments by three groups of auditors varying in experience; the average correlation for all three groups was .72.

These results suggest three conclusions: First, experts in a variety of non-financial domains often disagree; the consensus correlations range from .40 to .55. Second, the agreement between financial experts is higher, with correlations ranging from .68 to .83. Third, there is suggestive evidence that increased experience may lead to greater consensus. Still, most researchers conclude that because of the sizable disagreement between experts, there is reason to be concerned about the competence of experts. In the next section, I will explore the hypothesis behind this concern.

Experts-Should-Converge Hypothesis

The unimpressive consensus correlations for non-financial experts led many researchers to question the abilities of experts in general. Following Einhorn's logic, these investigators assume that agreement is a necessary condition for expertise. The lack of agreement, therefore, suggests that "experts are no damn good" (Gettys, personal communication, 1980). This analysis and interpretation of reliability data are derived from an implicit hypothesis about experts. The hypothesis, labeled the *Experts-Should-Converge* (ESC), is based on the following five arguments:

(1) For most tasks performed by experts, there is assumed to be a "gold standard" or unique "ground truth." If this truth is easy to access, we can get it directly. For expert tasks, however, the truth is outside the realm of common knowledge or direct sensory experience of most people. Thus, unique correct answers exist, at least in theory, but they are difficult to obtain.

(2) Because of their special skills and experience, experts should be able to tell the rest of us about this "ground truth." That is, experts can access what others cannot access.

(3) Since by definition there can be only one "ground truth," all experts should give us the single correct answer. The special abilities of experts should allow them to obtain the same truth.

(4) If experts disagree, then someone is wrong – they cannot all be correct. Some or all (or none) of them must not really be experts. Thus, disagreements are a reflection of incompetence.

(5) Since non-experts do not know which of the so-called "experts" are correct, the only safe course of action is to distrust all of them. That is, disagreement between experts implies that we should be suspicious of their claimed special abilities.

This hypothesis, of course, is not a formal chain of logic. But, it is implicit in the way that most researchers reason about results reporting disagreements between domain experts.

Contributing Factors

The ESC hypothesis is supported by two disconnected lines of thought, one of which affects theoretical reasoning and the other influences empirical research. First, research on decision making has been linked historically to theories from economics (e.g., expected utility theory) and statistics (e.g., Bayes theorem) (see Edwards, 1983). When applied to well-specified problems, such theories lead to unique solutions. Given a specified set of antecedent conditions, these formalisms provide point estimates of the optimal decision strategy. Interestingly, a great deal of laboratory work has been devoted to showing the inadequacy of such theories (Yates, 1990). Nevertheless, economics and statistics continue to provide a point of comparison in much of decision research.

This reasoning has also been applied to analyses of domain experts. Although most experts work on problems that are not well specified, researchers nonetheless assume that unique solutions exist – just as they do for simple cases. Some investigators have recently questioned the relevance of economic/statistical theories as a basis for making real-world decisions (Klein, Orasanu, Calderwood, & Zsambok, 1993). Other scholars have questioned the narrow assumptions required and have argued instead that these theories as not descriptive of real situations (Gigerenzer, 1993). While economic and statistical theories often cannot be applied to the domains where experts work, it is nonetheless commonly assumed that point solutions exist, at least in the abstract.

Second, nearly a century of experimental psychology (particularly in America) has been focused on analysis of the "Generalized Normal Human Adult Mind" – *GNHAM* (Edwards, 1983). According to Boring (1927), this emphasis dates to the beginnings of experimental psychology, particularly Wilhelm Wundt and Edward Titchener: "Titchener's interest lay in the generalized, normal, human adult mind that had also been Wundt's main concern." Later writers reflected this theme: "Psychology may gather its materials from many sources, but its aim is to understand the generalized human mind" (Heidbreder, 1933, p. 125-126).

In fact, there is no evidence that Wundt ever used this term (Shanteau, 1999). References to the concept, however, are common in Titchener's writing, e.g., "psychology is concerned with the normal, human, adult mind" (1916, p. 2). Modern historians of psychology now believe that Boring created a "myth of origin" (Samelson, 1974) to provide a justification for Titchener's (who was his mentor) place in psychology. As noted by Hebb (1972, p. 291), although Boring's history is "commonly considered the standard work, and beautifully clear it its exposition, this book is thoroughly misleading in its emphasis (on the relationship between) . . . Wundt and Titchener."

