

From: *Decision science and technology: Reflections on the contributions of Ward Edwards*, (1999) J. Shanteau, B. A. Mellers, & D. A. Schum (Eds). Boston: Kluwer. (pp. 105-130).

DECISION MAKING BY EXPERTS: THE GNAHM EFFECT

James Shanteau
Kansas State University

Psychological studies involving experts date back to the earliest days of experimental psychology. Research on domain experts has also been a fundamental part of the history of judgment and decision making (JDM). The purpose of this chapter is to look at how domain experts have been viewed in the decision making literature. The focus will be on an unappreciated historical bias derived from a misinterpretation of the foundations of experimental psychology.

This chapter will first focus on the contributions of five key psychologists: four (*Wilhelm Wundt, James McKeen Cattell, Edward Titchener, and Edwin Boring*) are historically significant and one (*Ward Edwards*) is the focus of this book. The discussion will then turn to the impact of an overlooked assumption (abbreviated as GNAHM) that arises from this history. Next, the literature on three central questions in the JDM research on domain experts will be reinterpreted in light of this assumption. The chapter concludes with an analysis, drawn from the ideas of Wundt and Edwards, about where research on experts should go in the future.

James McKeen Cattell

The first study of the psychological processes of domain experts apparently was

conducted in 1886 by James McKeen Cattell as part of his doctoral degree under Wilhelm Wundt, the founder of experimental psychology. Cattell investigated the individual differences in “association times” (reaction times) for two skilled observers with different domain expertise. One (B) was “a teacher of mathematics” and the other (C) “busied himself more with literature” (Cattell, 1887); thus they had expertise in contrasting domains. Cattell asked both observers to answer a common set of questions, some oriented towards math, some oriented towards literature, and others oriented towards different fields altogether. The questions were designed to “call to mind facts with which (both observers) were familiar.”

The results revealed that mean reaction times over all questions were similar for the two observers: .42 sec for B and .44 sec for C. However, there was an interaction between question type and domain of expertise. The results from Cattell’s study show a 2 x 2 crossover – expertise in a domain leads to faster response times for questions in that domain. As Cattell put it, “C knows quite as well as B that $5 + 7 = 12$, yet he needs 1/10 sec longer to call it to mind. B knows quite as well as C that Dante was a poet, but needs 1/10 sec longer to think of it.” In other words, expertise has an impact on the rate of thinking above and beyond simply getting the right answer.

In another phase of his research, Cattell assessed “the time it takes to form a judgment or opinion.” Two types of tasks were used: either estimating the number of lines on a card or deciding which of two eminent men were thought to be greater. The results replotted in Figure 1b again show a 2 x 2 crossover. Expertise in a domain leads to faster judgments or choices. “In judging as to the relative greatness of eminent men, the times were shortest...if the (observer) had already compared the men together.” Cattell (1887) saw this research as just the beginning and concludes, “I think it...desirable to still further increase the number of experiments.”

Despite this promising beginning, there has been little effort to followup this line of research. Important questions about the effects of expertise on the rate of decision making remain unexplored. Other aspects of decision making by experts have also received little attention. Moreover, few theories have been advanced and few unique methodologies have been developed to study the behavior of experts. One purpose of this chapter is to explore a previously unappreciated reason for the paucity of research. To understand this reason, however, it is first necessary to look at the efforts of four major historical figures in the development of experimental psychology.¹

Wilhelm Wundt

There is little dispute that Wilhelm Wundt is the founder of experimental psychology. “Wundt is the founder because...he brought the empirical methods of physiology to the questions of philosophy and also created a new, identifiable role—that of psychologist (Leahey, 1987, p. 182). Wundt made his vision clear in his seminal book, *Principles of Physiological Psychology* (1873), noting that it was his intention to establish psychology as a new domain of science.²

Wundt argued that the goal for psychology should be: “the construction of explanations of experience and the development of techniques for objectifying experience” (Blumenthal, 1975, p. 1081). The experiences to be analyzed ranged from memory and perception to social psychology and language. To study experience, Wundt advocated the use of ‘internal perception.’ Although this methodology

has often been labeled *introspection*, “Wundt did not use the kind of qualitative introspection in which subjects describe in detail their inner experiences.... The majority of Wundt’s studies involved objective measurements that used a variety of sophisticated laboratory equipment.... Wundt then inferred information about conscious processes and elements from these objective measures” (Schultz & Schultz, 1987, p. 65-66).

The ‘subjects’ in Wundt’s laboratory were far from the naive students commonly used in research today. Instead, highly trained observers (frequently Wundt himself or his graduate students) “were presented with carefully controlled, sensory events and required to describe their mental experiences. To yield valid (observations), Wundt insisted that certain rules be enforced: the observer had to be ‘master of the situation’...All observations were to be repeated many times; and finally experimental conditions were to be varied systematically” (Hothersall, 1984, p. 88). As Wundt argued, “we learn little about our minds from casual, haphazard self-observation...It is essential that observations be made by trained observers under carefully specified conditions for the purpose of answering well-defined question.” Wundt dismissed *self observation* as “contemplative meditation that led only to fruitless debates and the grossest self-deceptions” (Wundt, 1904, p. 7). Therefore, psychological insights came from skilled observers, not naive subjects. Indeed, Wundt would likely disagree with the current use of student subjects in research.³

Wundt chose the term *voluntarismus* (voluntarism) for his school of psychology. To understand causality, Wundt used the concepts of *purpose*, *value*, and *future anticipation*. His “central mechanism of psychological causality was *apperception*, which in modern terms translates roughly to selective attention” (Blumenthal, 1980, p. 30). Attention was more than just selective; “the mind is a creative, dynamic, volitional force...It must be understood through an analysis of its activity – its processes” (Hothersall, 1984, p. 89). Moreover, “the *act of choice* constituted...a special case of ‘volitional activity’.... This is the basis on which the later forms of volition, i.e., *choice and decision*, develop” (Danzinger, 1980b, p. 96, italics added). In short, Wundt developed a dynamic process theory of choice based on the concept of will operating selectively on goals, values, and future prospects.

Wundt proposed three principles for description of psychological processes. The first was the *principle of creative synthesis* – mental constructions produce new thoughts or values that are coherent units not attributable just to external events. Wundt’s second principle, the *principle of psychic relations*, “states that the significance of any mental event is dependent upon its context” (Blumenthal, 1980, p. 31). The third principle of *psychological contrasts* concerned contrast effects and opponent processes – opposite experiences produce intensified reactions. From these principles, Wundt concluded that “voluntary activity can be understood only in terms of the ends or goals toward which they are directed” (Blumenthal, 1980, p. 31).

It should be clear that Wundt’s view of psychology is surprisingly modern. As Fancher (1979, p. 148) concluded, current psychologists will “find that their studies of such cognitive phenomena as *information processing*, *selective inattention*, and *perceptual masking* are much in the tradition of Wundt’s studies.” According to Wundt (1892, p. 495), “the fundamental character of mental life...in all its phases it is process; an *active*, not passive, existence.” Thus, it can be argued that Wundt would have been comfortable with the focus of much of modern psychological research.⁴

Wundt taught over 24,000 students and supervised 186 doctoral students, including 16 from North America. Many of the earliest psychology laboratories in the U.S.A. were founded by Wundt's students. "Wundt so influenced the first generation of American psychologists that the majority of American students of psychology (today) can probably trace their historical lineage back to him" (Boring & Boring, 1948)." For instance, Ward Edwards received his degree under Boring, who was a student of Titchener, who studied with Wundt. In other words, with Edwards you can *shake the hand that shook the hand that shook the hand that shook the hand of Wundt*.

Edward Titchener

Wundt's most vocal advocate in the United States was Edward Titchener. After spending two years with Wundt, Titchener considered himself to be a true Wundtian – he "accepted Wundt's psychology without reservation" (Hothersall, 1984 p. 103). "Titchener pronounced himself to be Wundt's loyal follower and true interpreter. He proclaimed Wundt as the source of his psychology and the precursor who validated his credentials" (Anderson, 1980, p. 95). American psychologists seldom read Wundt in German because all they needed to know was provided by Titchener. Of course, few Americans then or now could read academic German. And unfortunately, what little was available in English often contained serious translation errors. It is noteworthy that "many more of Wundt's books were translated into Russian—even into Spanish—than were ever translated into English" (Blumenthal, 1980, p. 28).

