Modeling Expert Forecasting Knowledge for Incorporation into Expert Systems

Robert M. Hamm, PhD

Institute of Cognitive Science
University of Colorado
Boulder, CO 80309

July 7, 1991

ICS Tech Report 91-12

Modeling Expert Forecasting Knowledge for Incorporation into Expert Systems.

Abstract.

The use of continuous multivariate models to represent experts' knowledge of relations among a set of variables is reviewed. Such knowledge models can be incorporated in expert systems, complementing contingent rules, especially when representing experts' knowledge of functional relations among entities in uncertain domains. Past work has most commonly involved linear averaging models in static domains, although nonlinear models and dynamic domains are also possible. Detecting errors in continuous multivariate models requires a different approach than detecting errors in collections of if-then rules. Methods for eliciting expert knowledge include modeling judgments made in real or hypothetical situations, and use of expert's self-insight to directly assist in construction of the model. Procedures for managing each of these methods have been computerized and could be included as elicitation tools in expert system building environments.

Contents

Modeling Expert Forecasting Knowledge for Incorporation into Expert Systems.

1 Introduction.

This paper will review the use of continuous multivariate models to represent experts' knowledge of relations among a set of variables, because such models could usefully complement the contingent rules commonly used to represent expert knowledge. Computer systems that embody experts' knowledge have been produced for nearly two decades (Buchanan and Shortliffe, 1984), although the reliable extraction of knowledge continues to be a matter of concern (Bamber, 1990; Nazareth, 1989). Stewart and McMillan (1987) advocated the use of linear averaging models to represent judgments in expert systems. However, expert knowledge has generally been represented in these systems using contingent rules rather than continuous multivariate functions.

Experts' knowledge can take a variety of forms. Knowledge of what is can be captured in categories. Knowledge of the structural relations among categories can be expressed in spatial models (clusters and dimensions). Knowledge of what to do in specific situations can be captured using if-then rules. Finally, knowledge of causal connections, correlations, or constituent relations can be expressed using mathematical functions.

Computer programs that relate multiple variables using continuous mathematical functions are common, e.g., in simulations of power plant operations or in weather forecasting. The relations are usually based on the best scientific understanding, or on large data sets whose characteristics have been abstracted into models (Rouse, Hammer, and Lewis, 1989). But expressions of such relations could be useful even in the absence of an accepted scientific theory or a formal empirical analysis -- they can be based on experts' understanding (Fischhoff, 1989; Keeney and von Winterfeldt, 1989). That is, although it has not commonly been done, experts' knowledge can be represented in computer systems as continuous multivariate functions.

Experts can produce such models of their knowledge. For example, highway engineers who have worked with the capacity problem can describe the relation between lane width and vehicle bearing capacity in terms of a mathematical function with numerical parameters (Hammond, Hamm, Grassia, and Pearson, 1987). Such models of knowledge can be useful for expert forecasting systems. In the case of highway capacity, the relation has been analyzed empirically and is available in tables and formulas, so there is no need to extract such knowledge from experts. But in any given domain there might be questions that have not had the benefit of formal analysis, for which experts would be able to produce continuous multivariate models that express their knowledge of the relations in the domain.

Approaches similar to the methods of psychological research have been used to aid in the process of extracting expert knowledge. These include: open ended and directed interviews, to learn about all aspects of the expert's knowledge; the analysis of the expert's thoughts recorded during problem solving, to discover the expert's concepts and how they are used; methods for discovering the elements in the domain the expert knows (Rep method; Boose, 1988), and the key dimensions by which these elements are organized; similarity based judgment methods (multidimensional scaling, cluster analysis, pathfinding; Cooke and McDonald, 1987), for eliciting and representing the core components of knowledge; and various structured judgment or decision exercises, for detailed elucidation of particular

distinctions. These techniques have been reviewed by Cleal and Heaton (1988, Chapter 7) and Hoffman (1989). In addition, Hoffman (1987) and Burton, Shadbolt, Hedgecock, and Rugg (1988) are examples of empirical comparisons among different approaches. Complementing the above work, this paper will focus on approaches for representing experts' knowledge of continuous relations among a set of variables.

The purpose of this paper is to review the variety of such models, the methods that have been developed for using experts' knowledge to produce the structure and the parameters of such models, and the possibility of computer tools for managing the extraction of such knowledge for the purpose of building expert systems. Section 2 illustrates issues that can arise when continuous multivariate models are integrated within expert systems. Section 3 distinguishes the types of multivariate model likely to be most useful. Section 4 compares accuracy and error checking procedures in conditional rule models versus continuous multivariate models. Section 5 discusses methods that can be used with an expert to produce such models of his or her knowledge. Section 6 suggests incorporating existing computer tools for eliciting such models from experts into expert system building environments.

2 Integration of multivariate knowledge models into expert systems.

As an illustration, consider a formula for predicting whether a hurt ankle is actually broken (Diehr, Highley, Dehkordi, Wood, Krueger, Teitz, and Hermanson, 1988). The formula was derived by using linear discriminant analysis on 36 symptoms and signs for 587 ankle trauma patients. It takes the form of an index:

Index = -85 + 4.3*Age + 154*Abnormal_Color + 95*Tender_Bone
        + 121*Achilles + 74*Low_Pulse - 49*Previous_Sprain

where *Abnormal_Color* means (is "1" if) the skin color is pale or cyanotic, etc. Most of these predictors are yes/no but could as easily have several levels or be continuous. The model produces a continuous output from these inputs: that is, it can produce any value (within the possible range) on the outcome dimension. One would order an X-ray only if this index exceeded a given threshold.

Diehr et al's (1988) formula could be integrated into an expert system. It could be turned into a single purpose doctor's decision aid by programming a front end that asks for the required inputs, or into a nurse's assistant by programming a series of questions that seek to identify when its application is appropriate. In the first case, the doctor would know when it was appropriate to use the aid, while in the latter case the doctor's knowledge concerning the appropriateness of using the formula would also be incorporated in the program, as a set of conditional rules. Alternatively this model could be just one of several embedded in a larger expert system, such as one offering guidance on X-ray use when any limb in the body is injured. In such a larger expert system, there might be several models similar to that of Diehr et al (1988), each applying a formula with multiple inputs. The application of these formulas would be governed by a rule-based control structure. The source of the models could be scientific research, or expert judgment if the research were not available.