According to the GNHAM view, the goal of behavioral research is to investigate commonalities among humans, not differences between them. That is, the focus of psychology belongs on the generalized mind, not on individual minds. Because the Titchener-Boring view dominated (at least in America), research was directed to the search for "universal truths" of behavior. As a result, experimental psychology developed neither the paradigms nor the theories to deal with outlier behavior. And expertise, by definition, is outlier behavior. (For a further discussion of the influence of GNHAM on decision research, see Shanteau (1999).)

Research in decision making, therefore, has been influenced by two research streams that assumed (1) that decision problems should have unique correct answers and (2) that differences between individuals are not important to empirical investigations. The persistence of observed differences between experts provided evidence inconsistent with these views. Thus, disagreements between experts led to the conclusion that something must be wrong, i.e., expertise is a sham.

In the next two sections, I will explore 10 factors behind why researchers should not be surprised when experts disagree. This will be followed by an alternate hypothesis along with some supporting data. The chapter then concludes with a discussion of implications and conclusions.

Structural Factors

Analysis of the context in which most experts work provides five structural factors behind why experts may disagree. These factors reflect the situational constraints under which experts work.

(1) In the contexts where experts work, the "ground truth" is a fiction. Single-point optimal solutions do not exist. Despite the tremendous analytic ability of master players and the incredible computation speed of computer programs such as *Deep Blue*, for instance, the game of chess still does not yield optimal solutions. If this is true for a well-structure game such as chess, how can it be possible to find a "correct answer" in an ill-structured setting? The reason we need experts in the first place is that they offer us answers that we could not obtain any other way.

(2) A distinction can be made between the different levels of decisions made by experts. Using terminology from medicine, it is possible to distinguish between three levels: The first is *diagnosis*

(what is it?) based on categorization and/or classification. The second is *prognosis* (what is the likely outcome?) based on forecasting future scenarios. And the third is *treatment* (what to do about it?) involving selection of a course of action. There are thousands of diagnoses and hundreds of prognoses, but relatively few treatments. It should not be surprising, therefore, to find that experts might disagree at one level (diagnosis), but agree at another (treatment).

(3) Despite the assumption behind the ESC hypothesis, experts are seldom asked to make single-outcome decisions. The concept of a "point prediction" is largely a fiction created for the convenience of the researcher and is not descriptive of the tasks that experts do. As Golde (1970) noted, although "an expert does sometimes make decisions, his (her) role is usually much more of an advisor . . . (they) let me know the kinds of decisions or actions that I must take." In other words, the job of the expert is to clarify alternatives and describe possible outcomes for clients.

(4) As Klein (1993) emphasized, experts generally work in dynamic situations with frequent updating. Thus, the problems faced by experts are unpredictable, with evolving constraints. In such situations there are rarely any best or correct answers. Therefore, while the ESC model assumes a stationary target, the reality faced by experts is generally more like a moving target.

(5) A long-term perspective reveals that experts work in realms where the basic science is still evolving. For instance, the rapid changes in medicine mean that the current "best answers" are soon obsolete. Why should we expect experts to agree on a single "correct answer," say for the treatment of AIDS, when new knowledge will likely provide better solutions tomorrow?

Functional Factors

An analysis of the strategies used by experts to make decisions reveals at least five functional or process factors behind why experts may disagree. These factors have to do with how experts think about the decisions and judgments that make.

(1) Most experts operate as if they have flat loss functions for deviations from optimality. They see small deviations as having minor consequences. In comparison, researchers often operate as if they have steep loss functions. That is, they view any deviation from optimality, no matter how slight, as having large consequences. A similar argument can be made for how disagreements between experts are viewed.

(2) While most researchers view an "error" as any deviation between behavior and the "correct answer," experts have a different definition of error. As noted above, experts are usually more concerned about avoiding big mistakes, whereas researchers are looking for perfection. Thus, the same outcome could well be called an "error" by the researcher and a "success" by the expert. In other words, experts may see agreement where investigators see disagreement.

(3) In many (most?) settings, experts expect to disagree with each other. In a discussion among any two academics, for instance, we know that they invariably will find something to argue about. Even when they agree on 99% of the issues, they will quickly find the last 1% and disagree about that. Similarly, experts in most any field bypass items of agreement to focus instead on disagreements. Thus, experts view disagreements as a normal part of their job.