It came to be believed that Titchener's approach to psychology was essentially a mirror-image of that of his mentor Wundt – if you knew Titchener, then you knew Wundt. However, "recent research on Wundt's writings casts doubt on this conclusion. Titchener did not represent Wundt. Evidence suggests that he altered Wundt's positions to make them appear compatible with his own, to lend credibility to his own views by asserting that they were consistent with those of psychology's founder. Titchener apparently elected to translate only those portions of Wundt's publications that supported his own approach to psychology.... For 100 years, texts in the history of psychology, and teachers of the history courses, have been compounding and reinforcing the error under the imprimatur of (Titchener's) alleged expertise" (Schultz & Schultz, 1987, p. 58-59).

Titchener named his system of psychology *structuralism* because of his emphasis on "discovering the elemental structure of consciousness" (Benjamin, 1988, p. 209). "Although Wundt recognized elements or contents of consciousness, his overriding concern was their organization into higher-level cognitive processes through the principle of apperception" (Schultz & Schultz, 1987, p. 85). While Wundt emphasized the whole, Titchener focused on the parts – a view he adapted from James Mill's *Analysis of the Phenomena of the Human Mind* (Danzinger, 1980a). Wundt never described his psychology as structuralism. Indeed, after Wundt's death, his students named his movement *Ganzheit Psychology* or 'holistic psychology.'

Titchener considered subjects 'to be a recording instrument, objectively noting the characteristics of the object being observed. Human subjects were nothing more than impartial and detached machines' (Schultz & Schultz, 1987, p. 91). Titchener talked about mechanized observation – "the method of psychology is the same as

the method of physics.... strictly impartial and unprejudiced” (Titchener, 1910, p. 23). He often referred to his subjects as reagents – “a reagent is generally a passive agent, one that is applied to something to elicit certain reactions” (Schultz & Schultz, 1987).

While Wundt also using skilled observers, his subjects played a more active role in the research. One of Wundt’s laboratory rules was that highly trained observers had to be the ‘master of the situation.’ To yield valid introspections, Wundt’s subjects controlled the situation, whereas Titchener’s subjects were passive observers.⁵

Titchener’s view of the appropriate content of behavioral research was narrow and specific (Titchener, 1923). He was interested only in “pure psychology” with no room for animal psychology, mental testing, educational psychology, industrial psychology, social or ethnic psychology. These fields were considered impure because subjects in these settings could not use introspection. When discussing why he opposed animal research, Titchener (1916, p. 267) argued, “if animals thought, they could undoubtedly use their vocal organs for speech; and since they do not talk, they cannot either be thinking.” This is in contrast to Wundt’s more expansive views of psychology. “There are other source of psychological knowledge ...in which we may trace the operation of determinate psychical motives: chief among them are language, myth and custom...They are dependent upon universal psychological laws; and the phenomenon that are referable to those laws form the subject-matter of a special psychological discipline, *ethnic psychology*” (Wundt, 1904, p. 4).

In these and many other respects, Titchener presented a view of psychology that differed significantly from Wundt’s. Although Titchener made numerous contributions to the development of psychology in America, he was not an accurate spokesman for Wundt. “If one were now to take the trouble to compare the psychological system of Titchener with...Wundt, one would be able to establish that one was dealing with two quite distinct psychological systems, the one calling itself structuralism, and the other voluntarism” (Danzinger, 1980a, p. 73-74).

Edwin G. Boring

During his 33-year career, 58 students received their Ph.D. under Titchener. Some became major figures in the years following Titchener’s death in 1927. Three of these were Karl M. Dallenbach, Madison Bentley, and Margaret Floy Washburn. But Titchener’s best-known student was Edwin Boring, who was also “his most loyal student” (Benjamin, 1988, p. 210). Boring had a distinguished career at Harvard where he became “the historian” of experimental psychology.

It is through the words of Boring that most American psychologists learned about Wundt. “Titchener’s inaccurate version of Wundt’s system influenced contemporary psychologists not only because of the status Titchener achieved...but also because of the visibility attained by his student, E.G. Boring.... Thus, generations of students have been offered a portrait of Wundtian psychology that may be more myth than fact, more legend than truth” (Schultz & Schultz, 1987, p. 59).

According to Blumenthal (1975), many “myths of origin” about Wundt can traced directly to Boring’s two editions of the *History of Experimental Psychology* (1929, 1950).⁶ For nearly 70 years, Americans relied on Boring to tell Wundt’s

story. However, Boring's accounts were heavily slanted towards a position that supported Titchener's perspective. One example concerns Wundt's focus on *volition* as a motivational mechanism. Titchener (1908) largely dismissed this theme. Then, "without giving supportive citation, Boring (1950) states that Wundt had opposed the implication of an active volitional agent in psychology" (Benjamin, 1988, p. 197). Mischel (1970) has since shown that volition is, in fact, central to Wundt's approach. As another example, Titchener (1921, p. 169) held that the 10 volumes of *Völkerpsychologie* were the result of Wundt's weakness for 'troublesome subjects of a certain sort,' this was also a view reflected in Boring's accounts.

The 1950 edition of Boring's history of experimental psychology has been credited with playing a major role in establishing the legitimacy of psychology. It was published at a time that "psychology (was) hungry for scientific respectability" (Gillis & Schneider, 1966, p. 230). The argument for legitimacy came from three elements: (1) the claim of a scientific tradition dating back to Wundt that establishes the historical precedents of research psychology; (2) an emphasis on statistical analyses based on a (mis)reading of Fisher, Neyman-Pearson, and other statisticians; and (3) the institutionalization of the 'classical experimental design' involving a comparison between experimental and control groups (Gillis & Schneider, 1966). Together, they provided a myth of origin for psychology.⁷

Hebb (1972, p. 291) described the 1950 Boring history book as follows: "Commonly considered the standard work, and beautifully clear in its exposition, this book is thoroughly misleading in its emphasis...on Wundt and Titchener." It seems clear that the typical American view of Wundt is incorrect, largely because of the impact of Boring's accounts. Blumenthal (1979) concludes that Wundt was the "founding father of psychology that most psychologists have never known." It is no wonder that Max Wundt (Wundt's son) described the picture of his father's work in most psychology texts as "nothing more than a caricature" (Hothersall, 1984, p. 99).

Still, Boring can be seen as a product of his times. "Boring's description of Wundt was probably the most intelligible and most acceptable one for the great majority of American psychologists. For them, it served to crystallize a then coherent explanation of the history of psychology...that justified the course of psychology's progress" (Blumenthal, 1980, p. 40). Indeed, Boring's accounts were more accurate than many others who attempted to justify themselves by contrasting their approaches with a caricature of Wundt.

GNAHM

One central belief of Titchener concerned the *Generalized Normal Adult Human Mind* or GNAHM. According to Boring (1929), "Titchener interest lay in the generalized, normal, adult mind that had also been Wundt's main concern." Later writers reflected this theme: Wundt's "classical concept (was) that the major goal in psychology was to form generalizations about the normal human adult mind" (Helmstadter, 1964, p. 4). And "like Wundt, (Titchener) was interested in 'the generalized human mind' – not in particular human minds as such, nor in the individual differences.... Psychology may gather its material from many sources, but its aim is to understand the generalized human mind" (Heidbreder, 1933, p. 125-126). This theme is also reflected by Anastasi (1968), Kendler (1987), Maloney and Ward (1976), Sargent and Stafford (1965), and Watson (1968).

