AI Magazine Volume 8 Number 3 (1987) (© AAAI)

How Humans Process Uncertain Knowledge:

An Introduction for Knowledge Engineers

Robert F. Hink and David L. Woods

The question of how humans process uncertain information is important to the development of knowledge-based systems in terms of both knowledge acquisition and knowledge representation This article reviews three bodies of psychological research that address this question: human perception, human probabilistic and statistical judgment, and human choice behavior The general conclusion is that human behavior under uncertainty is often suboptimal and sometimes even fallacious. Suggestions for knowledge engineers in detecting and obviating such errors are discussed The requirements for a system designed to reduce the effects of human factors in the processing of uncertain knowledge are introduced.

What a chimera then is man! . Judge of all things . depository of truth, a sink of uncertainty and error, the glory and shame of the universe

—*Blaise Pascal,* Lettres Provinciales *(1656-1657)*

Currently, a vigorous debate is in progress within the AI community concerning how best to represent and process uncertain knowledge in knowledge-based systems This debate carries great importance because most human decisions are made under conditions of uncertainty.

Psychological research has revealed that human performance in the face of uncertainty is spotty at best. Humans display suboptimal choice strategies, miscalibrations in assessing probabilities, fallacious statistical inference, and inconsistencies in their preferences for uncertain outcomes. Moreover, both novices and experts are subject to these kinds of inaccuracies and errors.

This poor report card should be particularly distressing to knowledge engineers (KEs) who are confronted with the dilemma that no matter how uncertain knowledge is represented in an expert system, it is suspect if acquired from a human, even a human expert. Those who are trying to automate knowledge acquisition by building intelligent interfaces to knowledge engineering tools cannot be comforted by this news. Their interfaces would have to contain sophisticated and as yet unspecified metaknowledge about these particular human frailties in order to overcome the problem.

On the positive side, these weaknesses in human judgment and reasoning also present a challenge and an opportunity for knowledge-based systems If the systems could compensate for human error in handling uncertainty, superexpert performance might be achieved.

This article attempts to review what is known about how humans handle uncertainty. The primary objective is to give KEs a basis from which they can evaluate the accuracy, consistency, and correctness of the domain expert's (DE's) problem-solving performance when faced with uncertainty. A secondary objective is to present possible knowledge-acquisition techniques that might improve communication between the KE and DE regarding problem solving with uncertain knowledge. Finally, we present the requirements for an automated system designed to improve problem solving with uncertain knowledge.

Several caveats need to be addressed before proceeding further. First, we make no claim that this review is complete. As we stated earlier, the research in this area is extensive and includes at least three large areas of inquiry. One area deals with human perceptual capacity and performance. Another area focuses on comparisons of human judgment of probabilities and statistics with normative models. The third research area involves human choice behavior under uncertainty. Another caveat is that as with most areas of active inquiry, this field has its controversies. Because our primary aim is to inform knowledge engineers about what they can expect, rather than to sort out theoretical issues, we avoid discussion of these theoretical controversies.

The Findings

The studies reviewed in this section highlight the significant psychological research concerning how humans process uncertain information from the perspectives of perception, judgment, and choice behavior.

Man against the Bits: Perception

In 1948, C. E. Shannon (1948) published an article that gave birth to information theory. Shannon defined the amount of information carried by a signal in terms of probabilities. Precisely, he defined the amount of information, I (in bits), a signal carries as

$$I(x) = \mathrm{Log}_2\, P_2(x) / P_1(x),$$

where P_2 is the probability of correctly identifying the signal x after it is received, and P_1 is the probability of guessing x before it is sent. Therefore, the amount of information carried by a signal is related to the amount of uncertainty it dispels after it is received as opposed to before it is sent.

Armed with a measure of information and a concept of an information channel, psychologists began to embrace the theory. The idea was that the human could be viewed as an information channel receiving signals through the senses and sending signals in response.

Channel Capacity. Perhaps the first behavioral effect related to uncertainty which was observed is that task performance generally falls off with higher stimulus or response uncertainty. This decrement in performance was observed in learning paradigms, perceptual recognition tasks, and perceptual reaction-time studies.

Is this decrement in performance related primarily to stimulus uncertainty or response uncertainty? Garner (1975) in his classic review of this early literature argues convincingly that the effect is attributable mostly to response uncertainty. When stimuli are regrouped to elicit responses of differing average uncertainty, large effects are observed even though stimulus uncertainty remains constant. However, when response uncertainty is held constant, and stimulus uncertainty is varied by offering or withholding knowledge of upcoming stimuli, the effect is only minimal.

Another question addressed in this early work concerns the channel capacity of the human observer. Given that the human could be viewed as an information channel, then as stimulus uncertainty (that is, amount of information received) is systematically increased, the amount of information transmitted should increase in step until the channel capacity is reached. At this point, the amount of transmitted information levels off. The amount of transmitted information is, in essence, a correlation measure between stimulus and response.

The results from experiments to measure channel capacity show that the limit seems to fall somewhere between 2.3 and 3.2 bits when subjects make absolute judgments on a single stimulus dimension. That is, if one asks a subject to order stimuli according to the magnitude of a given stimulus dimension, the subject will be able to use only five to nine ranks efficiently.

In his famous and entertaining review, George Miller (1956) relates this narrow range of capacity for absolute judgments to a similar range for immediate memory span. Most people can remember about seven items for a period of several seconds. A local telephone number, for example, can usually be remembered long enough to find a pencil and write it down. Tricks can be learned to increase this span These tricks, or recodings as Miller calls them, allow people to represent longer lists as shorter lists composed of recoded chunks. A good example of recoding is the use of six octal digits to represent 18 binary digits. Good evidence exists that recoding is a sign of expert performance. The point is that even with recoding, only about seven chunks can be remembered. Although Miller stopped short of declaring the discovery of a universal constant of psychology, the implication was clear: the human information processor seems to be limited to dealing with only about seven mental entities at a time.

Signal Detection Theory. Not long after the advent of information theory, another theory came onto the scene. As with information theory, this new approach had its origins in engineering. It was called signal detection theory (SDT) (Tanner and Swets 1953, 1954). SDT is statistical rather than probabilistic in that sensory evidence for various stimulus conditions is assumed to be normally distributed (see figure 1). In its simplest form when a stimulus to be detected is present, the distribution representing this state of the world is offset some distance to the right (that is, increased sensory evidence) with respect to the distribution representing the absence of the target stimulus; variance is unaffected. The observer has knowledge of both these distributions. On a given trial, the observer receives a certain amount of sensory evidence; the task then is to decide from which distribution the evidence comes. This task is accomplished by strategically placing a criterion to maximize gain or minimize loss at least for the ideal observer. Any evidence that equals or exceeds the criterion level warrants a yes response; otherwise, a no response is given. In effect, the observer is performing a statistical hypothesis test on the sensory evidence received.

The observer's performance is usually evaluated in terms of P(Hit) and P(False Alarm). From these data, a measure of sensitivity (d') and the criterion level ß can be derived. The former shows how well the observer discriminated between the two stimulus conditions (that is, how far apart the distributions were). In fact, d'_2 is proportional to the amount of information transmitted. However, ß is a measure of response bias. One of the main virtues of SDT is that d' and ß can be measured independently.

How well do humans perform as assessed by the SDT model? The answer is not too well. The problem is not so much with d', which we already know from the work with information capacity has an upper bound, as it is with how observers locate their ß. Two important variables have an effect on ß: stimulus probability and payoff structure. When observers know that the target stimulus is likely to be presented, they are inclined to give the yes response, and ß is smaller (that is, less strict). Of course, the opposite would occur if the observers had prior knowledge that the target stimulus is presented only infrequently. In a similar fashion, when the payoff matrix offers incentives for responding yes, observers will lower their criteria; with corresponding disincentives, criteria become stricter.