A continuous multivariate model of an expert's knowledge is defined only for a particular, well specified domain. This can be considered its "scope": its input variables, its output, and the range of the input variables (and other non-modeled conditions) for which it holds. Common sense would prevent someone from applying the broken ankle model when a broken

knee is suspected, or when ankles are swollen due to heart failure. When such a model is included in a working expert system, care should be taken that it is applied only within its intended scope. That is, common sense needs to be programmed explicitly.

Boundary cases also need attention. Treatment of boundary cases includes consideration of (a) cutoffs: whether a model should be applied to cases that are at the edge of the zone for which it is defined, (b) overlap: what to do when two different models could each be applied to a case (analogous to rule conflict in rule-based systems), and (c) gaps: how to assure that there are no holes, i.e., cases for which no model has been provided. For example, in the integrated broken bone decision aid described above, there might be ambiguous cases where either the "ankle" or the "foot" model might apply. Ideally, the two models would make similar recommendations about whether to X-ray when applied to a boundary case; if not, the expert should know why the models make different recommendations, and give guidance governing the boundary cases that can be incorporated into expert system's rules for controlling model application. There is also the issue of what should be done when information is lacking concerning one or a few of the input variables: should the model not be applied, or should reasonable default assumptions be used?

3 Distinctions among types of continuous multivariate models of expert knowledge.

3.1 Functional versus structural models.

Functional models describe observed relations between input variables and output variables. *Structural* models are derived from theoretical knowledge or scientific understanding of the domain. The formal procedures for predicting highway capacity or broken ankles, described above, are examples of functional continuous multivariate models. A program for calculating the energy flow of a house would be an example of a structural continuous multivariate model. Although the transmission and conduction rates of individual materials would be empirically determined, the overall relations (e.g., that the amount of heat lost through a particular surface is a function of temperature differential, surface area, and heat conduction rate) would be theoretically based.

In principle, an expert could supply subjectively judged continuous multivariate models that are either functional or structural. Practically, if the model is structural the parameters are probably available already and do not need to be judged by the expert. Therefore it is functional models that experts would most often be asked to supply.

The functional/structural distinction is related to Steels' (1987) idea that expert systems are now capable of both surface reasoning (in terms of the situation's categories and the available actions) and deep reasoning (in terms of a model of the principles governing relations in the problem domain). An advantage of the deep models is the possibility of deriving additional rules or relations at the surface level. Continuous multivariate models could be used in either type of reasoning. For example, given a deep or structural model $y = a + b_1 x_1 + b_2 x_2 + b_3 x_3$, if it were necessary to infer x_2 from y, x_1, and x_3, a deep reasoning expert system could manipulate the first model to produce the needed model.

3.2 Averaging versus non-averaging models.

In an *averaging* model the impacts of all the input dimensions are averaged to produce the output measure. The inputs can be linear (e.g., $y = a + b_1 * x_1 + b_2 * x_2$) or nonlinear (e.g., $y = a + b_1 * x_1 + b_2 * x_2 + b_3 * x_2^2 + b_4 * x_2^3$). Alternative ways of combining the input dimensions include multiplication (e.g., $y = a * x_1 * x_2$) and other combinations of mathematical operations (e.g., $y = x_1 ** x_2$), as well as contingencies (e.g., if $x_1 < 1$, $y = b_1 * x_1$; if $x_1 > 1$, $y = b_2 * x_2$).[1] Advantages of averaging models are: they are flexible, that is, able to describe most relations; their behavior is robust to errors in the parameters; high values on one input dimension can compensate for low values on another; and they are easy for an expert to understand.

A separate question is what kinds of model best describe the human judgment process (N.H. Anderson, 1981; N.H. Anderson and Zalinski, 1988; B.F. Anderson, Deane, Hammond, McClelland, and Shanteau, 1981). It may be argued, for example, that those models most closely related to the thought processes that happen to be used in a domain are also easiest for people to use in expressing their knowledge of that domain (Hammond, 1982). If the thought processes typically used in a domain were known, it would provide a basis for selecting the form of model to be used in expressing an expert's knowledge.

3.3 Linear versus nonlinear averaging models.

There has been considerable discussion of which averaging models are best for expressing experts' knowledge (especially for multiattribute evaluation), and of the attendant methods for extracting the knowledge. Dawes (1979) advocates the linear averaging model: identify the predictors, establish measurements for each, combine in a weighted average (or a multilinear form; Keeney and Raiffa, 1976, pp. 288-297). This has several advantages. Linear models often produce surprisingly accurate descriptions even when people expect that nonlinear models will be required (Levi, 1989; Lusk, Stewart, Hammond, and Potts, 1990). They are simple to program. Whether the models' parameters are fitted to empirical data or by judgment, it is easy to explain them to the expert.

Models with nonlinear combinations of the input variables are also useful. There are a large variety of nonlinear models, including nonlinear averaging models, as well as methods for fitting them (Rouse, Hammer, and Lewis, 1989). These models require the same steps as linear models: identify the elements, characterize the relations and select a model form capable of handling these, then specify the particular numerical coefficients in the model. The difference is that the mathematics are more complex: it is difficult for most people to think about the specific nonlinear configurations of input variables that are to be averaged, and hence it is challenging for the expert to judge the parameters and produce the models. A wrong guess at a parameter might make the model behave very differently from what was intended.

There are different problems for which one or the other of these approaches is more appropriate (Hammond, 1983), based on the characteristics of the relations among the variables as well as the abilities of the experts. The boundaries between them are a matter of argument. This paper will focus on linear averaging models. If a persuasive case can not be made for their utility as a vehicle for expressing expert knowledge for expert systems, then nonlinear averaging models and the general class of nonlinear models are not likely to be found useful for that purpose. The paper will not consider the issue of recognizing what type

---

1. The distinction between averaging and multiplying models is not sharp: for example, there can be averages of products, etc.: $y = b_1 * x_1 + b_2 * x_2 + b_3 * x_1 * x_2$.

of model is appropriate in what situation (Leddo, Cohen, O'Connor, Bresnick, and Marvin, in preparation), which Bamber (1990) suggests is another problem for expert judgment.

3.4 Dynamic systems.

A more complicated use of multivariate continuous models in computer systems is dynamic models, in which there are feedback connections from the output variables of one equation to the input variables of another (Forrester, 1961). Sterman (1989) and Kleinmuntz (1990) have shown that people's understanding of the behavior of these systems can be quite inaccurate. Although the accuracy of experts' understanding of these systems has not been widely studied, the extraction of experts' knowledge of parameters of the defining models for dynamic systems does not seem promising at this time.