What are the origins of GNAHM? Boring (1929) cites Wundt. Edwards (1983) cites Wundt and Titchener. Others (e.g., Fancher, 1979) cite a German intellectual tradition of concern for the human mind in general. My search of a number of volumes by and about Wundt (e.g., Rieber, 1980) revealed no mention of the term. However, given that Wundt published over 491 items with a total of 53,735 published pages (Boring, 1950), it is virtually impossible to prove that he did not use the phrase. However, it is clear that Titchener referred to the concept repeatedly. "Psychology...is concerned with the normal, human, adult mind" (Titchener, 1916, p. 2). Titchener felt the focus of psychology should be on "the generalized mind, not with individual minds."

Regardless of the origins, the belief in GNAHM has become so widespread that it has become the accepted focus for psychological investigations. Moreover, few researchers today realize that there are alternatives. They fail to recognize that this unstated assumption colors nearly every aspect of their research perspective. And this has been going on for so long that it has become 'common sense.' For instance, the methods, analytic procedures, and theories used to study psychological processes are assumed to be generalizable to most, if not all, humans. In this view, undergraduates provide as good a source as any to examine behavior. Thus, GNAHM leads to research paradigms based on the investigation of psychological phenomenon derived from studies of 'typical students.'

As a result, experimental psychology has developed neither the methods nor the theories to deal with outlier behavior, such as experts. Instead, studies of expertise generally reflect GNAHM origins. The behaviors that have been found to characterize student subjects become the focus of research on experts. Rather than trying to discover what makes experts unique, most researchers rely on paradigms derived from studies of non-expert behavior. While this strategy sometimes can be revealing, it often overlooks some of the most intriguing aspects of expert behavior.⁸

As a consequence, behavioral researchers have not developed separate paradigms for investigating domain experts. Instead, theories and methods were borrowed from other areas. Because these paradigms are ill-suited to the study of experts, 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.

JDM and Experts

A survey of literature reveals that "the study of domain experts is intertwined with the development of the JDM field.... At nearly every point in the development of the field, the study of how experts make decisions and judgments has played an important role" (Shanteau & Stewart, 1992, p. 95-96). As examples, they cite research conducted using psychometric procedures (validity and reliability), linear models (multiple regression analyses), and heuristics and biases (e.g., representativeness).

Yet, the methods and conceptual approaches used to study experts were borrowed primarily from other paradigms. Few (if any) of the methods or concepts used in JDM studies were developed with experts in mind. As a consequence, expertise has been investigated using approaches intended to study typical GNAHM behavior. It should not be surprising, therefore, to find that conclusions from studies on experts tend to parallel those of non-experts.

That is not to say that paradigms originally developed within the GNAHM perspective cannot be adapted to study expert behavior. As illustrated below, it is possible to use these approaches to provide insights into expertise. The key is not the approach, *per se*. Rather, it is how the researcher uses the approach – any technique can be misused in some situations (Shanteau, 1977). The next three sections review research on experts conducted using three classical JDM approaches.

Before continuing, it is essential to comment on the definition of “expert.” I have argued elsewhere (Shanteau, 1992a) that experts can be viewed as those who have risen to the top in a domain, i.e., they are the best at what they do. But this definition is mute about which tasks lead experts to perform better or worse. In most domains experts gain their knowledge from a combination of formal training and on-the-job experience. What has yet to be determined, however, is how these two forms of knowledge combine and whether there is an interaction with tasks (Lopes, personal communication, 1997).

Psychometric Analyses

Traditional research

In the 1950’s and 60’s, systematic research on domain experts began with psychometric analyses of validity and reliability conducted on clinical psychologists. For instance, Goldberg (1959) had 22 subjects (4 practicing clinicians, 10 student interns, and 8 naive subjects) examine 30 Bender-Gestalt protocols (a test for cortical brain damage). The accuracy for all groups fell between 65 and 70%, where 50% is chance.

In an analysis of reliability, Oskamp (1962) found a correlation of .44 for repeated judgments of the same cases by clinicians. In related research, Oskamp (1965) reported that confidence of clinical psychologists increased as more information was available: the ratings for increasing amounts of information were 33, 39, 46, and 53%, respectively. However, the accuracy remained unchanged at 26, 23, 28, and 28%, where 20% is chance. Further, no difference was found between expert clinicians and students. The conclusion from such psychometric studies is that clinicians’ judgments are lacking in accuracy and consistency, that more information increases confidence but not accuracy, and that there is little difference in the judgments of clinicians and novices (Goldberg, 1968).

Similar validity and reliability results have been obtained for a variety of other expert domains, such as medical doctors (Einhorn, 1974), parole officers (Carroll & Payne, 1976), and court judges (Ebbesen & Konecni, 1975). Moreover, related research has shown that experience is often unrelated to judgment ability (Meehl, 1954) and that experts often perform little better than novices. Thus, these poor psychometric properties suggest a universal GNAHM finding; both experts and novices show the same poor performance.

GNAHM

Einhorn (1974) argued that reliability is a necessary condition for expertise. There are two ways to define reliability. The first involves *intra-judge* correlations (also known as *internal consistency* or *stability*); that is, when asked to judge the same case, an expert should make the same judgment. The other involves *inter-judge*

correlations (also known as *between-subject reliability* or *consensus*); different experts looking at the same case should arrive at the same judgment.⁹

The top three rows of Table 1 give the internal consistency results for three ‘classic studies’ of reliability. As can be seen, the correlations for repeated judgments of the same case by individual experts range between .4 and .5. Similar levels of internal consistency have been reported for other types of domain experts, such as licensed grain judges (Trumbo, Adams, Milner, & Schipper, 1962).

Table 1. Internal Consistency Reliability *r* Values for Domain Experts

<i>Domain</i>	<i>Consistency r</i>	<i>Author(s)</i>
Clinical Psychologists	.44	Oskamp (1962)
Medical Pathologists	.50	Einhorn (1974)
Physicians	<.40	Hoffman, et al (1968)
Auditors: Internal Controls	.81	Ashton (1974)
Auditors: Materiality	.90	Messier (1983)
Auditors: Audit Time	.86	Joyce (1976)
Auditors: Materiality	.83	Ettenson, et al (1987)
Livestock Judges	.96	Phelps (1977)
Weather Forecasters: Precipitation	.95	Stewart, et al (1997)
Weather Forecasters: Temperature	.99	Stewart, et al (1997)

These sorts of results led to the GNAHM conclusion that experts are not reliable. When similar analyses were conducted with expert auditors, however, the correlations were considerably better (as seen in the next four rows in Table 1). For instance, Ashton (1974) observed a mean correlation of over .8 between two presentations of 32 cases involving internal control. Researchers looking at other tasks have reported even higher results. Despite differences in task and experimenters, the findings are surprisingly consistent – between .8 and .9.

An even more impressive level of internal consistency can be seen for livestock judges. The *r* values obtained by Phelps (1977) for four top swine judges ranged from .95 to .97 for judgments of overall breeding quality. The average *r* values for other livestock judging tasks ranged from .8 to near 1.00 (perfect). Thus, livestock judges show quite high levels of intra-judge reliability.

However, Stewart, Roebber, and Bosart’s (1997) study of weather forecasters yielded the highest internal consistency levels yet recorded. Weather experts were found to have nearly perfect reliability for 24-hour temperature forecasts. Slightly lower values were obtained for 24-hour precipitation forecasts. It appears, therefore, that internal consistency for weather forecasters is near perfect.

Besides internal consistency, Einhorn (1974) also argued that experts should agree with each other. Table 2 provides results for *inter-judge* reliability (consensus). The top portion of the table shows the results from several “classic studies.” For example, Slovic (1969) and Goldberg and Werts (1966) reported average between-

judge correlation values of less than .40 for stock brokers and clinical psychologists, respectively. A somewhat higher value was observed by Einhorn (1974) for medical pathologists. Despite the impressively high internal consistency found for livestock judges, Phelps (1977) observed an average between-judge correlation of only .50. These values are similar enough to each other and to novices to support a GNAHM argument of mediocre inter-judge consensus across skill levels and domains.