This situation is as it should be. In fact, observers are quite good at locating the optimal ß in balanced situations (that is, P(target) = 0.5 and incentives = disincentives) (Green and Swets 1966; Ulehla 1966). When stimulus probabilities are unequal, most researchers report less criterion shift than is optimal, although this shift is in the appropriate direction (Ulehla 1966; Dorfman 1969; Thomas and Legge 1970; Thomas 1975; Craig 1976; Kubovy 1977; Healy and Kubovy 1978, 1981). In other words, observers tend to not go far enough in adjusting their criteria to the situation.

A moment's reflection reveals why conservative placement of ß is suboptimal. The criterion represents the subjective point of neutrality between the two responses. Where should the ideal observer locate the neutral point? When one is neutral, no reason exists for favoring one response over the other. With the payoffs balanced, the probability of receiving a certain level of evidence (E) from the signal-present distribution (S) relative to the evidence level from the signal-absent distribution (N) should be offset precisely by the prior probabilities at the point of neutrality:

$$ß = P(E|S) / P(E|N) = P(N) / P(S).$$

To place the criterion lower would result in too many false alarms, and to place it higher would result in too few hits.

A number of explanations have been offered to account for this conservatism, but perhaps the best-supported explanation involves a phenomenon called probability matching (Dorfman 1969; Thomas and Legge 1970; Thomas 1975; Craig 1976; Healy and Kubovy 1978, 1981). We discuss probability matching in detail when we cover choice behavior Suffice it to say at this point that *probability matching* is a strategy whereby people attempt to match their response probabilities to the corresponding stimulus probabilities.

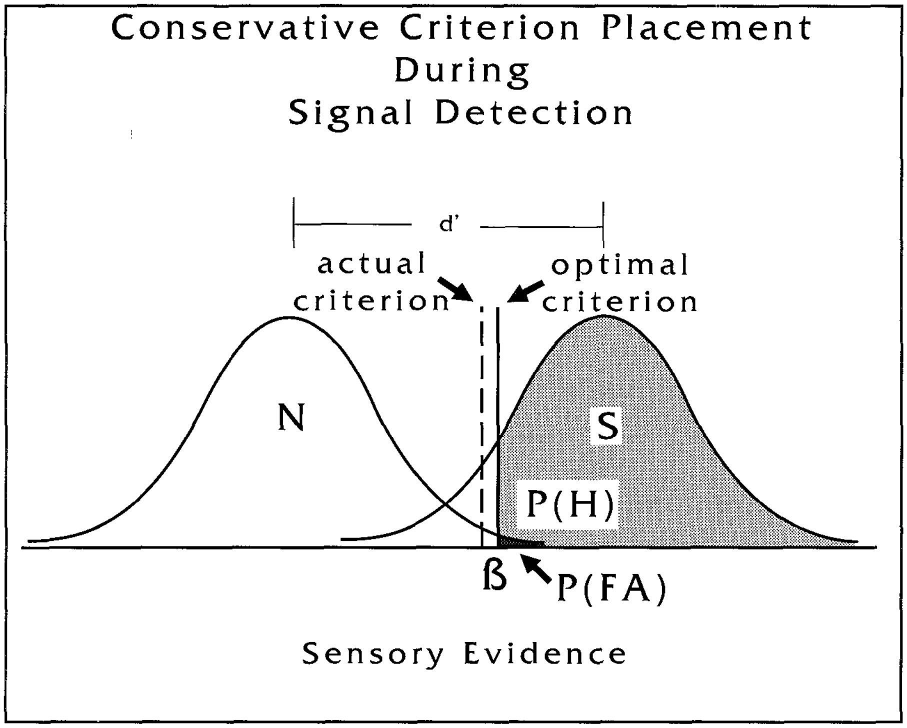

Figure 1. Conservative Criterion Placement.

The actual ß (dashed line) is placed to the left of the optimal ß (solid line) with unequal stimulus probabilities (P(N) = 87; P(S) = 12) The hit probability (P(H)) and the false alarm probability P(FA)) are indicated by the shaded areas under the curves

Probability matching leads to suboptimal behavior. Consider an observer trying to detect a signal that has a known prior probability of 25 percent (the optimal ß is 3). Assume that the observer has just had a run of rejections (for example, the no response has been given 10 times over the past 10 trials). Now, the observer might think it is time for a signal and as a consequence relax the criterion over several trials (for example, adjust ß to 2.5). Over this period, the observer is exposed to an increased probability of registering false alarms, and performance deteriorates. The process is asymmetric because the prior probabilities are unequal. In other words, the observer occasionally might set the criterion in anticipation of a stimulus event that should favor the more probable stimulus only less often than in the converse situation.

Another possible interpretation of the conservative ß placement is that the observer is miscalibrated for probability. That is, instead of judging ß = P(n) / P(s), as it should be for optimal performance, the observer sets ß = [P(n) - x] / [P(s) + x] simply because of a misjudgment of the objective probabilities. The topic of miscalibration is discussed in detail in the next section.

The picture to this point is that the human can be viewed as an information-processing system of limited capacity. In and of itself, this view is neither surprising nor damning. We know of no real system that has unlimited capacity. However for the KE, this limitation does pose a practical problem: how to elicit the sufficiently precise and accurate information concerning the DE's state of confidence regarding some fact or relationship. The absolute-judgment research suggests that the KE can expect only approximately 2.8 bits worth of precision. That is, the DE will use only about seven response categories when attempting to judge confidence in a given proposition, even though an infinitude of possible responses are offered. The SDT studies indicate that this confidence might be misplaced (that is, B might be taken as a measure of confidence).

Methods to Enhance Precision. Precision can be expanded by several means. One method is to make a series of absolute judgments by providing anchor points. This method is similar to recoding in that the judgments can be structured in a serial process so that the parallel-processing capacity is not overwhelmed by the information load. By offering the DE a point of reference that divides the range of confidence into two parts, the DE can first locate the confidence level above or below this so-called anchor point. Then the process can be repeated on the subrange containing the judged confidence level until the subrange becomes so small that the DE is unable to make further judgments. The problem is in providing meaningful anchor points. Of course, the natural points are absolute confirmation, absolute disconfirmation, and neutrality (that is, maximum uncertainty). Clearly, a means of providing additional anchor points is needed.

Some decision analysts have tried to represent anchor points graphically with the *probability wheel*, an adjustable pie chart (Spetzler and Stael von Holstein 1975). The DE is asked whether betting on the proposition under consideration or the probability wheel is preferred. This process is repeated with different "pie slices " When there is no preference, the level of confidence can be measured directly from the probability wheel.

The problem of accuracy is somewhat involved. We defer discussion of this issue until after we have considered how people make probabilistic judgments.

Man against the Models: Judgment

The study of probability and statistics has given us a variety of models that are designed to extract as much useful information as possible from uncertain or incomplete data. The research we are about to review focuses on whether humans process information as these models indicate that it should be processed.

Humans as Bayesians. One model well known to the AI community is the Bayesian probability theory. Suppose you are given the task of estimating the probability that a certain hypothesis holds given a set of data. The basic concept behind the Bayesian approach is that as additional information (E) becomes available, adjustments can be made to the prior probability of the hypothesis (H) to yield its posterior probability. The relationship is simply

$O(H|E) = LR(E|H) \times O(H)$,
where $O(H)$ is the prior odds that H holds
$LR(E|H) = P(E|H) / P(E|\text{not } H)$
$O(H|E)$ is the posterior odds of H given E,
and $O(x) = P(x) / (1 - P(x))$.