4 Comparison of if-then rules and continuous multivariate functions as representations of knowledge.

The kind of relation easily expressed by continuous multivariate models, including the linear averaging functional models focussed on here, is different from the kind of relation most easily expressed by if-then rules, as in the rule based production system architectures commonly used for programming expert systems. On the one hand, this is just a question of programming language, and the current languages are flexible enough to express any desired sort of relation. But the rule versus function distinction can be found (a) between types of relation in the domain, (b) between the corresponding types of knowledge that an expert might have, and (c) between the kinds of judgment needed to express the knowledge in a form usable by the program, and so it is worth a little more exploration.

Consider the production-system type rule "if $x + 2*y > 7$, then $z = 2$". Although the inequality on the left hand side includes a continuous expression, the rule has a categorical output. In contrast, in the following rule the output is a continuous function of the input variables: "If there exists information on x, and there exists information on y, then $z = 5*x + 2*y$." Here the output or the effect of executing the rule, z, is a continuous function of the inputs, x and y.

In some domains (such as the highway capacity and house energy flow examples, above) the relations among the concepts may be more appropriately expressed using continuous functions rather than collections of categorical contingencies. An income tax preparation aid, where steps are taken contingent on qualifying for particular exemption categories, is an example of a domain appropriately addressed with rules.

Uncertain connections between the inputs and outputs in a relation present very different problems for categorical rules and for continuous mathematical functions. Uncertainty makes the application of a contingent rule "wrong" some proportion of the time; hence the path the program takes following the use of that rule may also be wrong. One solution has been to address the uncertainty explicitly using probabilities (Cheeseman, 1985; Pearl, 1986) or degrees of confirmation (Chapters 10 and 11 of Buchanan and Shortliffe, 1984). Functional relations can handle uncertainty by producing a best guess plus error. A slight error in inputs produces a slight error in output. Experts asked to express their knowledge of uncertain domains in terms of one of these formalisms may find it easier to make judgments in terms of weighted averages rather than precise contingencies (Hammond, 1982).

If continuous multivariate models are appropriate for the domain, programming them as a mathematical expression is probably more efficient and more accurate than approximating the desired relations in steps using a coordinated set of categorical rules. Although rules with categorical outcomes can model such domains to arbitrarily any degree of accuracy, the higher the desired accuracy the more rules are required. Nonetheless, stepwise approximations have been used with satisfactory results, including in the calculation of highway capacity. Expressions of continuous relations are considered here primarily because they may provide better vehicles for expressing experts' knowledge than collections of rules would, not because of programming efficiency.

No matter what form of model is most appropriate for representing the domain, there is also the question of how well the expert can express his or her knowledge using this type of model. Expressing a relation in the form of a mathematical expression is a very different process than expressing it in terms of a collection of rules with categorical outputs. While both forms require the identification of the key dimensions, the mathematical functions require the judgment of numerical parameters that express general relations between variables, while contingent rules require the specification of particular combinations of levels of inputs, and the corresponding outputs. Therefore when deciding whether to use continuous multivariate representations of experts' knowledge of the functional relationships in a domain, one should consider not only programming efficiency and the appropriateness of the model type to the domain, but also the expert's ability to make the judgments that the model requires.

4.1 Accuracy of representations based on conditional rules versus continuous multivariate models.

There have been a limited number of comparisons of the accuracy of conditional rule versus weighted averaging representations of expert knowledge. Einhorn, Kleinmuntz, and Kleinmuntz (1979) studied two tasks. For each they produced conditional rule models based on analysis of verbal protocols, as well as linear models based on judgments of cases. The linear model of a clinical psychologist's judgments of the mental health of a set of college students described the psychologist's judgments better than the conditional rule model. In the second comparison, there was little difference in accuracy between the methods. A third comparison was provided by Larcker and Lessig (1983). For 25 of 31 subjects, the sets of rules derived from a guided retrospective process tracing procedure predicted the subject's judgments somewhat better than did a cross validated linear model. That rules were more accurate in one domain, averages in another, is perhaps a reflection of the different task domains.

In the domain of weather forecasting, a number of comparisons have been made between the performance of expert weather forecasters, linear models of the experts, and conditional rule based expert systems (Stewart, Moninger, Grassia, Brady, and Merram, 1988; Moninger, Flueck, Lusk, and Roberts, 1989). The expert systems and the linear models were not based on the same experts. Generally, the linear models of the experts and the rule based expert systems performed at about the same level, and were exceeded by only the best of the weather forecasters.

4.2 Error checking in rule-based and continuous multivariate models of knowledge.

Another contrast between representations of expert knowledge that use conditional rules versus continuous multivariate functions lies in the processes required to check for errors in the representation.

4.2.1 Checking conditional rule representations of expert knowledge.

In rule-based systems the knowledge representation is typically checked along with the programming, as a step in "a methodology of interviewing human experts and building prototype systems, incrementally redesigning, refining, and extending the prototype until it reaches the desired competence" (Gruber, 1989, p. 131). The expert helps correct the system and direct its expansion, by looking at the system's behavior when faced with each logically possible state of the world. Thus the expert is assumed capable of recognizing errors, and is asked to exhaustively review the possible combinations of conditions. Another aspect of checking, done by the programmer rather than the domain expert, is to determine whether rules are consistent and not redundant (Nazareth, 1989).

4.2.2 Checking multivariate functional models of expert knowledge.

Checking a continuous multivariate model of an expert's judgments about an uncertain domain requires a different approach. One can not simply select a new case, apply the model to it, compare the model's predictions with the expert's judgment, and adjust the model to accommodate for any observed mismatch. An individual case is not sufficient for evaluation of the multivariate model as a whole, nor does it provide an immediately applicable prescription for improving the model's accuracy. Unique features of the particular case, which influence the expert's judgment of the case, may not be captured by the general model of the expert's knowledge. Although the expert may disagree with the model's prediction in the one situation, the model may still provide the best general characterization of the expert's knowledge about a large set of situations. Adjustment of the model to increase its agreement with the expert's judgments in a test case might make it agree less in many other cases.

Because of noise in the domain and uncertainty in the expert's knowledge of the domain, predictions by a multivariate functional model of an expert's knowledge will have a component of error. Checking of expert judgment models must take this into account.[2] The expert could evaluate the model using judgment at a general level, asking how well the model expresses the relations he or she believes hold in the domain. Checking a model using general judgments would be particularly appropriate if such judgments had not been used in the original production of the model.