However, results for expert auditors are noticeably higher. Kida (1980) asked 27 audit partners to judge 40 profiles based on five accounting ratios; the average inter-judge correlation was .76. Similar findings were reported by Ashton (1974) for judged strength of internal controls and by Libby, Artman, and Willingham (1985) for judgments of control reliance.

Inter-auditor reliability has been observed to increase as a function of experience. In a study of materiality judgments, Krogstad, Ettenson, and Shanteau (1984) compared 11 accounting students with 10 audit seniors (mid-level professionals) and 10 audit partners; the mean correlations showed a systematic increase from .66 to .76 to .83. Similar results were reported by Messier (1983) in a comparison of partners with different levels of experience.

Table 2. Consensus Reliability r Values for Domain Experts

<i>Domain</i>	<i>Consensus r</i>	<i>Author(s)</i>
Medical Pathologists	.55	Einhorn (1974)
Stockbrokers	.32	Slovic (1969)
Clinical Psychologists	<.40	Goldberg & Werts (1966)
Livestock Judges	.50	Phelps (1977)
Auditors: Accounts Ratios	.76	Kida (1980)
Auditors: Internal Controls	.70	Ashton (1974)
Auditors: Control Reliance	.68	Libby, et al (1985)
Audit Students: Materiality	.66	Ettenson, et al (1987)
Audit Seniors: Materiality	.76	Ettenson, et al (1987)
Audit Partners: Materiality	.83	Ettenson, et al (1987)
Weather Forecasters: Precipitation	.94	Stewart, et al (1997)
Weather Forecasters: Temperature	.98	Stewart, et al (1997)

As with intra-judge consistency, however, the highest recorded levels of inter-judge reliability were observed for weather forecasters. Stewart, et al (1997) reported measures of consistency that approached 1.0 for 24-hour temperature forecasts and were slightly lower for 24-hour precipitation forecasts.¹⁰

Given the high intra-judge and inter-judge correlations observed for auditors and especially for weather forecasters, it is clear that low reliability for expert judges is not a GNAHM conclusion. Some tasks reveal high levels of agreement, both within and across experts.

Linear Model Analysis

Traditional research

Use of linear regression methods for modeling judgment began in the 1950's (Hammond, 1955; Hoffman, 1960).¹¹ Much of this early research involved modeling the use of the MMPI by clinical psychologists for diagnostic purposes (Meehl, 1954). Clinicians claimed that proper use of the 11 MMPI scales required extensive training and experience, and that simple linear models (such as a sum of weighted scores) could not capture the richness of the diagnostic process.

The clinicians' claims were challenged in a series of studies showing that: (1) judgments of experienced clinicians were little better than graduate students (Oskamp, 1967); (2) statistical diagnosis based on linear regression models outperformed diagnoses made by clinicians (Goldberg, 1969); (3) diagnoses derived from regression models of the clinicians' own judgments outperformed the clinicians (Goldberg, 1970); and (4) the performance of clinicians could be matched by a model that simply added the 11 scores using equal weights (Dawes & Corrigan, 1974). This work inspired a body of research that contributed to an unflattering view of the abilities of experts.

Common sense implies that the greater knowledge of experts should be reflected in the models of their judgments. However, research has shown that linear models of expert judgments are based on relatively little information. Despite the presence of numerous relevant cues, Hoffman, Slovic, and Rorer (1968) were able to model the judgments of medical radiologists using two to six cues (also see Einhorn, 1974). Ebbesen and Konecni (1975) found that only one to three were needed to model the bail set by court judges for defendants. And Slovic (1969) was able to model stockbrokers' judgments using six to seven cues. Thus, despite the availability of additional information in each case, the models of these experts in these studies were based on surprisingly few cues. This suggests that experts make important decisions without adequate attention to all the relevant information. If so, it should not be surprising to find that expert decisions are often viewed as seriously flawed (Dawes, 1988).

In all, the evidence from traditional research revealed that linear models of experts and nonexperts alike contained a small number of significant predictor cues. "A robust finding in research on human judgment is that relatively few cues account for virtually all of the systematic variance" (Reilly & Doherty, 1989, p. 123). Since the expectation of greater information use by experts was not confirmed, the GNAHM conclusion was drawn that experts are limited decision makers.

GNAHM

It makes sense that experts should be able to access more information than non-experts. JDM researchers have interpreted this maxim to imply that greater expertise should be reflected by a larger number of cues in linear models. That is, models of experts should contain more variables than models of novices. This was labeled the *Information-Use Hypothesis* by Shanteau (1992b).

Is this GNAHM-based hypothesis correct? A cursory examination of the literature suggests mixed support. Sometimes models of experts have been found to be based on relatively large numbers of variables, e.g., analyses of livestock judges led to models with 8 to 11 significant cues (Phelps & Shanteau, 1978). In contrast,

other studies reported that agricultural experts can be described with relatively few cues (Trumbo, Adams, Milner & Schipper, 1962).

One explanation for this seeming contradiction lies in the distinction between relevant and irrelevant information. Prior research has not evaluated the degree of relevance of the cues used by experts. However, ignoring irrelevant information is exactly what an expert should do, even if it leads to fewer significant cues. Thus, the information-use hypothesis must be modified to state that experts should only be influenced by information that is the most relevant or diagnostic.

Support for this modified information-use perspective has been obtained from studies of auditors. Ettenson, Shanteau, & Krogstad (1987) compared the materiality judgments of a proposed account adjustment made by 11 audit managers and 11 accounting students. Multiple regression analyses revealed 2.7 significant cues (out of 8) for managers and 2.6 for students. Although the models contained similar numbers of cues, the pattern of cue weights was quite different. For expert auditors, most of the weight was placed on one cue with little weight given to other cues. For students, in contrast, the weighting function was flatter; similar weights were given to the various cues. Students failed to recognize that one cue was by far the most relevant. Although the number of significant cues did not differentiate between experts and novices, the ability to identify critical cues did.

In a related study, Bamber, Tubbs, Gaeth, & Ramsey (1991) compared 94 experienced auditors and 97 inexperienced auditors. Both groups were asked to review two audit cases and to revise their assessment after receiving two additional pieces of information, one relevant and one irrelevant. Both groups responded with appropriate assessment revisions after seeing relevant information. However, inexperienced auditors made an adjustment after seeing irrelevant information, whereas the judgments of experienced auditors were largely unchanged.

The evidence from studies of experts in auditing and elsewhere is consistent: the judgments of novices can often be modeled with as many (or sometimes more) cues than experts. These findings do not support the information-use hypothesis – the judgments of experts do not always seem to be based on more information than non-experts. However, experts do differ from novices in *what* information is used. What distinguishes an expert is his/her ability to discriminate what is diagnostic from what is nondiagnostic. Non-experts lack the ability to separate relevant from irrelevant information. It is the type of information used (diagnostic versus non-diagnostic), not the amount, that distinguishes experts and novices (Jacavone & Dostal, 1992, independently arrived at the same conclusion in a nursing context).

Prior decision researchers failed to recognize that relevance is *context* dependent. What is diagnostic in one context may be nondiagnostic in another. Only an experienced decision maker can determine what is relevant (and what is not) for a given situation. That is why they are considered experts and why it is so hard for others (including researchers) to understand what they are doing. Thus, the ability to evaluate the diagnosticity of information in a given context is critical to expertise.

Heuristics and Biases

Traditional research

In the early 1970's, Tversky and Kahneman developed a research orientation that has dominated the *JDM* literature for over 25 years. They argued that humans make use of heuristics rules as mental shortcuts to reduce the cognitive complexity of

making probabilistic judgments. "In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors (biases)" (Tversky & Kahneman, 1974). To demonstrate the application of heuristics, such as *representativeness*, they developed a variety of empirical demonstrations that produced behavior that deviated from normative or correct answers. For instance, they found that student subjects often *ignore base rates* or prior information in making judgments.

The method commonly used to study heuristics and biases involves having undergraduates answer statistics questions presented as word problems. In the well-known *cab problem*, for example, verbal descriptions are provided of prior odds and the likelihood ratio (Kahneman & Tversky, 1972). The correct answer can be computed from an application of Bayes Theorem. Estimates of the posterior odds reveal that subjects behaved as if they *ignored base rates*.