The adjustment factor is the likelihood ratio (LR). This relationship can be applied recursively; as new evidence (E') becomes available, the old posterior odds become the new prior odds assuming conditional independence (that is, the introduction of new data does not affect the conditional probabilities involving the old data).

When human posterior probability estimates are compared to those estimates by the model, the prevalent finding is that the human estimates are too conservative (that is, too close to the prior probabilities) (Edwards and Phillips 1964). A long series of "bookbag-and-poker-chip" studies were reported that were designed to describe this phenomenon. In these experiments, subjects are given samples of objects, usually poker chips of two different colors, and knowledge of the conditional, as well as the prior, probabilities. They are then asked to estimate, not calculate, the probability that the sample was drawn from one of two populations, usually bookbags.

Here is a synopsis of the findings. Conservatism is least when the sample size is small (Peterson, Schneider, and Miller 1965). Payoffs for accuracy reduced conservatism (Phillips and Edwards 1966). Conservatism falls off as prior probabilities become extreme (Peterson and Miller 1965). Larger conditional probabilities result in less conservatism, even though the likelihood ratios, which are all that matter to Bayes' formula, are held constant (Beach 1968). Finally, as the proportion of objects in each population is made comparable, conservatism diminishes or even reverses (Peterson and Miller 1965; Phillips and Edwards 1966).

Several explanations have been offered to account for conservatism. Fischhoff and Beyth-Marom (1983) catalog seven possible points in the application of Bayesian reasoning where bias could enter: (1) inappropriate hypothesis formation, (2) misassessment of subjective probabilities, (3) misapplication of prior odds, (4) inappropriate or incomplete assessment of likelihood ratios, (5) misaggregation (that is, calculation error), (6) incomplete search for evidence, and (7) misinterpretation of analysis. Research efforts were under way to sort out where and to what degree biases were affecting the application of Bayesian logic when a new finding was uncovered that suggested humans are not conservative Bayesians. In fact, they are not Bayesians at all.

This finding is called the base-rate fallacy. Whereas conservatism might be considered a mild bias considering the adjustments are in the right direction, the base-rate fallacy involves what appears to be a total disregard for prior probabilities. Kahneman and Tversky (1972a) report a particularly dramatic case of this fallacy. Consider the following problem.

Two cab companies, the Blue and the Green, operate in a given city (according to the color of the cab each runs). Eighty-five percent of the cabs in the city are Blue, the remaining 15 percent are Green. A cab was involved in a hit-and-run accident at night. A witness identified the cab as a Green cab. The court tested the witness' ability to distinguish Blue and Green cabs under nighttime visibility conditions. It found that the witness was able to identify each color correctly about 80 percent of the time and confused it with the other color about 20 percent of the time. What do you think are the chances that the errant cab was indeed Green, as the witness claimed?

The vast majority of their subjects reported probabilities close to 80 percent; the Bayesian answer is 41 percent. Apparently, the subjects were not adjusting prior odds (that is, base rate); they simply ignored them. This situation is the base-rate fallacy. Tversky's and Kahneman's subjects might have read more into this particular problem than was warranted, although the effect has been demonstrated in many diverse contexts (compare Kahneman and Tversky 1972b; Fischhoff, Slovic, and Lichenstein 1979) and among lay people as well as professionals. Evidence exists of the fallacy not only among practicing physicians but also in the medical literature (compare Bar-Hillel 1980)

Tversky and Kahneman (1980) use causality to account for the effect. They argue that base-rate information is observed when it fits into a causal schema of the problem, as with the bookbag-and-poker-chip studies. When the base rate does not fit the schema, it is discounted or ignored. Bar-Hillel (1980) offers an alternative explanation. She believes that people order information according to its perceived relevance. Highly relevant information would dominate less relevant information, thereby accounting for the effect. Specificity to the individual case is what determines relevance. Although specificity might be a slightly broader concept than causality, both explanations seem consistent.

Calibration. Human behavior does not seem to conform well to the Bayesian model. As suggested earlier, some of the discrepancies can be accounted for by misjudgment (that is, miscalibration) of subjective probability. Calibration is typically assessed by having people assign numeric values that represent their degree of confidence in their responses to items from problem-solving or judgment tasks. These values, scaled appropriately, are the subjective probabilities. A person is considered perfectly calibrated when the probability of a correct response is equal to its subjective probability (that is, the level of confidence).

The pervasive finding of calibration studies is that judges show overconfidence with low-objective probabilities and underconfidence with high-objective probabilities (compare Lichtenstein and Fischhoff 1977; figure 2). Interestingly, the overconfidence bias diminishes in easy tasks, and experts tend to show less bias than nonexperts (compare Lichtenstein, Fischoff, and Phillips 1982).

These miscalibrations are affected by psychological variables. In assessing positive events, subjective probabilities are higher when the events described would personally affect the assessor rather than some unknown person; for negative events, the opposite is found (Zakay 1983). Subjective probabilities are influenced by payoff structures (Phillips and Edwards 1966) Physicians' probability assessments of diagnoses are affected by the severity of the corresponding treatment consequences (Betaque and Gorry 1971). In competitive tasks requiring skill, people are prone to overestimate their own abilities (that is, subjective probability of success) (Howell 1972).

Evidence exists that calibration can be improved through learning (Lichtenstein and Fischhoff 1980). The learning is very fast, although it does not seem to generalize well to other situations.

Properties of Subjective Probabilities. Do subjective probabilities conform to the same relationships as objective probabilities? One such relationship is the requirement that the sum of all probabilities from a collectively exhaustive, mutually exclusive set of events equals one. For children estimating relative frequencies, the result depends on the sample size. With small samples (less than five trials), the sum exceeds one; for larger samples (greater than five trials), the sum is less than one. Adults' subjective probabilities sum closer to one, though only inconsistently (see Lee 1971 for review).

Another relationship is the multiplicative rule for conjoined, independent events (that is, P(A&B) = P(A) x P(B)). The general finding is that subjective probabilities of independent events are not multiplied together to yield the subjective probability of the compound event. The compound event's subjective probability is usually too high (Cohen and Hansel 1958; Cohen, Dearnaley, and Hansel 1958).

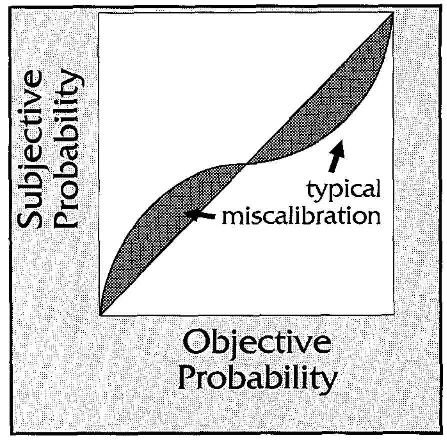

Figure 2. Miscalibration.

The shaded areas show the typical pattern of miscalibration between subjective and objective probabilities

For example, in Tversky and Kahneman's study (1983), 72 percent of the subjects ranked the statement "(Bjorn) Borg will lose the first set but win the match" as more probable than "(Bjorn) Borg will lose the first set." Even if the two events were viewed as perfectly correlated (that is, the independence assumption is violated), the statements should have been ranked equally. Apparently, the subjects were basing their estimates more upon what they knew about Bjorn Borg than upon what they knew about probability theory.