(4) Disagreements are often the route by which experts increase understanding of their field. By seeking out areas of disagreement, experts examine the limits of their own knowledge and stretch

their range of competency. Therefore, experts see disagreements as a key step in increasing their grasp of their field.

(5) Once a domain has advanced to the point where all issues are resolved, there are few disagreements among experts because there is nothing to argue about. When a field has developed to that degree, however, the answers are known and agreed upon. Thus, total agreement among experts is an indication that there is no longer much of a role for experts to play in that domain.

Domain Differences

We all know that experts in different domains perform different tasks. Yet decision researchers persist in treating all experts alike, so that the term "expert" is used generically. For instance, Kahneman (1991) concluded "there is much evidence that experts are not immune to the cognitive illusions that affect other people." Yet nearly all researchers are aware that at least some experts, e.g., weather forecasters, show little sign of biases or "cognitive illusions." Thus, despite the generalizations drawn about experts, we know there are many exceptions to the rule.

In an effort to account for these domain differences, I constructed Table 1 to differentiate between those domains where experts do well and those where experts do not. The table is based on a continuum from high to low competence (see Shanteau 1992a,b for earlier versions of this table). In the left column are those domains where experts make aided decisions using Decision Support Systems (DSS) or other computerized tools, e.g., in weather forecasting. The next column contains domains where experts make skilled but largely unaided decisions, e.g., livestock judges. The third column lists domains where experts show limited competence, e.g., clinical psychologists. The behavior of experts in the last column is close to random, e.g., stockbrokers.

It should be noted that assignment of domains within the table was based on a review of the literature, i.e., the assessment of competence is drawn from researchers who study each domain. Also three domains –nurses, physicians, and auditors – appear in several columns. That is because the literature in these fields provides mixed evidence of the competence of experts.

==

Insert Table 1 About Here

==

There are many ways to describe the differences in this table (see Shanteau, 1992a,b). For present purposes, it makes most sense to note that domains to the left side possess more stable (*static*) properties. That is, the stimuli and the problem "hold still" for experts to evaluate. The domains to the right side, however, involve more changeable (*dynamic*) properties. Thus, the stimuli and problem are less stable, harder to specify, and more like "moving targets." It makes sense, therefore, that expert agreement will be higher on the left side and lower on the right side.

To test this idea, Table 2 gives reliability values from studies of domain experts in the four categories of Table 1. Two domains are listed under each category, with the between-expert agreement (consensus) given as average correlations. As can be seen, the average consensus r value for weather forecasters is .95, whereas average values for livestock judges, clinical psychologists, and stock forecasters are .50, .40, and .32, respectively. Comparable results appear for other domains on the second line. The trend supports the prediction outlined above – better structured domains lead to high consensus and less structured domains to less consensus.

==

Insert Table 2 About Here

==

For comparison, the average within-expert reliability (consistency) correlations for these same domains are listed in Table 3. The trends are similar, with better structured domains leading to higher internal consistency. As expected, the consistency values (except for pathologists) are higher than the corresponding consensus values in Table 2. In two domains (livestock judges and polygraphers), there are notable discrepancies between the consensus and consistency correlations; the reasons for these differences are not clear.

==
Insert Table 3 About Here
==

Decision Researchers' View of Experts

Compared to other fields of inquiry, decision researchers have taken an idiosyncratic view of the abilities of experts. Investigators in artificial intelligence, expert system design, cognitive science, systems analysis, and computer science have all concluded that experts are superior decision makers. That is why knowledge engineers build computer simulations around what experts know. Further, most domain-specific researchers (such as in medicine and weather forecasting) view experts as possessing unique information essential for making good decisions. In short, investigators in these fields see human expertise as something to be emulated.

In contrast, decision researchers, especially in America, have concluded that experts are flawed and prone to making simple errors (e.g., Kahneman, 1991). Moreover, experts and novices are viewed as sharing the same shortcomings. For instance, Tversky (quoted in Gardner, 1985, p. 360) stated, 'whenever there is a simple error that most laymen fall for, there is always a slightly more sophisticated version of the same problem that experts fall for."