Similar findings have been observed for a variety of other word problems. For instance, consider the *lawyer-engineer* problem: "Dick is a 30 year old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues." The description was designed to be uninformative about Dick's occupation. When asked whether Dick was more likely to be an engineer or a lawyer, most students judged the probability to be .5, even when they were told the overall proportion of engineers was .7 or .3. In other words, students ignored the base rate.

In support of their approach, Tversky and Kahneman (1974) argue that heuristics and biases are *universal cognitive illusions*. "There is much evidence that experts are not immune to the cognitive illusions that affect other people (Kahneman, 1991)." Similarly, Tversky (quoted in Gardner, 1985) 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." Thus, Kahneman and Tversky drew the GNAHM conclusion that biased behavior from the use of heuristics applies to all.¹²

GNAHM

Is the GNAHM argument that experts ignore base rates (just like students) correct? This question has been studied extensively in the auditing literature. Using a variation of the lawyer-engineer problem, for instance, Joyce and Biddle (1981) reported that 132 auditors under-utilized base rate information in judging management fraud. They found, however, that professional auditors did better than students. In a followup study, however, Holt (1987) concluded that it was the wording of the problem, rather than the use or misuse of base rates, that led to the Joyce and Biddle conclusion.

To see if task familiarity influenced reliance on base rates, Kida (1984) evaluated the use of data by 73 audit partners and managers. The results revealed that "most auditors attended to both base rates and indicant (case specific) data" (Smith & Kida, 1991, p. 475). Thus, "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" (Smith & Kida, 1991, p. 480).

To examine the use of heuristics by experts more broadly, I reviewed 20 studies of audit judgment (Shanteau, 1989). These studies were intended to examine one or more of the heuristics and biases proposed by Tversky and Kahneman. My

conclusion was that the results reported were often surprisingly close to normative. Consider the following examples: Bamber (1983) reported that auditor managers were not only sensitive to the reliability of information, they often over compensated. Kinney and Uecker (1982) observed results contrary to the anchoring-and-adjustment heuristic in two experiments. Biddle and Joyce (1982) failed to find the effects predicted by the availability heuristics. Gibbins (1977) reported that 40% of auditors' responses followed representativeness, about half made the normative response, and the remainder were inconsistent with either. Comparable findings were observed by Waller and Felix (1987) and Abdolmohammadi and Wright (1982).

Despite these results, many audit researchers nonetheless emphasized the presence of heuristics and biases. When Biddle and Joyce (1982) failed to support anchoring and adjustment, they concluded that "some as yet unidentified heuristic was at work" (p. 189). Similarly, Ashton (1983, p. 35) states, "despite the mixed nature of the overall results in the heuristics area, findings such as these suggest that auditors often have difficulty understanding the implications of sample information." The possibility that the heuristics approach may be inappropriate apparently wasn't considered.

These cases reflect a tendency in auditing (and other areas) to define success of a study by whether biases are observed or not. This "bias" toward emphasizing poor decision behavior also has been reported in the psychological literature (Christensen-Szalanski & Beach, 1984). My conclusion (Shanteau, 1989, p. 170) was that "heuristics and biases research...has been of limited relevance for auditing research."

Independent of my review, Smith and Kida (1991) examined 25 studies of heuristics and biases in the accounting and auditing literature. They systematically considered research on each of the various heuristics proposed by Kahneman and Tversky. With anchoring and adjustment, for instance, they concluded that auditors show considerable bias when presented with unfamiliar, student-like problems. "However, when the experimental tasks were more analogous to typical audit judgments, and hence more familiar to the (expert) subjects, anchoring results were often not found or were significantly mitigated" (Smith & Kida, 1991, p. 477).

Looking at auditing studies of representativeness, these authors report: "In experiments asking (accounting) students to perform abstract and unfamiliar tasks, the results more strongly support the earlier findings in the heuristics literature, whereas in experiments asking experienced auditors to complete familiar, job-related tasks, the results...correspond more closely to normative principles than the behaviors of naive student subjects" (p. 480).

After their review of these and other heuristics, Smith and Kida concluded the "biases found readily in other research are not evident in the judgements of professional auditors" (p. 485). Based on the reviews conducted by both Smith and Kida (1991) and Shanteau (1989), it seems clear that the effects observed for students in studies of heuristics and biases do not necessarily generalize to auditing experts.

It is worth noting that similar conclusions have been drawn in other domains. For instance, Schwartz and Griffin (1986, p. 82) concluded after a review of the literature in medical decision making that use of heuristics in medicine is limited. "Decision heuristics...appear more likely to create biases in the psychology laboratory than in the (medical) clinic." Other investigators have arrived at similar

conclusions (Anderson, 1987; Cohen, 1981; Edwards, 1983; Edwards & vonWinterfeldt, 1986; Gigerenzer & Hoffrage, 1995; Jungermann, 1983; Shanteau, 1978; Wallsten, 1983; and Wright, 1984). As Smith and Kida (1991, p. 486) conclude, “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.... Accordingly, findings from heuristic studies that use students performing generic tasks of limited consequence do not necessarily generalize well to this expert decision making context.”

It is important to note, however, that not all experts are immune to the use of heuristics. Although quite knowledgeable about their fields, many domain experts seem unable to translate that knowledge into appropriate decisions. Judgments of probability present a particular difficulty for many experts, just as for naive subjects. For instance, Slovic and Monahan (1995) found that both expert clinicians and naive participants were inappropriately influenced by scale format in making assessments of probability of danger from mental patients. Thus despite their superior knowledge and experience, clinical psychologists show the same biases as unskilled subjects.

Obviously, more research is needed to determine the conditions under which experts may be expected to perform well. One possibility (Slovic, personal communication, 1997) is that experts do well in repetitive tasks where there is high quality feedback – such as weather forecasting and auditing. In contrast, experts do poorly when probability forecasts are required with low fidelity feedback – such as clinical psychology and legal judgment. There is clearly a need to develop an assessment of task characteristics to make some sense out of such differences.

Conclusions

What direction should JDM research on expertise take to resolve the issues raised in this chapter? For answers, we will turn first to directions laid down a century ago by Wundt. Then we will consider a more modern set of ideas from Ward Edwards.

It is worth recalling that Wundt stressed the role of *skilled participants* in behavioral research. He had no interest in studying the behavior of naive subjects. While this might be considered an extreme position today, it has merit. If we want to gain maximum insight into an issue, we should work with those who have the best perspective on the issue. For most problems in decision making, that implies conducting research with individuals who have direct experience, which often means experts. Thus, Wundt’s views imply that JDM researchers should focus on the behavior of skilled decision makers.

Wundt’s methodology involved objective measures of internal perception under controlled conditions involving sophisticated laboratory equipment. This approach still makes sense today. Our goal should be the same as Wundt’s – to make inferences about conscious processes from objective measurements. Why not use reaction time measures, as Cattell did in 1887, to study comparative judgments of domain experts? In particular, the interaction involving domains shown in Figure 1 should be examined further.

Wundt’s emphasis on *apperception* or selective attention also seems appropriate to understanding how (and when) experts achieve superior performance. Recently, many investigators have concluded that the superiority of domain experts depends on their ability to focus on relevant information (Benner, Tanner, & Chesla, 1996;

Ettenson, Shanteau, & Krogstad, 1987; Jacavone & Dostal, 1992; Mosier, 1997; Phelps & Shanteau, 1978; Schwartz & Griffin, 1986; Shanteau, 1992b). The goal in future research should be to learn how experts make these discriminations and to find ways to enhance the process.

As with so much in the judgment and decision making research, Ward Edwards was the first to note the significance of the concerns expressed here. In 1983, he saw the danger to JDM research by over reliance on GNAHM thinking. According to Edwards, investigators operated as if “any individual’s mind was as much a representative of the ‘generalized normal adult human’ as any other” (p. 507). Thus, “studies of unmotivated subjects and use of non-expert subjects are enshrined in psychological tradition.... I have been in disagreement with this line of research and thought for some time, and now I regret my own role in starting it off in the early 1960’s.... it is time to call a halt” (p. 508).