Another important process for checking the accuracy of continuous multivariate models of experts' knowledge is "checking the math". This involves assuring not only that there have been no slips in writing the numbers, but also that all procedures use numbers correctly according to the rules of measurement (see Krantz, Luce, Suppes, and Tversky, 1971; N.H. Anderson, 1981; Stewart, 1988). Usually such checking is the responsibility of the programmer rather than the expert. Methods that assure that the procedures adhere to these principles will be discussed below.

5 Methods for producing a model of an expert's judgment.

Over the years a number of methods have been developed for deriving continuous multivariate models, particularly linear averages, that represent people's knowledge about a domain. Each of these might be useful for a knowledge engineer modeling an expert's knowledge. The

---

2. Ravinder and Kleinmuntz (1991) and Fischer (1991) give general advice about improving the accuracy of additive decompositions of multiattribute utility that is also applicable to linear averaging models of expert judgment: spend the most effort improving the precision of parameters with the greatest impact.

methods to be discussed involve fitting models to experts' intuitive judgments of typical cases, versus asking the experts to define the models directly.

5.1 Observation of expert's judgments.

Many expert practitioners are problem solvers who deal on a daily basis with concrete situations (Schon, 1988), yet are uncomfortable with abstract academic characterizations of their domain. Hence they may be able to predict outcomes for cases described in terms of their essential features, but yet be unable to produce parameters to specify models of the relations between those features and the outcomes. For example, an expert emergency room physician may be quite comfortable judging whether a bone is broken, but may have no basis for expressing his or her knowledge by producing a parameter in a linear discriminant model such as the broken ankle index (see above).

A method for producing continuous multivariate models representing the knowledge of such an expert is to "bootstrap" the expert, that is, to use best-fit algorithms on a fairly large number of the expert's judgments to produce a model that predicts the judgments as a function of the elements of the objects judged (Hammond, McClelland, and Mumpower, 1980; Rouse, Hammer, and Lewis, 1989). These judgments may be produced in the daily practice of the expert's trade, or produced especially for the knowledge engineer.

5.1.1 Unobtrusive observation: Judgments from the expert's daily practice.

If the expert's judgments about a number of cases are available, these can be used to produce the model. The judged cases should be drawn from the class to which the model is intended to apply, and they should be present in the judged sample in representative proportions, else the model would not be valid for the situations in which it is to be applied.

Anand (1990) proposes such an approach for a medical context. He envisions an expert system that would have access to a large data base of patient histories. These histories contain both facts about the patients (inputs to the judgment model) and the doctors' diagnoses (the output). Anand's (1990) model could produce probability estimates for diagnoses in a new case, based on the previous doctors' diagnoses of the old cases. This is notable because the basis for the judgment model (the data base) is a part of the same computer system that includes the expert system, and as knowledge or experience with a type of disease is accumulated the expert system could update its judgment model automatically.

A system such as Anand's (1990) would treat a number of experts as if they have one judgment policy. The resulting model can be said to describe the judgment practice of the institution which the individuals represent if the individuals participate in the study proportionately to their responsibility for the institution's decisions (Bursztajn, Gutheil, Hamm, Brodsky, and Mills, 1988). However, if judged cases were contributed by members of two camps with different judgment policies, the model would be some sort of average of the two policies, possibly less valid than the policy of either camp. If there were an advance in treatment or a shift in the characteristics of the disease it would be necessary to modify the program else, being based on a large number of prior cases, its recommendations would be out of date. Any expert system would need such revision; the problem here is that people might rely on the automatic updating, and not recognize that it would not handle discontinuous changes in knowledge.

The biggest drawback to using existing databases of experts' judgments to infer judgment models is that in most domains it is not possible to find adequate records of a collection of similar cases. Case information is usually not complete, and data bases do not include all the

inputs a model may require.[3] Further, case information may not be objective, as in the adversary legal system, or in social work where case histories are often "reconstructed" after a resolution rather than being recorded as they happen (L. Dalgleish, personal communication).

An alternative to using existing databases of experts' case judgments is to make fresh observations. However, this may take a long time and the observations may not cover the whole range of situations needed for the model.

5.1.2 Systematic observation: Expert's judgments on special tasks.

Having volunteer experts judge a set of hypothetical cases can provide the needed judgments relatively quickly.[4] Experts would be asked to respond to the hypothetical situations as if they were encountered in their daily practice. The cases can be described with the inputs in "raw" form (as they would be encountered in daily practice), or pre-measured.

Judging hypothetical cases in such controlled tasks allows for uniform conditions of judgment. The characteristics of the set of cases judged can be systematically manipulated for various purposes. A uniform or representative distribution of values over each dimension can be provided. The dimensions can be made to vary independently to allow for precise specification of the model. Alternatively, the dimensions can be made to have intercorrelations characteristic of the domain so that the expert experiences the task as realistic (Hammond, McClelland, and Mumpower, 1980). Judgment of cases allows fitting of individuals' judgment policies; if different policies are found for different individuals, some basis must then be found for selecting among them or combining them.

An advantage of having experts judge individual cases rather than provide parameters for a model is that the variety of cases may serve as a fairly thorough set of cues for stimulating the expert's memory of pertinent knowledge (see Bamber, 1990). In contrast, asking the expert to give a model describing his or her knowledge (see below) induces an abstract consideration of the problem which may access less of the expert's knowledge.

The paradox of using controlled procedures for observing experts' judgments is that, although the procedures aim to eliminate extraneous sources of variation in the judgments, they can introduce distortions. Even high status experts may judge differently when they know their performance is being watched. And people develop efficiencies when judging repeated cases of the same type: they do it faster, more consistently, and possibly more simply than when judging real life cases (Slovic, Lichtenstein, and Edwards, 1975).

The choice whether to present case information to the expert in abstract or "raw" form depends on several factors. Experts can often make quicker judgments of situations in which the inputs have already been processed, without loss of accuracy. For example, in a study of highway engineers' judgments of highway characteristics (safety, capacity, aesthetics), two methods were used to develop "bootstrapped" models of each engineer's judgments. In one procedure, the engineer looked at pictures of the highway (a moving film strip, consisting of photographs taken every 50 feet) and perceived the pertinent features himself. In the other procedure, the relevant features were measured numerically and presented graphically in the form of a profile (8 to 10 features simultaneously displayed as bar graphs). Judgments were completed more quickly with the abstract bar graph display (Hammond, et al, 1987). When the engineers'

---

3. An interesting exception is bank officers' judgments of credit risk: a West German bank had computer records of hundreds of thousands of cases, which permitted derivation of judgment models (Roland Sholz, personal communication, June, 1990).