Investigators have overlooked the fact that in most real-world problems, unique solutions do not exist. Instead, there are multiple solution paths. For instance, in medicine there may be many ways to treat an illness – when one approach does not work, a physician seeks another. It should not be surprising, therefore, to find experts disagreeing about which is the appropriate course of action to take. Other inquiry systems, therefore, accept multiple points of view across experts as inevitable. In contrast, decision researchers with their simplified, single-answer view of the world find a multiple-solution perspective difficult to understand.

Multiple-Solution Hypothesis

The position taken here is that previous researchers have unknowingly adopted an ESC view of expertise. According to ESC, disagreement between experts is a sign that something is wrong. This leads to the conclusion that experts are not as skilled or as competent as they claim to be.

In this section, I will propose an alternate hypothesis that is closer to the view of how experts see themselves. There are five arguments behind this *Multiple Solution Model* – MSM:

(1) The primary job of an expert is not to make decisions but to help clients reach a broadly defined target state. For example, the goal of the client may be to design a better investment portfolio or to find a better loan strategy. These goals do not involve single answers, but instead require something more elaborate from the expert, such as a strategic plan.

(2) To reach the client's goal state requires dealing with multiple, constantly changing, dynamic factors. As noted by Klein, et al (1993), the situations faced by experts are different and more complex than the simplified settings studied in research laboratories. Thus, experts work on problems that are considerable more complex than those studied in lab settings.

(3) Using their knowledge and experience, the role of the expert is to recognize patterns and find consistencies in a dynamic problem space. The expert's job is to clarify the issues for the client. In other words, the challenge for an expert is "to make sense out of chaos."

(4) Based on their experience and insights into the nature of problems, experts try to help clients clarify their thinking. For instance, an expert will often identify several alternate paths to the desired goal states. The expert's role is to lay out the options and the consequences in a clear and comprehensible fashion for the client.

(5) In the end, it is the client, not the expert, who actually makes most decisions. The expert offers insights and observations, but the client makes and implements the final choice(s). Thus, the final responsibility for the decision rests on the client, not the expert.

The view is nicely summarized by the management consultant Golde (1970): "We seem to expect too much and the wrong things of our experts." As expressed by MSM, experts generally act more like knowledgeable consultants. Rarely do they function as the "all-knowing" single-answer decision makers envisioned by most researchers. Instead, experts help clients by giving them the insights and information needed to make their own decisions.

When experts disagree, therefore, it is because they see different paths to the client's goal state. In turn, savvy clients may seek out the views of various experts precisely because they want different perspectives on their problems. Thus, disagreements between experts are not only expected, but may actually be useful.

Conclusions

By relying on comparisons to economics/statistical analysis and by incorporating GNAHM assumptions into their research designs, decision investigators have unknowingly adopted a distorted view of what experts do. The remainder of the chapter will look further at the implications arising from these two assumptions.

Economic/Statistical Thinking

By drawing a parallel to economic/statistical theory, decision researchers have adopted a "single correct answer" approach to assessing expertise. When an expert (or anyone else) gives an answer different from the "correct answer," he/she is said to have a 'bias" (Tversky & Kahneman, 1974). And when two or more experts give different answers, the claim of experts to special competence is questioned (Einhorn, 1974).

The position here is that researchers have relied on an incorrect view of how experts function. For instance, the environment in which experts work is much different from that incorporated into traditional laboratory research. The complex, changeable environment that experts operate in is considerably more complicated than the small, stable context constructed by investigators. In reality, problems rarely are simple enough to lead to a single correct answer. Instead, there are multiple answers (or at least multiple routes to answers). If so, it should not be surprising to find that experts often take different approaches (as envisioned by MSM) to making recommendations.

The underlying problem is that researchers misunderstand what experts do and what is expected of them. Investigators seem to think that experts see the world as they do – with simplifying assumptions and normative solutions. However, experts generally have a different world view – with many complexities and contingencies, but few optimal solutions. In addition, experts have a flexible approach to adaptation (e.g., hedge clipping) and are better at managing uncertainty.

From the MSM perspective, disagreements between experts are to be expected. Although researchers view disagreements as evidence of incompetence, experts see disagreements as more-or-less inevitable, and even as a useful, part of the job.

GNAHM Research Paradigm

By adopting the GNAHM approach to research, most investigators have largely ignored individual differences. The average results is emphasized, variability is not. More important, researchers almost never look at the distribution of responses over subjects. Consequently, exceptions are overlooked and statements made about results that imply that all people follow the same pattern of behavior.