He 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 generalization is made” (p. 509). This parallels Wundt’s *principle of psychic relations* – the significance of any mental event is dependent upon its context. In short, decisions and choices must be studied in context.

Edwards’ second message is that sometimes “experts can in fact do a remarkably good job of assessing and working with probabilities. Two groups of studies seem to show this. One is concerned with weather forecasters.... A second group of studies seems to say that physicians also work well with probabilistic reasoning tasks of kinds familiar to them” (p. 511).

Edwards drew three conclusions that summarize the arguments offered here. “One is that, as a practical matter, the rejection of human capacity to perform probabilistic tasks is extremely premature.... Obviously, the experimenters themselves, using tools and expertise, were able to perform such tasks rather well.... A second conclusion is that the ‘generalized normal adult human mind’ is simply the wrong target for research on human intellectual performance. We must recognize that minds vary, that tools can help, that expertise can help” (p. 511). His final conclusion is “that 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 exactly what kinds of people and tasks deserve our attention” (p. 512). These arguments remain as challenges for JDM researchers today and in the future.

Acknowledgment to Ward Edwards

The work described in this chapter was stimulated by an original suggestion from Ward Edwards in 1992. Moreover, he provided valuable input at every step as these ideas developed since. In a very real sense, therefore, this project is a direct result of Ward’s efforts to advance understanding of a concern basic to psychology and decision research.

However, my interaction with Ward predates this project by many years. Our first contact was in 1970 when I was a postdoc at the University of Michigan. Shortly after arriving in Ann Arbor, Ward invited me to give a lunchbag presentation about a study I had conducted on probability revision. The results showed that, contrary to the ordinary version of Bayes theorem, subjects were

averaging present and prior probabilities. Ward was the first in the room to grasp the meaning of the results. But then he offered an alternative version of Bayes that was consistent with averaging.

This example illustrates three of Ward's most valuable traits. First, he has consistently provided enthusiastic encouragement to young researchers in judgment and decision making. Second, he understands and promotes new ideas and research directions. And third, he offers insightful reinterpretations that force you to think more deeply. For these and many other reasons, it is my honor to dedicate this chapter to Ward Edwards, the founder of behavioral decision research.

Author Notes

This research was supported, in part, by grant 96-12126 from the *National Science Foundation* and by support from the *Institute for Social and Behavioral Research* at Kansas State University. I am grateful to Paul Slovic and Lola Lopes for their helpful comments on the manuscript. I am also indebted to Gerd Gigerenzer and Franz Samelson for their insights into the misconceptions of Wilhelm Wundt held by most psychologists. Finally, I wish to acknowledge the continuing inspiration and support offered by Ward Edwards throughout all stages of this project.

Correspondence concerning this research should be addressed to James Shanteau, Institute for Social and Behavioral Research, Ahearn 202, Kansas State University, Manhattan, KS 66506-0307.

Notes

¹ Cattell plays a curious role in this history. After completing his studies with Wundt, he worked for a time in England with Francis Galton (Charles Darwin's half cousin), the originator of research on individual ability. It is sometimes assumed that Cattell was originally inclined to follow Wundt in the study of GNAHM behavior, but that Galton somehow convinced him to investigate individual differences instead. The reality, as the opening section shows, was that Cattell (1885) was already interested in individual differences at the time he was a student in Germany (Sokal, 1981).

It is also noteworthy that a number of Wundt's other students went on to have productive careers in fields as varied as psychopathology, classical conditioning, perception, child psychology, physiology, applied psychology, psycholinguistics, comparative psychology, statistics, education, and social psychology. These are all fields divergent from perception and sensation. As Murphy (1929, p. 173) stated, "Wundt's students began even while still with him the study of problems which were both envisaged and prosecuted with originality and relative independence." Thus, Cattell's lifelong focus on individual differences was well within the tradition of Wundt's students pursuing their own objectives.

² In the mid 19th century, the adjective "physiologis chen" (*physiological*) came to mean "experimental." Thus, there was discussion of "physiological pedagogy," "physiological aesthetics," and "physiological psychology" (Blumenthal, 1980). Wundt was inspired by the research methods of physiologists, such as Müller and Helmholtz. However, "Wundt saw psychology as a separate science in which it was

not necessary to refer to physiological processes.... Hence, his psychology was not physiological psychology in the current sense of that term” (Watson, 1979, p. 131).

³ As Gigerenzer notes in a personal communication (1994), the “subjects” in Wundt’s research were Ph.D.s or professors. “The longstanding Wundtian tradition (was) that experimental subjects had to be experts.” So an argument can be made that the first studies using experts were in fact conducted by Wundt. However, Cattell appears to have been the first to conduct research that examined the content or domain knowledge of experts. In contrast, Wundt asked his subjects to use their experience to gain further insights into basic psychological processes, such as perception.

⁴ Wundt’s long academic career can be divided into three parts. In the first part, he studied attention, learning, memory, reaction time, and mental associations in addition to perception. “Out of this work came what is often considered the most important book in the history of psychology, *Grundzüge der Physiologischen Psychologie (Principles of Physiological Psychology)*, published in two parts in 1873 and 1874” (Schultz, 1969, p. 45). In his middle career during the 1880’s, Wundt turned his efforts to philosophical psychology, with books entitled *Logic*, *Ethic*, and *System of Philosophy*. “The last of Wundt’s three lifetime projects was to create social psychology” (Miller, 1962, p. 23). He did this through 10 volumes on *Völkerpsychologie (Ethnic Psychology)* covering the study of languages, art, myths, social customs, law, morals, and culture. Thus, Wundt’s actual career was far broader than the narrow stereotypic view typically taught to American psychology students.

⁵ Titchener was responsible for other ‘myths of origin’ concerning Wundt. Titchener advocated a methodology based on introspective techniques under carefully controlled conditions. “To ensure his students’ accuracy in describing their consciousness experiences, Titchener drilled them in what he called ‘hard introspective labor.’ Certain introspections were defined as correct, and certain others as in error, with the final authority being Titchener himself.” (Hothersall, 1984, p. 105). Titchener considered most ordinary untrained (or common sense) observations “worthless;” they were “usually inaccurate” and almost involved what he called the *stimulus error* – confusing the mental process with the object being observed.

Although Titchener attributed his methodology to Wundt, “Titchener’s form of introspection was, however, quite different from Wundt’s.... Titchener used detailed, qualitative, subjective reports of the subjects’ mental activities during the act of introspection” (Schultz & Schultz, 1987, p. 90). This contrasts with Wundt’s approach based on laboratory instruments and his focus on objective quantitative measurements, such as reaction time. Wundt’s label for his technique is *selbst-beobachtung* or “self-observation.” This term is usually translated as “introspection,” an unfortunate choice because it implies “a type of armchair speculation, which was certainly not what Wundt meant” (Hothersall, 1984, p. 88). He dismissed such introspection as “contemplative mediation” that led only to pointless debate and the grossest self-deception (Wundt, 1904, p. 7).

⁶ This reflects a trend noted by the psychohistorian Samelson (1974) for social scientists to create ‘myths of origins’ to justify present positions as being the course of history. Obviously, psychology was not immune to the development of, and belief in, these myths.

⁷ See Gigerenzer (1993) for a discussion of how psychologists misinterpreted the arguments of statisticians. This ultimately led to the current reification of null hypothesis significance testing in psychology.