Judgments of Statistical Parameters. In regard to statistical, rather than probabilistic, judgments, a number of studies show that people do not appear to independently judge the mean and variance of a sample. For a given variance, the mean can be estimated fairly accurately (Beach and Swensson 1966). However, the variance of these estimates increases with sample variance and sample size (Spencer 1961; Beach and Swensson 1966). In addition, estimates of variability decrease as the mean increases (Hofstatter 1939; Lathrop 1967) as though the coefficient of variation were being estimated.

In order to determine how people estimate dispersion, Beach and Scopp (1967) tried to find the exponent that best simulated the estimates of their subjects. The normative exponent for variance, of course, is two. Low values tend to emphasize the effect of small

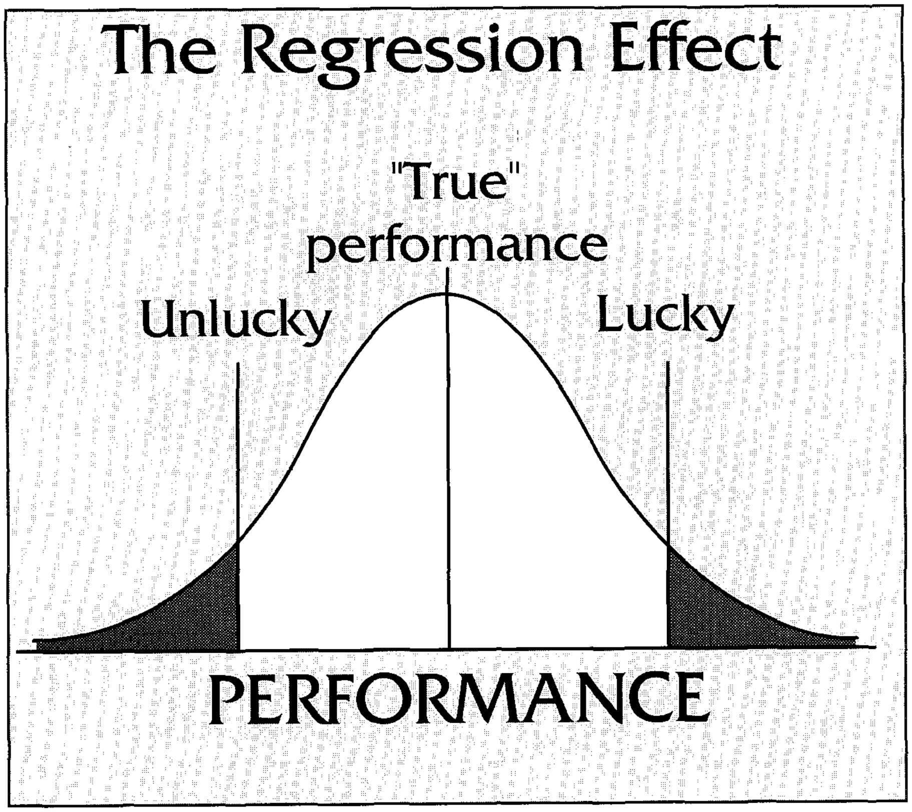

Figure 3. The Regression Effect

The "true" level of performance is shown as the mean; deviations to the right (lucky) or the left (unlucky) are random The shaded areas indicate the probabilities of exceeding the deviations on subsequent trials Because these areas are small compared with the rest of the area under the curve, subsequent scores are likely to be closer to the mean

deviations; high values emphasize large deviations. They found an exponent of 0.39, indicating that in their study people were strongly influenced by small deviations.

Correlation is a statistical concept that has received much attention from cognitive psychologists mainly because it provides a model and a measure of relationship. The cognitive process of uncovering relationships is fundamental to learning and thinking. Ironically, humans seem to estimate correlation from data samples rather imperfectly in certain circumstances. Alloy and Tabachnik (1984) give an excellent survey of the research in this area. They discuss work that shows humans can estimate correlation, especially positive correlation, accurately except when their prior expectations contradict the objective data. In this case, the prior expectations seem to dominate the judgments.

An example of the effect of prior expectations is the so-called illusory correlation. This phenomenon occurs when judges overestimate correlation even in cases where none exists or where only negative correlation is present Evidence of illusory correlation has been found in a variety of situations among professionals and nonprofessionals (Smedslund 1963; Chapman and Chapman 1967, 1969). Studies show that illusory correlations arise and persist because of a tendency to discount or ignore disconfirming (that is, noncorrelational or anticorrelational) evidence contrary to prior expectations. This effect has been used to account for the persistence of social and ethnic stereotypes (Hamilton and Rose 1980).

Numerous explanations exist of the genesis of illusory correlation. One suggestion is that it is a vestige of biological adaptation (McArthur 1980). The argument is that the need to discover biologically important relationships, such as between the ingestion of certain foods and subsequent illness, biased the cognitive system to overdetect relationship. Another possibility is that the practice of undervaluing disconfirming evidence is learned (compare Einhorn and Hogarth 1978). Indeed, there are many real-life circumstances where one cannot readily obtain disconfirming data. Consider the case of a job applicant who is rejected because of the lack of a particular qualification. How would the employer ever assess the value of this qualification unless the person without it was observed? An extreme example is the case of superstitious behavior in which it is deemed unwise to seek disconfirming evidence because of some untoward event occurring on perhaps only a single trial.

Correlation appears to be a difficult concept for people. Inhelder and Piaget (1958) reported that the concept of correlation does not develop until age 14 or 15. Wason (1960) showed that humans tend to not seek disconfirming evidence even when to do so is more efficient. Einhorn and Hogarth (1978) in their review of the literature expand on this point. They indicate that the formal notion of experimental control (that is, the search for disconfirming evidence) came late in the history of scientific thought. Also, the need for placebo conditions and double-blind designs in clinical trials is a recent development (Shapiro 1960). In any event, the rather disconcerting conclusion is that people appear to have perfect vision for what confirms their beliefs and severe myopia for what disconfirms them.

A concept closely related to correlation is *regression*. The basic idea of regression is that nonrandom and random processes can be separated using statistical methods. The former is represented by a mean value and the latter by a deviation from the mean. Usually, the random process is assumed to be normally distributed so that large deviations are less probable than small deviations (compare figure 3).

The regression effect can be expressed simply as a tendency for observations to cluster about the mean. For example, consider a piano student performing a composition for the teacher. The student does remarkably well, and the teacher offers appropriate praise. When the student performs the piece a second time, the student does less well. This lesser performance could be a result of the regression effect. The first effort was not representative of the true level of proficiency; subsequent regression toward the mean occurred. Now assume the student performs extremely poorly and is scolded by the teacher. On the following trial, the student's performance improves again because of the regression effect. The teacher, however, is left with the distinct impression that punishment is far more effective than praise, when, in fact, the changes in performance were out of the student's control. Again, attempts to find disconfirming evidence guard against such misconceptions.

Admittedly, the regression effect is subtle. However, even with sophisticated subjects given hints and prodding, susceptibility to the regression effect seems to persist Kahneman and Tversky (1973) described an experiment with psychology graduate students who were to give their 95 percent confidence interval around an IQ score of 140. They were told explicitly that the 140 score was the sum of a true score and a random error. Because this score is well above the population mean, one could infer that the error component is large (that is, there should be a significant regression effect). Seventy-three of the 108 subjects reported confidence intervals symmetric around the 140 score, ignoring the random element. The authors conclude that there is a tendency for people "...to predict as if the input information were error free...."