4. The similar judgment tasks used with the Rep method (Boose, 1988) and psychological scaling (Cooke and McDonald, 1987) serve a different purpose — to discover the structure of the elements in the domain, rather than to elicit the causal or contributory relations among those elements.

judgments were evaluated with respect to objective standards, the judgments of the bar graph information were more accurate for two of the characteristics (aesthetics and capacity), but the film strips produced more accurate safety judgments (Hamm, in press).

Recent work comparing raw versus premeasured information about cases has shown a surprising level of disagreement at the perceptual measurement stage among atmospheric scientists (Lusk and Hammond, 1991). In producing an expert system, one might wish to decompose the judgment task (see Stewart, 1991) to extract judgment policies in areas of agreement, and to isolate areas of disagreement for further study.

5.2 Use of expert's self insight to guide construction of model.

Judging a large number of cases to provide data for an algorithmic model-fitting process is laborious,[5] and an available expert may not be willing to do it. Another approach is to get the expert's direct assistance in building a continuous multivariate model. Steps in this process could include: determining the areas of knowledge for which such a model is appropriate (Bamber, 1990), selecting the variables that should be included as input in the model, determining the structure of the relations between input and output variables (e.g., linear averages), and specifying numerical coefficients for the relations in the vocabulary of the particular model (Rouse, Hammer, and Lewis, 1989). The expert could also help by evaluating the model (Fischhoff, 1989).

The various possible approaches differ in who has control over the basic elements of the process: selecting the variables, determining the basic mathematical structure of the model, and producing the particular parameters for a given model form. The expert may be offered full control over the form of the model, or at the other extreme may have responsibility only for naming parameters in a model whose variables and organizing principles have been selected by the knowledge engineer.

5.2.1 Basing model elements on the scientific literature.

Hammond, Anderson, Sutherland, and Marvin (1984) used experts to select the variables to be included in a model for forecasting the health effects of a nuclear bomb manufacturing plant on its neighbors. Five experts with different public stances on the issue of the safety of the Rocky Flats plant participated in a several-stage process that was managed by researchers who visited the experts individually (to avoid arguments) and carried each expert's recommendations to the others. The experts defined the scope of the problem (the effects of inhaled plutonium dust, as modified by the external factor of smoking), selected the relevant variables, determined the form of the model, and agreed on the parameters for that form of model. The rules imposed on the experts included that if there was disagreement about the variables, model forms, or parameters, they would support their positions with arguments based on research that was available in the literature. Differences supported by published research were resolved by going back and forth until the parties agreed on a parameter.

Scientific backing was required by Hammond et al (1984) as a means to resolve disagreements among experts. But there are domains in which scientific research has not been done or has not produced an accepted or unified theory, and the best available knowledge is therefore expert judgment. In such domains, the experts might need to construct models of their knowledge from scratch, rather than modify accepted models.

---

5. The number of cases required may be as large as (N + 1) * (M + 1) - 1, for N cases described on M attributes (N.H. Anderson and Zalinsky, 1988).

5.2.2 Expert's unaided construction of the model.

Experts often are well qualified to produce models that express the joint continuous influences of multiple input variables upon an output variable. Their experience sensitizes them to the important factors, the effects of each factor, and the interactions among the factors. Their training and everyday practice may involve the use of mathematical expressions describing these relations, so that they are competent to construct a model describing their knowledge.

Giving experts responsibility for producing the model of their knowledge respects their expertise and hence may motivate careful work. However, there is the possibility of error. The expert may misunderstand some aspect of the domain, and the knowledge engineer would be unlikely to catch this error (unlike in the previous approach where reference to scientific authority and mutual criticism among experts served to prevent errors). The expert might omit a variable from the model that he or she would easily recall if the situation actually arose (Fischhoff, Slovic, and Lichtenstein, 1978). There is a "data type" problem -- that a variable may be defined differently in the formula the expert hands over to the knowledge engineer than it is in the rest of the expert system. And there is the possibility of misconstructing the model from the point of view of measurement theory -- performing unjustified operations on variables, such as adding a constant to a ratio scale variable or taking ratios of interval level variables. Finally, experts can "slip", e.g., make a clerical error or forget a minus sign (cf Norman, 1981).

The importance of experts' error in expressing their knowledge is illustrated by the experience of Hammond et al (1987). When highway engineers were asked to produce formulas representing the influences of a number of highway features upon highway aesthetics, safety, or capacity, over a third of the formulas contained errors that the researchers could discover later, and correction of these errors increased the models' average accuracy by up to 15% (Hamm, in press). To protect against such error, the expert should be involved in the process long enough to catch mistakes, i.e., slips between what he or she intended the model to do and what it does, and the knowledge engineer should look for inconsistency between the model and the expert's expressed intentions.

The highway engineer study also provided comparisons between the models that the expert produced directly and those that were produced by modeling the expert's judgments of a number of individual highways. The results were different for each task. After the engineers' slips had been corrected, the models that they wrote for highway safety predicted true highway accident rate less accurately than the models fit to their judgments of the safety of individual highways. The two types of aesthetics model were equally accurate. For predicting highway capacity, the models the engineers wrote were significantly more accurate than the models based on the engineers' capacity judgments based on *film strip* displays. However, the models bootstrapped from the engineers' capacity judgments from *bar graph* displays were more accurate than the models the engineers wrote (Hamm, in press).

The efficacy of these alternative methods may depend upon the availability of the experts' knowledge of the domain. If they "know what they know", they may be able to write it directly. On the other hand, if they are not conscious of all the information that they use when thinking about the domain, it may be better to have them make judgments about a systematically produced set of cases, and then to use statistical procedures to derive the relations they use in their judgments.

Related to this is the question of diagnostic versus predictive relationships. In domains where there are negative feedback relations, i.e., compensatory processes, the predictive relation that a cue (input) has to an output may be the opposite from the diagnostic relation between the two variables. For example, when considering the effects of variations in highway layout upon

highway safety, wider lanes *predict* safer highways because wide lanes give more room to maneuver. But highway departments have known this for years, and have spent their limited budgets in making wide lanes in locations where they are most needed. Therefore, wide lanes in currently existing highways *diagnose* that there was a need for wide lanes; hence the wider the lane, the more dangerous the highway. In the highway engineer study, engineers used the "lane width" cue diagnostically when judging individual highways from film strip displays, but they used it predictively when writing formulas (Hamm, in press).

5.2.3 Knowledge engineer guides the expert's construction.

If the expert can not accurately use mathematical language to express his or her knowledge of the relations among the elements in the domain, the knowledge engineer may need to provide structured guidance to assure that the model has the intended meaning.