⁸ As pointed out by Cronbach (1960) in his essay, ‘The two disciplines of scientific psychology,’ a schism separates two branches of behavioral research. On the one hand, laboratory researchers focus on controlled experimentation looking for global effects – the ‘Wundt tradition.’ On the other hand, individual difference researchers examine the variation of behavior within designated groups – the ‘Galton tradition.’ Investigators from both perspectives view the breach as having historical justification. The following excerpt from Kelly (1967, p. 87-88) examines the schism:

“The early experimental psychologists were primarily concerned with discovery of general laws to describe the exact relationships between variables.... In their search for general laws, they were looking for relationships that are true for all persons. To the extent that individual differences among subjects showed up in the results of an experiment, they were regarded as errors rather than phenomena of intrinsic interest. Even today, the most widely used statistical technique – analysis of variance – for inferring that a true difference exists between groups of subjects tests the differences of means and overlooks marked individual differences among the members of each group.”

“By contrast, following the lead of Galton in England and James McKeen Cattell in the United States, another group of psychologists became intrigued with the range of variety of individual differences, the origins of such differences, and the relation among these differences. Unfortunately, there was all too little communication between the two groups of psychologists. Each tended to develop its own theories, methods, and body of knowledge.”

Both approaches, however, accept the basic GNAHM argument. Experimental researchers focus on average results; outlier behavior is generally excluded from the group averages. Individual difference researchers, while sensitive to subject variation, focus on distributions of scores; normal distributions are assumed with an emphasis on the overall distribution. In their own ways, both experimental and individual difference researchers are concerned about the characteristics of groups of subjects. In this partition, there is no clear home for research on experts.

⁹ The question of validity (accuracy) will not be pursued because correct answers seldom, if ever, exist in the domains under consideration. The concept of a *gold standard* or other absolute point of reference does not apply to tasks performed by experts; in fact, we often need to rely on experts precisely because we do not have correct answers (Shanteau, 1992a; Gigerenzer & Goldstein, 1996).

¹⁰ Theoretically, the inter-judge correlations should not be higher than the intra-

judge correlations. Due to the rounding of average results, however, some correlations in Table 2 are slightly higher than comparable values in Table 1.

¹¹ In what appears the first regression analysis of expert judgment, Wallace (1923) reanalyzed the data that Hughes (1917) collected on corn judges. Using a precursor of path analysis, Wallace found that these experts relied primarily on ear length; however, kernel weight was most predictive of yields. Despite its historical precedence, this application of linear model analysis to the study of experts was not followed up.

¹² It should be pointed out that base rates or prior information are vitally important for expertise. For experts to ignore base rates would be a most serious error. To concentrate only on case-specific information is equivalent to making decisions without using past knowledge or experience. This contradicts one of the most fundamental characteristics of experts. Use of prior experience is a defining ingredient of expertise. Indeed, the words “expert” and “experience” have the same Latin root. Thus, when an expert ignores base rates, there is good reason to question his/her competence.

References

- Abdolmohammadi, M., & Wright, A. (1987). An examination of the effects of experience and task complexity on audit judgments. *Accounting Review*, 1-13.
- Anastasi, A. (1968). *Psychological testing* (3rd Ed.). NY: Macmillan Publishers Co.
- Anderson, N. H. (1987). A cognitive theory of judgment and decision. In, Brehmer, et al. (Eds.), *SPUDM proceedings*. Amsterdam: North-Holland Press.
- Anderson, R. J. (1980). Wundt's prominence and popularity in his later years. *Psychological Research*, 42, 87-101.
- Ashton, R. H. (1974). An experimental study of internal control judgments. *Journal of Accounting Research*, 12, 143-157.
- Ashton, R. H. (1983). *Research in audit decision making: Rationale, evidence, and implications*. Vancouver: Canadian Certified General Accountants' Monograph 6.
- Bamber, E. M. (1983). Expert judgment in the audit team: A source reliability approach. *Journal of Accounting Research*, 21, 396-412.
- Bamber, E. M., Tubbs, R. M., Gaeth, G., & Ramsey, R. J. (1991). Characteristics of audit experience in belief revision. In, *USC audit judgment symposium*. Los Angeles: University of Southern California.
- Benjamin, L. T., Jr. (1988). *A history of psychology: Original sources and contemporary research*. NY: McGraw Hill.
- Benner, P., Tanner, C. A., & Chesla, C. A. (1996). *Expertise in nursing practice: Caring, clinical judgment, and ethics*. NY: Springer.
- Biddle, G. C., & Joyce, E. J. (1982). Heuristics and biases: Some implications for probabilistic inference in auditing. In, *Symposium on audit research IV*. Urbana: University of Illinois.
- Blumenthal, A. L. (1975). A reappraisal of Wilhelm Wundt. *American Psychologist*, 30, 1081-1086.
- Blumenthal, A. L. (1979). The founding father we never knew. *Contemporary Psychology*, 24, 449-453.
- Blumenthal, A. L. (1980). Wilhelm Wundt and early American psychology: A clash of cultures. In R. W. Rieber (Ed.), *Wilhelm Wundt and the making of a scientific psychology*. NY: Plenum Press.
- Boring, E. G. (1929). *A history of experimental psychology*. NY: Appleton & Co.
- Boring, E. G. (1950). *A history of experimental psychology* (2nd ed). NY: Appleton.
- Boring, M. D., & Boring, E. G. (1948). Masters and pupils among the American psychologists. *American Journal of Psychology*, 61, 527-534.
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 62, 193-217.
- Carroll, J. S., & Payne, J. W. (1976). The psychology of parole decision processes: A joint application of attribution theory and information-processing psychology. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social psychology*. Hillsdale, NJ: Erlbaum.
- Cattell, J. M. (1885). The inertia of the eye and the brain. *Brain*, 8, 295-312.
- Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, XI, 63-65.
- Cattell, J. M. (1887). Experiments on the association of ideas. *Mind*, XII, 68-74.
- Christensen-Szalanski, J. J. J., & Beach, L. R. (1984). The citation bias: Fad and fashion in the judgment and decision making literature. *American Psychologist*, 75-78.

- Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavior and Brain Sciences*, 317-331.
- Cronbach, L. J. (1960). The two disciplines of scientific psychology. *American Psychologist*, 12, 671-684.
- Danziger, K. (1980a). Wundt and the two traditions of psychology. In R. W. Rieber (Ed.), *Wilhelm Wundt and the making of a scientific psychology*. NY: Plenum.
- Danziger, K. (1980b). Wundt's theory of behavior and volition. In R. W. Rieber (Ed.), *Wilhelm Wundt and the making of a scientific psychology*. NY: Plenum Press.
- Dawes, R. M. (1988). *Rational choice in an uncertain world*. San Diego, CA: Harcourt Brace Jovanovich.
- Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81, 95-106.
- 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.
- Edwards, W. (1983). Human cognitive capacities, representativeness, and ground rules for research. In P. Humphreys, O. Svenson, & A. Vari (Eds.), *Analyzing and aiding decision processes*. Budapest: Akademiai Kiado.
- Edwards, W., & von Winterfeldt, D. (1986). On cognitive illusions and their implications. *Southern California Law Review*, 59(2), 401-451.
- Einhorn, H. (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.
- Fancher, R. E. (1979). *Pioneers of psychology*. NY: Norton.
- Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. NY: Basic Books.
- Gibbins, M. (1977). Human inference, heuristics, and auditors' judgment processes. In *CICA Auditing Research Symposium*. Toronto: CICA.
- Gigerenzer, G. (1993). The superego, the ego, and the id in statistical reasoning. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues*. Hillsdale, NJ: Erlbaum.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650-669.
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102, 684-704.
- Gillis, J., & Schneider, C. (1966). The historical preconditions of representative design. In K. R. Hammond (Ed.), *The psychology of Egon Brunswik*. NY: Holt.
- Goldberg, L. R. (1959). The effectiveness of clinician's judgments: The diagnosis of organic brain damage from the Bender-Gestalt Test. *Journal of Consulting Psychology*, 23, 25-33.
- Goldberg, L. R. (1968). Simple models or simples processes? Some research on clinical judgments. *American Psychologist*, 23(7), 483-496.
- Goldberg, L. R. (1969). The search for configural relationships in personality assessment: The diagnosis of psychosis vs. neurosis from the MMPI. *Multivariate Behavioral Research*, 4, 523-536.
- Goldberg, L. R. (1970). Man vs. model of man: A rationale, plus some evidence, for a method of improving clinical inferences. *Psychological Bulletin*, 73, 422-432.