Heuristics of Probabilistic Judgments. The work of Kahneman and Tversky deserves special consideration because their ideas provide a useful framework with which to integrate and amplify much of what has been discussed. In their view, humans use heuristics to process uncertain information (Tversky and Kahneman 1971, 1973, 1974, 1980, 1983; Kahneman and Tversky 1972a, 1972b, 1973, 1982; Kahneman, Slovic, and Tversky 1982). One heuristic is *representativeness* which is based on the assumption that the more an object (or event) typifies a corresponding class (or process), the higher the probability of a relationship between the two. Representativeness seems to have a certain commonsense plausibility about it However, Kahneman and Tversky invoke representativeness to explain a wide variety of errors and fallacies We now examine their explanations in detail.

They explain the base-rate fallacy with representativeness People focus on how closely a hypothesis matches the facts given at the near exclusion of what is known about the tenability of the hypothesis in general (that is, the base rate). Even when uninformative though case-specific facts are given, people tend to ignore the base rate

Another frequently observed error is the inattention to sample size. For example, when asked to judge the probabilities of attaining more than 60 percent male births from both a small (15 births per day) and a large (45 births per day) hospital, most respondents gave the same values. Of course, probability should have been higher for the small hospital because the proportions are more variable as a result of its smaller sample size (for example, consider a very small hospital with 1 birth per day; the probability should be about 50 percent). Representativeness suggests that people ignore sample size because they are focusing on the relationship between the sample parameter and the corresponding population parameter

Representativeness also explains why most people when asked to generate random sequences tend to produce too many short runs. The explanation is that people expect randomness to be exhibited over the short term as well as the long term. In a similar fashion, representativeness explains the gambler's fallacy, which comes from a misguided belief that the termination of a long run is a predictable event

People often show a propensity for making cavalier predictions based on only scant information. Kahneman and Tversky describe an experiment that illustrates this tendency. Two groups were given accounts of lesson presentations of several student teachers. One group evaluated the performance of each student teacher on a percentile basis; the second group similarly was to predict performance five years hence. The range of evaluations matched the range of predictions, but given the low predictability value of the accounts, one would expect the predictions to have a much tighter range (that is, a clustering around the mean value) because of the regression effect.

The illusion of validity is the tendency for people to express overconfidence in their predictions. This effect is not surprising given that representativeness controls both prediction and its associated confidence. Overconfidence is observed in predictions based on redundant information contrary to statistical theory. Presumably, redundancy reifies the representative quality of the information and, thereby, enhances confidence. However, redundancy adds nothing to the predictive value of the information.

Kahneman and Tversky use representativeness to suggest why the regression effect seems so elusive. They submit that the regression effect operates contrary to what people expect according to the representativeness heuristic Extreme observations are expected to be representative of their underlying process; regression toward the mean is incompatible with this notion.

A second heuristic is called *availability.* The logic of availability is that easily recallable information has higher associated probability than less easily recallable information In general, frequently presented material is both more probable and more easily recalled (compare Kintsch 1970). However, other factors such as salience,

...the human can be viewed as an information-processing system of limited capacity.

recency, and primacy affect how easily information can be recalled. Therefore, reliance on this heuristic can lead to biases and errors.

Biases from the availability heuristic often appear in tasks where frequency is to be estimated. In estimating the length of lists with people's names, for example, lists with more famous and familiar names are usually judged to be longer Also, people estimate the frequency of English words starting with *R* or *K* higher than words with these consonants as the third letter, even though the opposite is true. Searches for words in memory with a first-letter key are easier than those with a third-letter key

Kahneman and Tversky explain the illusory correlation in terms of availability. They suggest that the frequency of co-occurrences of two events can be overestimated if there is a strong associative bond between these events in memory Therefore, events strongly associated in memory are judged to occur often together, even though this assumption might not be valid.

A third heuristic presented by Kahneman and Tversky is *adjustment and anchoring*. The procedure involved in adjustment and anchoring is to estimate or compute an initial starting point and then make adjustments away from the initial value. On the surface, this heuristic seems reasonable. The concept behind Bayesian probability theory, regression analysis, analysis of variance, and factor analysis resembles the adjustment and anchoring procedure. However, as we soon see, people tend to make conservative adjustments.

Adjustment and anchoring could account for why people often overestimate the probability of conjunctive events and underestimate the probability of disjunctive events (Cohen, Chesnick, and Haran 1972). Assume that an individual examining a conjunctive (or a disjunctive) event begins by looking at the probability of an elementary event (for example, the coin comes up heads) Given that this person then adjusts down for the conjunctive event (or up for the disjunctive event) but insufficiently, the probability estimate would be too high (or too low).

In a similar fashion, adjustment and anchoring is used to explain miscalibration of subjective probabilities. As stated earlier, usually infrequent events are overestimated, and frequent events are underestimated. Kahneman and Tversky suggest that people might use the neutral confidence point as an anchor point; insufficient adjustments from this point would cause conservatism (that is, the pattern described).

...decision makers seem to excercise good sense by evaluating risk. On the other hand, decision makers seem to gamble inconsistently and subobtimally.

Methods for Unbiasing Judgments. The research comparing human judgment against prescriptions of normative models clearly shows that humans do not behave in accordance with the models. Humans also do not seem to have an intuitive understanding of stochastic processes.

For the KE, the effect of these human frailties can be minimized, though usually not completely eliminated The best way to reduce the effects is to avoid using probabilistic or statistical judgments by the DE as much as possible. Of course, if the hard data are available, they should be used. When the data are unavailable, the necessary probabilities and statistics can be estimated from the DE's subjective probabilities to primitive events.

The problem then reduces to minimizing miscalibrations. Several approaches could be taken at this point. DEs could be trained to become better calibrated (Lichtenstein and Fischhoff 1980). Another technique is to require the DE to offer reasons for why the event(s) under consideration should not or could not occur (Fischhoff 1982). The purpose is to have the DE focus attention explicitly on negative scenarios, evidence, or possibilities, thereby reducing overconfident judgments. Some researchers have suggested that better calibration can be achieved by asking for subjective probabilities associated with fixed values from the underlying distribution rather than for the values associated with fixed probabilities of the distribution (Seaver, von Winterfeldt, and Edwards 1978) Whichever method is employed, the KE should never ask leading questions, lest the KE's own biases return in disguise.

Spetzler and Stael von Holstein (1975) discuss how the interview process should be structured in order to reduce or remove biases in estimating probabilities. They break down the process into five phases: (1) motivating, (2) structuring, (3) conditioning, (4) encoding, and (5) verifying. The motivating phase has two purposes: to familiarize the DE with the procedure and its importance and to uncover any biasing related to the process, such as any implicit payoff structure associated with responses.

The structuring phase is designed to define and structure the uncertain quantities. The aim is to eliminate any ambiguities so that the DE knows exactly what is being assessed. The DE should be encouraged to identify what is relevant to the task.

The conditioning phase is concerned with reducing potential bias. The DE is asked to justify the response given both in terms of explicit information provided by the KE and of background information The KE should be alert for signs of representativeness and availability from these justifications. In both cases, the DE should be queried about other possible cases which might influence the estimate but which have gone unmentioned.

During the encoding phase, the DE makes the estimates Spetzler and Stael von Holstein advocate the use of the probability wheel, although other psychophysical techniques are available. The discriminations should become increasingly difficult. Also, each quantity being estimated should be randomly selected to avoid order or carryover effects. A cumulative probability distribution generated from the responses highlights any inconsistencies or gaps in the data.

The verification phase is used to provide feedback to the DE. The cumulative distribution is shown to the DE for comment. If the DE is unsatisfied with the plot, the procedure might have to be repeated. Another method of verification is to construct pairs of bets based on the distribution. These bets are offered to the DE, who is to express a preference if any. This method checks the consistency of the data.