5.2.3.1. Provision of structure. In one form of guided expert model building, the knowledge engineer determines the variables to include in the model and the structure of the relations among them (usually a linear averaging model). The expert supplies only the parameters in the model, specifying the tradeoffs among the input factors. This approach might be used when the knowledge engineer has produced a model with one expert, and wants additional input from another expert without having to start from scratch. The second expert is then asked only to judge the parameters for the model. The "simple multi-attribute rating technique" (Edwards and Newman, 1982) also takes this approach when eliciting value tradeoffs from citizens, because information is needed from many citizens using the same model form. When expert bank officers used similar techniques to produce models of their decisions to offer credit to loan applicants, the models were more accurate than models based on the same individuals' judgments of hypothetical loan applicants (Stillwell, Barron, and Edwards, 1983).

A drawback of requiring the expert to produce only the numerical parameters for a predefined model is that the expert may not make the numerical judgments carefully. Not "owning" the model, the expert may not be motivated to ensure that the parameters accurately reflect the tradeoffs he or she understands to be true. And not having been involved in selecting the basic form for the model, even the well motivated expert may not understand the meaning of the parameters enough to judge them accurately.

A number of studies have cast doubt upon the accuracy of people's direct judgments of parameters for linear averaging models intended to express their judgments in continuous multivariate domains. Work by Gabrielli and von Winterfeldt (1978; see von Winterfeldt and Edwards, 1986, p. 368) and Stewart and Ely (1984) has shown that when people name weights expressing the input dimensions' relative importance in influencing the output dimension, they are not very responsive to the range over which the input dimensions vary, even though this is crucial to the weights' interpretation. This can occur even when the relation of the range information to the parameters has been explained carefully (see also Anderson and Zalinsky, 1988; Goldstein and Beattie, in press). People are inaccurate at summarizing their judgments in terms of the relative weights they put on different factors (Roose and Doherty, 1976), although they can recognize weight-profile descriptions of their own judgment policies (Reilly and Doherty, 1989). Also, weights elicited for all input dimensions simultaneously are "flatter" than weights elicited for the same dimensions organized as a hierarchical tree (Stillwell, von Winterfeldt, and John, 1987), and a factor's prominence in a hierarchical tree and the amount of detail used to describe it can increase the numerical weight that people assign to it (Borcherding and von Winterfeldt, 1988; Weber, Eisenfuhr, and von Winterfeldt, 1988).

Experts' difficulty expressing their knowledge using numerical parameters may be due to the fact that the language of multivariate models may not correspond to how they think about the

domain (Fischhoff, 1989). If the terms of the model and the operations in the model building process are unfamiliar, experts may treat the exercise of naming numbers as something they must go along with but that has no important consequences. Allowing the experts to have more of a role in defining what is modeled and how it is measured may produce models in which their judgments are more accurate.

5.2.3.2. Guiding the model building process to assure adherence to principles of measurement. The alternative to eliciting the expert's judgments of parameters within a prestructured model is to involve the expert in the structuring of the model. The knowledge engineer can guide the expert to assure that (a) the terms of the model are consistent with the rest of the expert system, (b) numbers used in the model have the required scale properties, and (c) only legal operations are applied to the numbers (Krantz, Luce, Suppes, and Tversky, 1971; Luce and Krumhansl, 1988). In effect, the knowledge engineer would be performing the same function as a decision analyst (see Keeney, 1977) or judgment analyst (see Stewart, 1988), assuring that the numerical model says what the expert means. The process would be similar to the structuring of a value tree (von Winterfeldt and Edwards, 1986, Chapter 2), except it deals with experts' general knowledge of the relations in a domain, rather than their knowledge of the multiple aspects that contribute specifically to value in the situation.

I will now sketch an approach to working with the expert to structure a model of his or her knowledge of the relations in a domain. This involves working from the top down and from the bottom up in alternation. The approach was used with a subset of the engineers in the Hammond et al (1987) study. The steps are summarized in Table 1, and presented in detail in Hamm (1990). From the top or the general perspective, the scope of the model is defined and the relevant variables are selected. From the bottom, the numerical scales for measuring each variable are defined, assuring consistency with the rest of the system and establishing constraints on the operations in which the variables are to be involved. The expert also decides whether the effects of any input variables on the formula's output variable are influenced by any of the other variables, i.e., whether the input variables interact in influencing the output variable.

****

Insert Table 1 about here.

****

From the top again, the variables are grouped into subsets that logically go with one another. Interacting variables are included within the same group so that the relations between groups are as simple as possible. (McLeish and Cecile, 1990, have used a similar heuristic approach for producing models with conditional independence.) Then the groups, as well as ungrouped individual variables, are arranged into a hierarchical tree structure, in which the model's output variable is the top node and the branches from each node [to lower variable-groups (nodes) or to individual variables (terminal nodes)] represent the expert's way of thinking about what influences or constitutes that node. The tree can be arbitrarily deep, depending on the number of variables and on how the expert wants to structure it.

After the model tree is structured, the expert and the knowledge engineer work bottom up again, to specify the mathematical form for combining the branches at each node of the tree. The knowledge engineer assures that the operations used are appropriate for the scales of the input variables (terminal nodes), and helps the expert define the scale type of the output of each node, so that the next higher level can be built appropriately. Once the formulas describing all the branch points in the tree have been specified, the total model can be written simply by substituting the lower nodes' formulas into the formula for the top node.

This guided model building approach, in which the knowledge engineer constructs the formula with input from the expert at all key decision points, preserves several advantages of the methods discussed above. As when the expert builds the model without guidance, the expert is involved here at every step and so has deep understanding of the model, as well as the motivation to think hard that comes from being involved. As when models are fit to experts' case judgments, the parameters in the models will be relatively valid because the structure of the procedure and the knowledge engineer's involvement make it easy to prevent slips and errors.

Guided model building has several possible disadvantages. Writing a formula offers limited memory cuing compared to judging hypothetical cases. Success depends on the expert being able to express his or her domain knowledge when making judgments about the following model options: the existence and form of interactions; the appropriateness of various model forms (e.g., multiplicative or additive combination); the relative importance of different factors involved in an averaging formula (weights) or a multiplying formula (powers); and the functional form of the relation between two variables.

The above procedure was used with 3 engineers in the Hammond et al (1987) study to build models, embodied in formulas, to express their knowledge of the effects of the offered input variables (highway features) on the characteristics of highways (safety, capacity, or aesthetics). The other 18 engineers built models unaided. All engineers were given the variables that were available for use in the model, with a definition of the variable and a metric for measuring it. The models resulting from the guided procedure were no more accurate (measuring accuracy as the correlation between the model's prediction and an objective criterion measure) than the models the engineers produced without guidance, after the unguided models were corrected for obvious slips (Hamm, in press).