Aside from risk preferences (risk adverse vs. risk seeking), there are few individual-subject variables that have received much attention in the decision-making literature. And even risk preferences are more talked about than studied. Yet there is an emerging research stream showing consistent individual differences in decision making strategies. For instance, Yates, Lee, and Bush (1997) found sizable and stable cross-cultural differences between subjects coming from different backgrounds. Such investigations, however, have yet to become of mainstream interest.

As a field, decision research has focused on the behavior of student subjects. Rather than using tools to discover what makes experts unique, researchers rely on paradigms derived from studies of non-expert behavior. Consequently, distinct decision-making paradigms for investigating domain experts do not exist. Instead, the methods to study experts are borrowed from standard procedures. Because these borrowed paradigms are often ill-suited to investigate expertise, it should not be surprising to find that our understanding of experts has not advanced at the same rate as our understanding of non-expert subjects.

That is not to say that research methods developed from a GNAHM perspective cannot be adapted to study expert behavior. The problem is not the methodology, per se. Rather, it is how researchers use the methodology – any technique can be misused (Birnbaum, 1973).

The dangers of reliance on GNAHM thinking for researchers have been recognized for some time (Edwards, 1983). However, the warnings have not been heard. Edwards (p. 509) offered two messages: "One is that psychologists have failed to heed the urging of Egon Brunswik (1955) that generalizations from laboratory tasks should consider the degree to which the task . . . resembles or represents the context to which the generalizations are made. The other message is that experts can in fact do a remarkably good job of assessing and working with probabilities" (e.g., in weather forecasting).

Final Comments

To gain insights into expertise, it is necessary to understand what it is that domain experts do. In my experience, decision researchers rarely take the opportunity to gain an in-depth understanding of how experts think. If they did, they would learn that single-solution approaches to defining "correct" and "incorrect" answers do not work. They would also learn that the GNAHM perspective is a trap that limits understanding of individual differences generally and expertise specifically.

Therefore, disagreement between experts should not be viewed as a source of concern about the competence of experts. It is time for researchers to rethink their view that lack of agreement leads to supposed incompetence of experts. Persistence in such beliefs says more about the biases of investigators than it does about experts. If disagreement between experts is not a focal issue, then what should be? Let me suggest three useful goals for future research on expertise.

First, various investigators have concluded that the superiority of domain experts depends on their ability to distinguish between relevant and irrelevant information (Ettenson, Shanteau, & Krogstad, 1987; Jacavone & Dostal, 1992; Mosier, 1997; Schwartz & Griffin, 1986; Shanteau, 1992b). One goal in future research should be to learn how experts make these discriminations and to find ways to enhance the process.

A second goal should be to understand the kinds of intellectual and physical tools used by experts to enhance their judgments. Experts seldom, if ever, make unaided judgments of the sort emphasized in laboratory research. In fact, researchers make use of the very tools denied their subjects. "The experimenters themselves, using tools and expertise, are able to perform (laboratory) tasks rather well" (Edwards, 1983, p. 511). The type of tools used by experts needs to be better understood.

The final goal should be to develop insights into domain differences. As argued by Edwards (1983, p. 512), "we have no choice but to develop a taxonomy of intellectual tasks themselves. Only with the aid of such a taxonomy can we think with reasonable sophistication about how to identify among the myriad types of experts and the myriad types of tasks . . . just exact what kinds of people and tasks deserve our attention." The analyses in Tables 1, 2, and 3 offer one perspective on such a taxonomy.

Research on such goals will help us expand our understanding of expertise. In contrast, concern about the supposed incompetence of experts based on disagreement offers little opportunity for enhancing our understanding of expertise. As argued here, the future of research on experts lies in other directions. Specifically, we should focus our efforts on analyses of relevance/irrelevance, tool usage, and domain differences. These are directions that future researchers should explore.

References

Ashton, R. H. (1974). An experimental study of internal control judgments. *Journal of Accounting Research, 12*, 143-157.

Birnbaum, M. (1973). The devil rides again: Correlation as an index of fit. *Psychological Bulletin*, *79*, 239-242.

Boring, E. G. (1927). *A history of experimental psychology*. NY: Appleton-Century-Crofts.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology'. *Psychological Review, 62*, 193-217.