Man against the Odds: Choice

As foreshadowed by the preceding discussion, the topic of choice behavior in the face of uncertainty is inextricably bound to the specific topic of gambling behavior. The relationship between the two can be traced back at least to the nineteenth century with Daniel Bernoulli's (1954) solution of the St. Petersburg paradox. Assume we offer you a chance to play a game. The game is simply that we flip a coin until a head appears. We pay $\$2_{n-1}$, where n is the trial on which the first head appears. How much would you be willing to pay in order to play this game? The view of how choices are made had been that people always choose between alternatives in order to maximize expected value. The expected value for this game is infinite given a true coin. Indeed, an infinite amount is only the expected return; you might win even more! You should be willing to put up all your assets and all you could borrow according to the expected value rule. Would you? Most people would not. This deviation from the expected value rule is the St. Petersburg paradox.

Maximizing Utility. Bernoulli's solution was that people do not maximize expected value but rather maximize utility, which is the subjective value that individuals place on commodities, opportunities, or states of the world. It is a measure of desirability. Bernoulli's solution was important for several reasons. First, it shifted the emphasis from objective measures (that is, expected value) to subjective measures (that is, utility) in explaining choice and decision-making behavior. Second, it implied that choosing or deciding between alternatives is an individual affair; different people have different utility functions. Third, it suggested that risk (that is, uncertainty in outcomes) is something people tend to avoid. Therefore, in the case of the St. Petersburg paradox, people are not willing to put up all their worldly possessions against the flip of a coin, no matter how high the expected value of the game, because there is too much variance in the outcomes.

	Choice 1	
Gamble 1	100%	$1,000,000
Gamble 2	10%	$5,000,000
	89%	$1,000,000
	1%	$0
	Choice 2	
Gamble 3	11%	$1,000,000
	89%	$0
Gamble 4	10%	$5,000,000
	90%	$0

These gambles can be viewed as though they represent lottery prizes, with ticket numbers associated with the various prizes:

	Ticket Numbers	
1	2-11	12-100
	Choice 1	
Gamble 1		
$1,000,000	$1,000,000	$1,000,000
Gamble 2		
$0	$5,000,000	$1,000,000
	Choice 2	
Gamble 3		
$1,000,000	$1,000,000	$0
Gamble 4		
$0	$5,000,000	$0

Table 1.
An Example of Savage's Axiom.

The utility function represents an ordering of preferences. The shape of the utility function can be illuminating in assessing risk-taking behavior. A concave-down function is said to indicate risk aversion. Why? The slope of the utility function can be viewed as a measure of sensitivity to variation in the underlying commodity. When the curve is concave down, the slope is a decreasing function of this commodity. In other words, an increment of payoff gives less than a commensurate increment in utility because the extra payoff carries extra variation (that is, increased risk). In an analogous fashion, concave-up functions indicate risk-seeking behavior (for example, gambling), and a linear function indicates risk neutrality (that is, an expected value maximizer)

Most people are predominately risk averse, as reflected by their utility functions. How then can a given person be willing both to buy insurance and to bet on the horses? Friedman and Savage (1948) proposed that such apparently inconsistent behavior could be explained by a utility function with two concave-down segments and an interposed concave-up segment. People can be both risk averse and risk preferring.

Inconsistent Choice Behavior. Utility can account for many apparent inconsistencies, but some effects remain difficult to explain no matter what theory is invoked. Lichtenstein and Slovic (1971, 1973) uncovered such an example. They asked their subjects to choose between bets with high probabilities of winning but low payoffs and those with low probabilities of winning but high payoffs. They then asked the subjects to bid for the bets. Curiously, they found that when the high-probability bets were chosen, they later lost to the high-winning bets in terms of bids. The authors suggest that these reversals might be another instance of the adjustment-and-anchoring heuristic of Kahneman and Tversky in operation.

Another inconsistency is related to an axiom of utility theory. The axiom states that when deciding between two alternative gambles with a common outcome, the value of this outcome cannot affect the choice (that is, it could be anything). This postulate is known as Savage's (1954) *independence principle*, because the choice is independent of the value of the common outcome. Although this axiom seems logically sound, Slovic and Tversky (1974) found that people often violated it. They gave their subjects choices to check for the consistent application of Savage's axiom (see table 1).

When viewed in this manner, it is clear that for tickets numbered 12 and greater, nothing is contributed to the decision-making process; the outcomes are identical The independence principle says you can ignore them. The decision comes down to a choice between a risky gamble for $5,000,000 and a less risky gamble for $1,000,000. Consistency demands that either gamble 1 and gamble 3 (for the risk averters) or gamble 2 and gamble 4 (for the risk takers) are chosen The majority of the subjects (17 of 29) mixed their choices.

No good explanation seems to exist for why some people violate the independence principle, which seems plausible if not logically compelling Perhaps people view certainty in a qualitatively different way than they view uncertainty. After all, a gift of $1 million does not seem quite comparable to a gamble for $5 million regardless of the odds. Clearly, Savage's axiom does not capture the way some people behave, but it is difficult to argue that this axiom is at variance with the way people should behave.

Probability Matching. Perhaps the most studied gambling behavior is probability matching To demonstrate this phenomenon, people are asked to guess the outcomes of independent Bernoulli trials (that is, stable probabilities) given a particular probability structure. Of course, the optimal strategy is to guess the most probable outcome. However, most people match their response probabilities to the corresponding outcome probabilities (compare figure 4). This strategy is suboptimal because whenever subjects guess the long shot, they are exposing themselves needlessly to a loss. This probability-matching tendency is a manifestation of the gambler's fallacy.

Interestingly, people can be induced to perform optimally. One way is to increase the number of outcomes (Gardner 1958). Another way is to emphasize gambling rather than the problem-solving aspects of the task in the instructions (Goodnow 1955).

Resolving Inconsistencies. This sampling concludes our survey of the literature. Our image of the humans making choices in the face of uncertainty is somewhat perplexing. On the one hand, decision makers seem to exercise good sense by evaluating risk. After all, not many sane people would be willing to risk everything to play the St Petersburg paradox game. On the other hand, decision makers seem to gamble inconsistently and suboptimally.

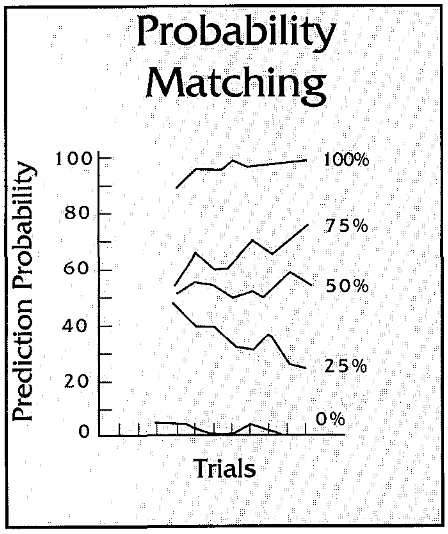

Figure 4. Probability Matching.

These schematic data are typical of actual results from probability-matching experiments Each plot shows the probability of choosing a given stimulus whose prior probability is displayed on the right over five trial blocks (1 unit = 5 trials)

This inconsistent behavior presents a dilemma to the KE. If apparent inconsistencies are detected in the DE's decision-making behavior, are they real or merely results of poorly articulated problem solving? If real, how can they be resolved? In order to detect real inconsistencies, the KE must obtain detailed justifications for each decision made. Restructuring questions by asking about what factors would have to be different to alter the decision can be helpful in this regard The KE should be wary of possible confabulations *Confabulations*, a term from clinical medicine, are fabricated explanations for an event or behavior that seems to defy rational explanation in order to cover up some weakness or failing.