Although the number of subjects using the guided procedure was small, it is remarkable that fixing dumb errors, rather than using the formal method to assure adequate measurement, made the biggest difference in model accuracy. However, such a guided procedure may be more useful in other situations. For example, engineers may have more mathematics skill than other experts and hence benefit less from the procedure. The guidance may be less crucial in situations where linear averaging models are appropriate, due to the robustness of these models to parameter variation.

The guided procedure produced models that were better calibrated, which is not reflected in the use of correlation as a measure of accuracy (Hamm, in press). Also, the guided formulas were more complex. This suggests that the complexity of a model that an expert produces may be influenced by motivation or the availability of model producing tools or assistance. Politzer (1991) suggests that in complex situations complex models produce unexpected (but accurate) predictions. If so, a structured guidance procedure would help produce complex models in those situations where they are needed.

5.3 Impact judgments.

A hybrid method asks the expert to consider one or a few cases and report the impact of each of the particular inputs on the output in each case (Hamm, Bursztajn, Mills, Appelbaum, and Gutheil, 1981; similar ideas were explored by Zhu and Anderson, 1991). In combination with separate information about the measures of each case on all input variables and about the level of each input variable that would have a "neutral" impact, the expert's impact judgments can be used to produce estimates of parameters in linear averaging models. This has three advantages. The numerical "impact" judgments have very specific referents and hence are unlikely to be governed by generic judgments of factor importance that ignore the range over

which the factor varies (Stewart and Ely, 1984). The experts think about actual cases rather than thinking at an abstract level. And the expert considers only a few cases, rather than a large number. Although research has not been done to assess the quality of experts' impact judgments in comparison with judgment and self-report methods, preliminary results[6] suggest that in the absence of direct supervision of the expert, the parameter estimates derived with this technique may be quite noisy.

6 The possibility of computer tools for expressing expert knowledge as multivariate models.

Computer tools have been developed to assist knowledge engineers in extracting expert knowledge -- interactive programs that actively elicit knowledge from experts and produce the appropriate code (Gruber, 1989, pp 137-145). Different modes of representing knowledge require different elicitation tools, so a repertoire of model-based elicitation tools is required (Gruber, 1989; Cooke and McDonald, 1987). It would be useful for such a repertoire to include tools for eliciting continuous multivariate models from experts. This section reviews existing or potential computer procedures for managing the various extraction processes reviewed in Section 5, that could easily be incorporated into the repertoire.

For discovering relevant entities and the similarity relations among them, the AQUINAS computer tool uses the Rep procedure for eliciting personal constructs (Bradshaw and Boose, 1990). The structure of causal relations among relevant concepts could be elicited using a computerized version of the procedure in Table 1, above, or of the procedures reviewed by Keller and Ho (1988). If the causal relations are circular, they could be represented as a dynamic system using I-THINK or STELLA (Richmond, Peterson, and Vescuso, 1987).

Once the structure of connections among the pertinent concepts has been identified, estimates of the parameters are needed. If this is to be done by having the expert judge many concrete situations and fitting models to those judgments, the POLICY program (Rohrbaugh and Schuman, 1988), the CONJOINT procedure in SPSS (SPSS, Inc., 1990), or the AVERAGE program (Zalinski and Anderson, 1986) could be used as a model for the needed procedures.

If the expert is to directly estimate the parameters in the model, Garthwaite and Dickey's (1991) approach to eliciting experts' estimates of regression coefficients could be used if the experts are competent to judge probabilistic bounds on the parameters. Heuristic procedures requiring ratio judgments of factors' relative importance are offered by Forman, Saaty, Seeley, and Whittaker's EXPERT CHOICE (but see Dyer, 1990). For multivariate models of value or utility, there are many decision analysis programs that help experts structure the domain and produce subjective tradeoff parameters, which could be adapted to the more general purpose addressed here (e.g., McNamee and Celona, 1987). White (1990) reviews progress in integrating these procedures with expert systems (e.g., AXOTL of Bradshaw and Boose, 1990).

Several computer programs build continuous multivariate models of the sort discussed here in a very flexible manner, using linear programming to extrapolate from experts' imprecise judgments of parameters (Cohen, Laskey, and Tolcott, 1987; Moskowitz, Wong, and Chu, 1989; White, Sage, and Dozono, 1984). For example, the user could make vague judgments of model parameters (e.g., saying that the parameter for input variable A is more than twice the parameter for variable C), and the program would identify the range of possible parameters

---

6. Work in progress by Hamm, Gutheil, and Bursztajn.

for the model, consistent with the judgments made so far. These programs are written to produce multiattribute utility models: the output measure is an object's desirability. However, the programs' internal rules, which enforce consistency with measurement theory principles, could easily be adapted for a more general class of models. As such, they would offer a flexible environment for rapidly building a calibrated model of the expert's knowledge, using judgments as precise as the expert is willing to make.

Within such an environment, it would be easy to alternate between "top-down" and "bottom-up" perspectives as recommended in Section 5.2.3.2, above. Narula and Weistroffer (1989) describe a similar procedure that provides feedback following the expert's judgment of parameters in dynamic system models, allowing subjective evaluation of the adequacy of the set of parameters. STELLA could similarly manage the production and simulation of dynamic system models in which nodes are governed by multivariate models with judged parameters. Tatman and Shachter (1990) provide an environment for dynamically managing influence diagrams that could serve the same function.

In sum, procedures for constructing continuous multivariate models when appropriate have long been known and computer tools are now available for applying these procedures. Any of these could be included in the repertoire of elicitation tools provided by an expert system building shell.

7 Conclusions.

This paper has argued that continuous multivariate models, especially linear averages, are an appropriate form of representation for experts' knowledge of functional relations among multiple elements in uncertain domains. Their differences from conditional rule representations -- in the kinds of judgment required of the experts, in the way they handle uncertainty, and in the processes required for checking for errors -- offer a basis for choosing between the two forms of representation, even though no consistent performance advantage has been demonstrated for either form.

Choice of method for extracting experts' knowledge into continuous multivariate models for use in expert systems depends on a number of factors, including the availability of experts and the types of judgment task that they are willing or able to engage in. When experts have sophisticated insight into what they know, they can help model the knowledge directly, although special procedures may be needed to assure the quality of the model. When experts lack this insight or when expert judgments about many cases in the domain are available, models can be produced from statistical analysis of the relation of the judgments to case features. Procedures for managing each of these methods have been computerized and could be included as elicitation tools in expert system building environments.