Edwards, W. (1983). Human cognitive capacities, representativeness, and ground rules for research In Humphreys, P., Svenson, O., & Vari, A. (Eds.), *Analyzing and aiding decision processes*. Budapest: Akademiai Kiado.

Einhorn, J. (1974). Expert judgment: Some necessary conditions and an example. *Journal of Applied Psychology, 59*, 562-571.

Ettenson, R., Shanteau, J., & Krogstad, J. (1987). Expert judgment: Is more information better? *Psychological Reports, 60*, 227-238.

Goldberg, L. R., & Werts, C. E. (1966). The reliability of clinicians' judgments: A multitrait-multimethod approach. *Journal of Clinical Psychology, 30*, 199-206.

Gardner, H. (1985). *The mind's new science: History of cognitive revolution.* NY: Basic Books.

Gigerenzer, G. (1993). The superego, the ego, and the id in statistical reasoning, in Keren, G. & Lewis, C. (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues.* Hillsdale, NJ: Erlbaum.

Golde, R. A. (1969). *Can you be sure of your experts?* NY: Award Books.

Hamilton, R. E., & Wright, W. F. (1982). Internal control judgments and effects of experience: Replications and extensions. *Journal of Accounting Research, 20*, 756-765.

Hebb, D. O. (1972). *Textbook of psychology* (3rd Ed.). Philadelphia: W. B. Saunders Co.

Heidbreder, E. (1933). *Seven psychologies*. NY: Appleton-Century Co.

Hoffman, P., Slovic, P., & Rorer, L. (1968). An analysis of variance model for the assessment of configural cue utilization in clinical judgment. *Psychological Bulletin, 69*, 338-349.

Jacavone, J., & Dostal, M. (1992). A descriptive study of nursing judgment in assessment and management of cardiac pain. *Advances in Nursing Science, 15*, 54-63.

Kahneman, D. (1991). Judgment and decision making: A personal view. *Psychological Science, 2*, 142-145.

Kida, T. (1980). An investigation into auditors' continuity and related qualification judgments. *Journal of Accounting Research, 8*, 506-523.

Klein, G. A., Orasanu, J., Calderwood, R., & Zsambok, C. E. (1993). *Decision making in action: Models and methods*. Norwood, NJ: Ablex Publishing Corp.

Libby, R., Artman, J. T., & Willingham, J. J. (1985). Process susceptibility, control risk, and audit planning. *The Accounting Review, 60*, 212-230.

Lykken, D. T. (1979). The detection of deception. *Psychological Bulletin, 80*, 47-53.

Messier, W. F. (1983). The effect of experience and firm type on materiality/disclosure judgments. *Journal of Accounting Research, 21*, 611-618.

Mosier, K. L. (1997). Myths of expert decision making and automated decision aids. In C. Zsambok & G. Klein (Eds.), *Naturalistic decision making*. Hillsdale, NJ: Erlbaum.

Phelps, R. H. (1977). *Expert livestock judgment: A descriptive analysis of the development of expertise*. Unpublished doctoral dissertation. Manhattan: Kansas State University.

Phelps, R. H., & Shanteau, J. (1978). Livestock judges: How much information can an expert use? *Organizational Behavior and Human Performance, 21*, 209-219.

Raskin, D. C., & Podlesny, J. A. (1979). Truth and deception: A reply to Lykken. *Psychological Bulletin, 86,* 54-59.

Samelson, F. (1974). History, origin myth and ideology: "Discovery" of social psychology. *Journal for the Theory of Social Behavior, 4,* 217-231.

Shanteau, J. (1989). Cognitive heuristics and biases in behavioral auditing: Review, comments, and observations. *Accounting, Organizations, and Society, 14,* 165-177.

Shanteau, J. (1992a). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes, 53*, 252-266.

Shanteau, J. (1992b). How much information does an expert use? Is it relevant? *Acta Psychologica, 81*, 75-86.

Shanteau, J. (1999). Decision making by experts: The GNAHM effect. In Shanteau, J., Mellers, B., & Schum, D. (Eds.), *Decision science and technology: Reflections on the contributions of Ward Edwards*. Norwell, MA: Kluwer Academic Publishers.

Slovic, P. (1969). Analyzing the expert judge: A descriptive study of a stockbroker's decision processes. *Journal of Applied Psychology, 53*, 255-263.