- Goldberg, L. R., & Werts, C. E. (1966). The reliability of clinicians' judgments: A multitrait-multimethod approach. *Journal of Consulting Psychology, 30*, 199-206.
- Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review, 62*, 255-262.
- Hebb, D. O. (1972). *Textbook of psychology* (3rd Ed.). Philadelphia: W. B. Saunders.
- Heidbreder, E. (1933). *Seven psychologies*. NY: Appleton-Century Co.
- Helmstadter, G. C. (1964). *Principles of psychological measurement*. NY: Appleton.
- Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin, 62*, 255-262.
- 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.
- Holt, D. L. (1987). Auditors and base rates revisited. *Accounting, Organizations, and Society, 12*, 571-578.
- Hothersall, D. (1984). *History of psychology*. Philadelphia: Temple University Press.
- Hughes, H. D. (1917). An interesting corn seed experiment. *Iowa Agriculturalist, 17*, 424-425.
- 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.
- Joyce, E. J. (1976). Expert judgment in audit program planning. *Journal of Accounting Research, 14*, 29-60.
- Joyce, E. J., & Biddle, G. C. (1981). Are auditors' judgments sufficiently regressive? *Journal of Accounting Research, 19*, 329-349.
- Jungermann, H. (1983). The two camps of rationality. In R. W. Scholz (Ed.), *Decision making under uncertainty*. Amsterdam: Elsevier.
- Kahneman, D. (1991). Judgment and decision making: A personal view. *Psychological Science, 2*, 142-145.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430-454.
- Kelly, E. L. (1967). *Assessment of human characteristics*. Belmont, CA: Brooks/Cole.
- Kendler, H. H. (1987). *Historical foundations of modern psychology*. Philadelphia: Temple University Press.
- Kida, T. (1980). An investigation into auditors' continuity and related qualification judgments. *Journal of Accounting Research, 8*, 506-523.
- Kida, T. (1984). The effect of causality and specificity on data use. *Journal of Accounting Research, 22*, 145-152.
- Kinney, W. R., & Uecker, W. (1982). Mitigating the consequences of anchoring in auditor judgments. *The Accounting Review, 57*, 55-69.
- Krogstad, J. L., Ettenson, R. T., & Shanteau, J. (1984). Context and experience in auditors' materiality judgments. *Auditing: A Journal of Practice & Theory, 4*, 54-73.
- Leahey, T. H. (1987). *A history of psychology: Main currents in psychological thought* (2nd ed). Englewood Cliffs, NJ: Prentice-Hall.

- Libby, R., Artman, J. T., & Willingham, J. J. (1985). Process susceptibility, control risk, and audit planning. *The Accounting Review*, 60, 212-230.
- Maloney, M. P., & Ward, M. P. (1976). *Psychological assessment: A conceptual approach*. NY: Oxford University Press.
- Meehl, P. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- Messier, W. F. (1983). The effect of experience and firm type on materiality/disclosure judgments. *Journal of Accounting Research*, 21, 611-618.
- Miller, G. A. (1962). *Psychology: The science of mental life*. NY: Harper & Row.
- Mischel, T. (1970). Wundt and the conceptual foundations of psychology. *Philosophical and Phenomenological Research*, 31, 1-26.
- Mosier, K. L. (1997). Myths of expert decision making and automated decision aids. In C. Zsombok & G. Klein (Eds.), *Naturalistic decision making*. NJ: Erlbaum.
- Murphy, G. (1929). *An historical introduction to modern psychology*. NY: Harcourt, Brace & Co.
- Oskamp, S. (1962). The relationship of clinical experience and training methods to several criteria of clinical prediction. *Psychological Monograph*, 76.
- Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology*, 29, 261-265.
- Oskamp, S. (1967). Clinical judgment from the MMPI: Simple or complex? *Journal of Clinical Psychology*, 23, 411-415.
- Phelps, R. H. (1977). *Expert livestock judgment: A descriptive analysis of the development of expertise*. Unpublished dissertation. Manhattan: Kansas State Univ.
- Phelps, R. H., & Shanteau, J. (1978). Livestock judges: How much information can an expert use? *Organizational Behavior and Human Performance*, 21, 209-219.
- Reilly, B. A., & Doherty, M. E. (1989). A note on the assessment of self-insight in judgment research. *Organizational Behavior & Human Performance*, 21, 209-219.
- Rieber, R. W. (1980). *Wilhelm Wundt and the making of a scientific psychology*. NY: Plenum.
- Samelson, F. (1974). History, origin myth and ideology: "Discovery" of social psychology. *Journal for the Theory of Social Behavior*, 4, 217-231.
- Sargent, S. S., & Stafford, K. R. (1965). *Basic teachings of the great psychologists* (Rev Ed.). Garden City, NY: Dolphin Books.
- Schwartz, S. & Griffin, T. (1986). *Medical thinking: The psychology of medical judgment and decision making*. NY: Springer-Verlag.
- Schultz, D. P. (1969). *A history of modern psychology*. San Diego: Harcourt.
- Schultz, D. P., & Schultz, S. E. (1987). *A history of modern psychology* (4th ed). San Diego: Harcourt Brace Jovanovich.
- Shanteau, J. (1977). Correlation as a deceiving measure of fit. *Bulletin of the Psychonomic Society*, 10, 134-136.
- Shanteau, J. (1978). When does a response error become a judgmental bias? *Journal of Experimental Psychology: Human Learning and Memory*, 4, 579-581.
- 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., & Stewart, T. R. (1992). Why study expert decision making? Some historical perspectives and comments. *Organizational Behavior and Human Decision Processes*, 53, 95-106.
- Slovic, P. (1969). Analyzing the expert judge: A descriptive study of a stockbroker's decision processes. *Journal of Applied Psychology*, 53, 255-263.
- Slovic, P., & Monahan, J. (1995). Probability, danger, and coercion: A study of risk perception and decision making in mental health law. *Law and Human Behavior*, 19, 49-65.
- Smith, J. F., & Kida, T. (1991). Heuristics and biases: Expertise and task realism in auditing. *Psychological Bulletin*, 109, 472-489.
- Sokal, M. M. (1981). *An education in psychology: James McKeen Cattell's journal and letters from Germany and England, 1880-1888*. Cambridge, MA: MIT Press.
- 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. (1908). *The psychology of feeling and attention*. NY: Macmillan.
- Titchener, E. B. (1910). *A textbook of psychology*. NY: Macmillan.
- Titchener, E. B. (1916). *A beginner's psychology*. NY: Macmillan.
- Titchener, E. B. (1921). Wilhelm Wundt. *Am. Journal of Psychology*, 32, 161-178.
- Titchener, E. B. (1923). *An outline of 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.
- Wallace, H. A. (1923). What is in the corn judge's mind? *Journal of the American Society of Agronomy*, 15, 300-324.
- Waller, W. S., & Felix, W. L., Jr. (1987). Auditors' covariation judgment. *The Accounting Review*, 62, 275-292.
- Wallsten, T. S. (1983). The theoretical status of judgmental heuristics. In R. W. Scholz (Ed.), *Decision making under uncertainty*. Amsterdam: North-Holland.
- Watson, R. I. (1968). *The great psychologists: From Aristotle to Freud* (2nd Ed.). Philadelphia: J. B. Lippincott Co.
- Watson, R. I. (1979). *Basic writings in the history of psychology*. NY: Oxford.
- Wright, G. (1984). *Behavioral decision theory: An introduction*. Beverly Hills: Sage.
- Wundt, W. (1873). *Grundzüge der physiologischen psychologie*. Leipzig: Engelmann.
- Wundt, W. (1892). *Vorlesungen über die menschen und theirseele*. (Trans. Creighton & Titchener). *Lectures on human and animal psychology*. NY: Macmillan.)
- Wundt, W. (1904). *Principles of physiological psychology* (4th ed) (trans. Titchener). NY: Macmillan.