Having uncovered a genuine inconsistency, the KE has several options. One option is to point out the problem to the DE and let the DE resolve the inconsistency. Another approach is to model the decision making mathematically. Evidence exists that linear models (regression or discriminant analysis models) often perform quite well (Dawes 1982). The DE's role in constructing the model would be to identify relevant variables. This approach does have merit in that it allows the humans to focus on what they do best, detect relationships, and it lets the machines to do what they do best, integrate information.

Getting It Automated

The foregoing review emphasizes the recurring observation that humans often introduce biases, distortions, or even errors in their processing of uncertain knowledge. One major conclusion from this observation is that the processing of uncertain knowledge could be improved if it could be objectified. An expert system seems to be a most suitable vehicle for acquiring the necessary objective information as a product of its use. This view has been expressed elsewhere (Cohen 1985).

A first step toward building a system capable of automatically acquiring such information is the design and implementation of a usage log, a database that would contain detailed records of each consultation with the system (sufficient to reconstruct the consultation). The usage log could be used to periodically test the appropriateness and validity of the knowledge represented in the system. Formal knowledge representations—whether production rules, frames, semantic nets, or whatever—are embodiments of associations between facts, objects, or concepts. The degree of association between these entities usually takes the form of likelihood ratios, certainty factors (CFs), or some other form of uncertainty representation. Because the values of these measures of association are typically based on the DE's subjective assessment, they are subject to possible bias. The usage log provides the objective data from which these assessments can be verified or corrected.

The obvious question is how to integrate objective data with subjective assessments. This question is nontrivial. One possible approach is to put the subjective assessments on the same dimension as the objective data. The use of contingency tables is a convenient way to accomplish this end in many cases. For example, consider a simple rule from a rule-based system. This rule has a single antecedent, A (or a conjunction of antecedents considered as a simple antecedent for the purposes of this discussion), and a consequent, C. Now, rather than directly ask the DE for a likelihood ratio or a CF to value the strength of the association between A and C, the KE would ask the DE to fill in the 2 x 2 contingency table based on the DE's experience or knowledge. Each cell of the table represents one of the four possible states of truth. The table might appear as shown in table 2.

The values in each cell are relative estimates of the DE's confidence in the corresponding state of truth. Ideally, each value would be the number of cases observed.

More typical measures of uncertainty can easily be derived from such tables. The likelihood ratio is 3.33 (that is, 50/60 / 10/40) for this example. The CF, defined as

CF = P(C|A) - P(C) / 1 - P(C) when A supports C
and
CF = P(C|A) - P(C) / P(C) when A supports not C,
would have a value of 0.58 (that is, 50/60 - 60/100 / 1 - 60/100).

Assume that the system is run, and it is determined by independent means that the consequent of the rule was false but that the antecedent was true. The table would be updated as shown in table 3.

The likelihood ratio now is 3.11, and our table-based CF is 0.56. In this manner, the system could acquire objective data and update subjective assessments.

Several points deserve comment. First, as the system gains experience, its measures of association converge toward the true values (that is, values supported by the objective observations). Second, with initially confident subjective estimates (that is, large cell values), this convergence takes longer. Third, this methodology naturally incorporates several of the psychological techniques for unbiasing subjective assessments of uncertainty, such as focusing on disconfirming conditions and using concrete measures (for example, frequencies rather than probabilities). Finally, limitations of precision with human judgments are overcome; precision is limited only by the amount of data acquired.

	A	not A
C	50	10
not C	10	30

	A	not A
C	50	10
not C	11	30

Tables 2 and 3.
Original Contingency Table (above).
Updated Contingency Table (below).

An additional benefit in acquiring objective data through system usage is that statistical information could be derived from such data about the efficiency of the knowledge base. Are sets of rules consistently firing together? Are some rules never firing? Are some rules always firing? This kind of information would be invaluable to a KE trying to tune an expert system.

One limitation to the approach proposed here is that not all uncertain knowledge conforms well to a contingency table analysis. For example, how should a DE analyze this rule: "If the ozone layer depletion continues at its present rate, then the polar ice caps will melt by the year 2050." This rule represents a scenario that would occur only one time if at all. Moreover, how could the system ever update the DE's estimate? In such cases, one can fall back on the reliability of the model used to make the prediction and the plausibility of its underlying assumptions. The acquisition of verifying data would be of little use because the knowledge could be tested only once at most.

Clearly, building such a system is an ambitious project. Besides a usage log, for example, the system would require support facilities, including an editor to maintain the log. However, the stage is set for knowledge-based systems to begin assuming the tasks for which they are best suited and for which the human is least well equipped to handle. The processing of uncertain knowledge seems to be such a task

Acknowledgments

The authors thank Clay Clayworth for his artful contributions to the manuscript This work was supported by the Veterans Administration Research Service and Expert Systems Design, Inc

References

Alloy, L B., and Tabachnik, N 1984 Assessment of Covariation by Humans and Animals: The Joint Influence of Prior Expectations and Current Situational Information *Psychological Review* 91:112-149

Bar-Hillel, M 1980 The Base-Rate Fallacy in Probability Judgment *Acta Psychologia* 44:211-233

Beach, L R 1968. Probability Magnitudes and Conservative Revision of Subjective Probabilities. *Journal of Experimental Psychology* 77:57-63.

Beach, L R., and Scopp, T. S 1967. Intuitive Statistical Inferences about Variances. Seattle,Wash Mimeograph

Beach, L R., and Swensson, R G 1966 Intuitive Estimation of Means *Psychonomic Science* 5:161-162

Bernoulli, D 1954 Exposition of a New Theory on the Measurement of Risk Translated by L Somer. *Econometrica* 22:23-26

Betaque, N E , and Gorry, A 1971

Automating Judgmental Decision Making for a Serious Medical Problem *Management Science* B17:421-434.

Chapman, L J., and Chapman, J. P 1969. Illusory Correlation as an Obstacle to the Use of Valid Psychodiagnostic Signs *Journal of Abnormal Psychology* 74:271-280

Chapman, L J, and Chapman, J. P 1967 Genesis of Popular but Erroneous Psychodiagnostic Observations *Journal of Abnormal Psychology* 72:193-204

Cohen, J; Chesnick, E. I; and Haran, D 1972. A Confirmation of the Inertial-Psi Effect in Sequential Choice and Decision *British Journal of Psychology* 63:41-46.

Cohen, J.; Dearnaley, E J; and Hansel, C E. M 1958 The Mutual Effect of Two Uncertainties *Durham Research Review* 2:215-222

Cohen, J, and Hansel, C E M. 1958 The Nature of Decision in Gambling *Acta Psychologia* 13:357-370

Cohen, P. R 1985 *Heuristic Reasoning about Uncertainty: An Artificial Intelligence Approach.* Boston: Pitman.

Craig, A. 1976 Signal Recognition and the Probability-Matching Decision Rule. *Perception and Psychophysics* 20:157-162

Dawes, R. M 1982 The Robust Beauty of Improper Linear Models in Decision Making In *Judgment under Uncertainty: Heuristics and Biases*, eds. D Kahneman, P Slovic, and A Tversky, 391-407 New York: Cambridge University Press

Dorfman, D D. 1969. Probability Matching in Signal Detection *Psychonomic Science* 17:103.

Edwards, W, and Phillips, L D. 1964. Man as Transducer for Probabilities in Bayesian Command and Control Systems. In *Human Judgment and Optimality*, 360-401 eds M. W. Shelly, II, and G. L Bryan New York: Wiley

Einhorn, H J, and Hogarth, R. M 1978 Confidence in Judgment: Persistence of the Illusion of Validity *Psychological Review* 5:395-416.