Stewart, T. R., Roebber, P. J., & Bosart, L. F. (1997). The importance of the task in analyzing expert judgment. *Organizational Behavior and Human Decision Processes, 69*, 205-219.

Titchener, E. B. (1916). *A beginner's psychology.* NY: Macmillan.

Trumbo, D., Adams, C., Milner, M., & Schipper, L. (1962). Reliability and accuracy in the inspection of hard red winter wheat. *Cereal Science Today, 7*, 62-71.

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

Yates, J. F. (1990). *Judgment and decision making*. Englewood Cliffs, NJ: Prentice-Hall.

Yates, J. F., Lee, J. W., & Bush, J. G. (1997). General knowledge overconfidence: Cross-national variations, response style, and "reality." *Organizational Behavior and Human Decision Processes, 70*, 87-94.

Author Note

Preparation of this manuscript was supported, in part, by National Science Foundation Grant DMI 96-12126 and by support from the Institute for Social and Behavioral Research at Kansas State University. The author wishes to thank Ward Edwards, Julia Pounds, and Rob Ranyard for their helpful comments on earlier versions of the manuscript.

Correspondence concerning this chapter should be addressed to James Shanteau, Department of Psychology, Bluemont Hall 492, Kansas State University, Manhattan, KS 66506.

Table 1

Progression of Domains from High to Low Performance

Highest Levels of Performance.................................Lowest Levels of Performance

Aided Decisions	*Competent*	*Restricted*	*Random*
Weather Forecasters	Chess Masters	Clinical Psychologists	Polygraphers
Astronomers	Livestock Judges	Parole Officers	Managers
Test Pilots	Grain Inspectors	Psychiatrists	Stock Forecasters
Insurance Analysts	Photo Interpreters	Student Admissions	Parole Officers
Physicists	Soil Judges	Intelligence Analysts	Court Judges
Nurses	Nurses	Nurses	
Physicians	Physicians	Physicians	
Auditors	Auditors	Auditors	

Note: The space in the table separates domains (top) where the competence of experts can be classified into one category from domains (bottom) where the evidence of competence is varied. For example, various studies have reported that the behavior of auditors ranges from moderately to extremely competent.

Table 2

Reliability (Consensus) Values for Experts

Highest Levels of Performance................................Lowest Levels of Performance

Aided Decisions	*Competent*	*Restricted*	*Random*
Weather Forecasters	Livestock Judges	Clinical Psychologists	Stockbrokers
$r = .95$	$r = .50$	$r = .40$	$r = .32$
Auditors	Grain Inspectors	Pathologists	Polygraphers
$r = .76$	$r = .60$	$r = .55$	$r = .33$

Note: The values cited in this table were drawn from the following studies (from left to right): Stewart, Roebber & Bosart (1997); Phelps & Shanteau (1978); Goldberg & Werts (1966); Slovic (1969); Kida (1980); Trumbo, Adams, Milner & Schipper (1962); Einhorn (1974); and Lykken (1979).

Table 3

Reliability (Internal Consistency) Values

Highest Levels of Performance................................Lowest Levels of Performance

Aided Decisions	*Competent*	*Restricted*	*Random*
Weather Forecasters	Livestock Judges	Clinical Psychologists	Stockbrokers
$r = .98$	$r = .96$	$r = .44$	$r = <.40$
Auditors	Grain Inspectors	Pathologists	Polygraphers
$r = .90$	$r = .62$	$r = .50$	$r = .91$

Note: The values cited in this table were drawn from the following studies (from left to right): Stewart, Roebber & Bosart (1997); Phelps & Shanteau (1978); Goldberg & Werts (1966); Slovic (1969); Kida (1980); Trumbo, Adams, Milner & Schipper (1962); Einhorn (1974); and Raskin & Podlesny (1979).

Biographical Sketch

James Shanteau is Professor of Psychology at Kansas State University, where he is Director of the Institute for Social and Behavioral Research. He has held visiting positions at the National Science Foundation, as well as the Universities of Michigan, Oregon, Colorado, and Cornell. He is Co-Founder and Past President of the Society for Research on Judgment and Decision Making. He is a Fellow of the American Psychological Association and the American Psychological Society. His research interests include decision making by domain experts, models of judgment and decision making, consumer health-care choices, and quantitative models of behavior.