Fischhoff, B. 1982 Debiasing In *Judgment under Uncertainty: Heuristics and Biases*, eds D Kahneman, P Slovic, and A. Tversky, 422-444. New York: Cambridge University Press

Fischhoff, B, and Beyth-Marom, R 1983 Hypothesis Evaluation from a Bayesian Perspective. *Psychological Review* 90:239-260

Fischhoff, B.; Slovic, P.; and Lichtenstein, S 1979 Subjective Sensitivity Analysis. *Organizational Behavior and Human Performance* 23:339-359

Friedman, M, and Savage, J. 1948 The Utility of Choice Involving Risk *Journal of Political Economy* 56:279-304.

Gardner, R. A 1958 Multiple-Choice Decision Behavior *American Journal of Psychology* 71:710-717

Garner, W R 1975. *Uncertainty and Structure as Psychological Concepts.* Huntington, NY: Robert E. Krieger

Goodnow, J J 1955. Determinants of Choice Distribution in Two-Choice Situations. *American Journal of Psychology* 68:106-116

Green, D M, and Swets, J A 1966 *Signal Detection Theory and Psychophysics.* Huntington, NY: Robert E Krieger

Hamilton, D L, and Rose, T L. 1980 Illusory Correlation and the Maintenance of Stereotypic Beliefs. *Journal of Personality and Social Psychology* 39:832-845

Healy, A F., and Kubovy, M. 1981. Probability Matching and the Formation of Conservative Decision Rules in a Memory Analog of Signal Detection *Journal of Experimental Psychology: Human Learning and Memory* 7:344-354

Healy, A F, and Kubovy, M. 1978 The Effects of Payoffs and Prior Probabilities on Indices of Performance and Cutoff Location in Recognition Memory *Memory and Cognition* 6:544-553

Hofstatter, P R. 1939 Ueber die schaetzung von gruppeneigenschaften *Zeitschrift fuer Psychologie* 145:1-44.

Howell, W. C. 1972 Compounding Uncertainty from Internal Sources *Journal of Experimental Psychology* 95:6-13.

Inhelder, B, and Piaget, J 1958 *The Growth of Cognitive Thinking from Childhood to Adolescence* New York: Basic Books.

Kahneman, D.; Slovic, P; and Tversky, A, eds 1982. *Judgment under Uncertainty: Heuristics and Biases.* New York: Cambridge University Press

Kahneman, D., and Tversky, A 1982 On the Study of Statistical Intuitions *Cognition* 11:123-141

Kahneman, D., and Tversky, A. 1973 On the Psychology of Prediction *Psychological Review* 80:237-251

Kahneman, D, and Tversky, A 1972a. On Prediction and Judgment *ORI Research Monograph* 12

Kahneman, D, and Tversky, A 1972b Subjective Probability: A Judgment of Representativeness *Cognitive Psychology* 3:430-454

Kintsch, W 1970 Models for Free Recall and Recognition. In *Models of Human Memory*, 331-373 ed D A Norman New York: Academic

Kubovy, M 1977 A Possible Basis for Conservatism on Signal Detection and Probabilistic Categorization Tasks *Perception and Psychophysics* 22: 277-281

Lathrop, R G 1967 Perceived Variability. *Journal of Experimental Psychology* 23:498-502

Lee, W. 1971. *Decision Theory and Human Behavior* New York: Wiley

Lichtenstein, S., and Fischhoff, B 1980. Training for Calibration. *Organizational Behavior and Human Performance* 26:149-171

Lichtenstein, S, and Fischhoff, B 1977 Do Those Who Know More Also Know More about How Much They Know? *Organizational Behavior and Human Performance* 20:159-183

Lichtenstein, S; Fischhoff, B; and Phillips, L D 1982 Calibration of Probabilities: The State of the Art to 1980 In *Judgment under Uncertainty: Heuristics and Biases*, 306-334 eds. D Kahneman, P Slovic, and A. Tversky New York: Cambridge University Press

Lichtenstein, S, and Slovic, P 1973 Response-Induced Reversals of Preference in Gambling: An Extended Replication in Las Vegas. *Journal of Experimental Psychology* 101:16-20.

Lichtenstein, S, and Slovic, P 1971. Reversals of Preference between Bids and Choices in Gambling Decisions. *Journal of Experimental Psychology* 89:46-55

McArthur, L Z 1980 Illusory Causation and Illusory Correlation: Two Epistemological Accounts *Personality and Social Psychological Bulletin* 6:507-519

Miller, G A 1956 The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review* 63:81-97.

Peterson, C R, and Miller, A J 1965 Sensitivity of Subjective Probability Revision *Journal of Experimental Psychology* 70:117-121.

Peterson, C R.; Schneider, R. J.; and Miller, A J 1965 Sample Size and the Revision of Subjective Probability *Journal of Experimental Psychology* 69:522-527

Phillips, L D, and Edwards, W 1966 Conservatism in a Simple Probability Inference Task *Journal of Experimental Psychology* 72: 346-354

Savage, L J 1954 *The Foundations of Statistics* New York: Wiley

Seaver, D A; von Winterfeldt, D; and Edwards, W 1978 Eliciting Subjective Probability Distributions on Continuous Variables *Organizational Behavior and Human Performance* 21:379-391.

Shannon, C E 1948 A Mathematical Theory of Communication *Bell System Technical Journal* 27:379-423, 623-656.

Shapiro, A K. 1960. A Contribution to a

History of the Placebo Effect *Behavioral Science* 5:109-135

Slovic, P, and Tversky, A 1974 Who Accepts Savage's Axiom? *Behavioral Science* 19:368-373

Smedslund, J 1963 The Concept of Correlation in Adults *Scandinavian Journal of Psychology* 4:165-173

Spencer, J 1961 Estimating Averages *Ergonomics* 4:317-328

Spetzler, C S, and Stael von Holstein, C 1975 Probability Encoding in Decision Analysis *Management Science* 22:340-358

Tanner, W P, Jr, and Swets, J A 1954 A Decision-Making Theory of Visual Detection *Psychological Review* 61:401-409

Tanner, W P, Jr, and Swets, J. A 1953 A New Theory of Visual Detection, Technical Report, 18, Electronic Defense Group, Univ of Michigan

Thomas, E A C 1975. Criterion Adjustment and Probability Matching *Perception and Psychophysics* 18:158-162

Thomas, E A C, and Legge, D. 1970 Probability Matching as a Basis for Detection and Recognition Decision *Psychological Review* 77:65-72

Tversky, A, and Kahneman, D 1983. Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment *Psychological Review* 4:293-315

Tversky, A, and Kahneman, D 1980 Causal Schemas in Judgment under Uncertainty In *Progress in Social Psychology*, vol I, 49-72 ed M Fishbein Hillsdale, NJ: Lawrence Erlbaum

Tversky, A, and Kahneman, D 1974 Judgment under Uncertainty: Heuristics and Biases *Science* 185:1124-1131

Tversky, A, and Kahneman, D 1973 Availability: A Heuristic for Judging Frequency and Probability *Cognitive Psychology* 5:207-232

Tversky, A, and Kahneman, D 1971 Belief in the Law of Small Numbers *Psychological Bulletin* 76:105-110

Ulehla, Z J 1966 Optimality of Perceptual Decision Criteria *Journal of Experimental Psychology* 71:564-569

Wason, P C 1960 On the Failure to Eliminate Hypotheses in a Conceptual Task *Quarterly Journal of Experimental Psychology* 12: 129-140

Zakay, D 1983 The Relationship between the Probability Assessor and the Outcome of an Event as a Determiner of Subjective Probability *Acta Psychologia* 53:271-280

For free information, circle no. 60