8 References.

Anand, P. (1990). Towards second generation expert systems. *International Journal of Systems Research and Information Science, 4*, 81-85.

Anderson, B.F., Deane, D.H., Hammond, K.R., McClelland, G.H., and Shanteau, J.C. (1981). *Concepts in Judgment and Decision Research: Definitions, Sources, Interrelations, Comments*. New York: Praeger.

Anderson, N.H. (1981). *Foundations of Information Integration Theory*. New York: Academic Press.

Anderson, N.H., and Zalinski, J. (1988). Functional measurement approach to self-estimation in multiattribute evaluation. *Journal of Behavioral Decision Making, 1*, 191-221.

Bamber, D. (1990). Knowledge acquisition for the development of expert systems: An analysis (Technical Document 1322). San Diego, CA: Naval Ocean Systems Center.

Boose, J.H. (1988). Uses of repertory grid-centred knowledge acquisition tools for knowledge-based systems. *Int. J. Man-Machine Studies, 29*, 287-310.

Borcherding, K., and von Winterfeldt, D. (1988). The effects of varying value trees on multiattribute evaluations. *Acta Psychologica, 68*, 153-170.

Bradshaw, J.M., and Boose, J.H. (1990). Decision analysis techniques for knowledge acquisition: combining information and preferences using *Aquinas* and *Axotl*. *Int. J. Man-Machine Studies, 32*, 121-186.

Buchanan, B.G., and Shortliffe, E.H. (Eds.) (1984). *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Reading, MA: Addison-Wesley Publishing Co.

Bursztajn, H., Gutheil, T.G., Hamm, R.M., Brodsky, A., and Mills, M.J. (1988). *Parens patriae* considerations in the commitment process: The "thank you" theory of commitment from the physician's and the patient's perspectives. *Psychiatric Quarterly, 59*, 165-181.

Burton, A.M., Shadbolt, N.R., Hedgecock, A.P., and Rugg, G. (1988). A formal evaluation of knowledge elicitation techniques for expert systems: Domain 1. In D.S. Moralee (Ed.), *Research and Development in Expert Systems IV* (pp. 136-145). Cambridge: Cambridge University Press.

Cheeseman, P. (1985). In defense of probability. In *Proceedings of the International Joint Conference on Artificial Intelligence*. Los Angeles, CA.

Cleal, D.M., and Heaton, N.O. (1988). *Knowledge-Based Systems: Implications for Human-Computer Interfaces*. New York: Wiley/Halsted Press.

Cohen, M.S., Laskey, K.B., and Tolcott, M.A. (1987). A personalized and prescriptive decision aid for choice from a data base of options (Technical Report 87-18). Reston, VA: Decision Science Consortium, Inc.

Cooke, N.M., and McDonald, J.E. (1987). The application of psychological scaling techniques to knowledge elicitation for knowledge-based systems. *Int. J. Man-Machine Studies, 26*, 533-550.

Dawes, R.M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*, 571-582.

Diehr, P., Highley, R., Dehkordi, F., Wood, R., Krueger, L., Teitz, C., and Hermanson, B. (1988). Prediction of fracture in patients with acute musculoskeletal ankle trauma. *Medical Decision Making, 8*, 40-47.

Dyer, J.S. (1990). Remarks on the analytic hierarchy process. *Management Science, 36*, 249-258.

Edwards, W., and Newman, J.R. (1982). *Multiattribute Evaluation*. Beverly Hills, CA: Sage Publications.

Einhorn, H.J., Kleinmuntz, D.N., and Kleinmuntz, B. (1979). Linear regression *and* process-tracing models in judgment. *Psychological Review, 86*, 465-485.

Fischer, G.W. (1991). Commentary on Ravinder and Kleinmuntz. *Journal of Behavioral Decision Making, 4*, 98-99.

Fischhoff, B. (1989). Eliciting knowledge for analytical representation. *IEEE Transactions on Systems, Man, and Cybernetics, 19*, 448-461.

Fischhoff, B., Slovic, P., and Lichtenstein, S. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 330-334.

Forman, E., Saaty, T., Seeley, M, and Whittaker, R. *Expert Choice*, a proprietary software package. Pittsburgh, PA: Decision Support, Inc.

Forrester, J. (1961). *Industrial Dynamics*. Cambridge: MIT Press.

Gabrielli, W., and von Winterfeldt, D. (1978). Are importance weights sensitive to the range of alternatives in multiattribute utility measurement? (Research Report No. 78-6). Los Angeles: Social Science Research Institute, University of Southern California.

Garthwaite, P.H., and Dickey, J.M. (1991). An elicitation method for multiple linear regression models. *Journal of Behavioral Decision Making, 4*, 17-32.

Goldstein, W.M., and Beattie, J. (in press). Judgments of relative importance in decision making: The importance of interpretation and the interpretation of importance. In D.R. Brown and J.E.K. Smith (Eds.), *Frontiers in Mathematical Psychology*. New York: Springer-Verlag.

Gruber, T.R. (1989). *The Acquisition of Strategic Knowledge*. Boston: Academic Press, Inc.

Hamm, R.M. (1990). Materials for guiding elicitation of self-reported formulas describing knowledge in a multi-object domain (Technical Report 90-14). Boulder, CO: Institute of Cognitive Science, University of Colorado.

Hamm, R.M. (in press). Accuracy of alternative methods for describing experts' knowledge of multiple influence domains. *Bulletin of the Psychonomic Society*.

Hamm, R.M., Bursztajn, H., Mills, M., Appelbaum, P., and Gutheil, T.G. (1981). Regression and introspection in the analysis of clinical decision making: The case of the 3-day paper. Boston: Massachusetts Mental Health Center.

Hammond, K.R. (1982). Unification of theory and research in judgment and decision making. Boulder: Center for Research on Judgment and Policy, University of Colorado.

Hammond, K.R. (1983). Teaching the new biology: Potential contributions from research in cognition. In C.P. Friedman and E.F. Purcell (Eds.), *The New Biology and Medical Education: Merging the Biological, Information, and Cognitive Sciences* (pp. 53-64). New York: Josiah Macy, Jr., Foundation.

Hammond, K.R., Anderson, B.F., Sutherland, J., and Marvin, B. (1984). Improving scientists' judgments of risk. *Risk Analysis, 4*, 69-78.

Hammond, K.R., Hamm, R.M., Grassia, J., and Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-17*, 753-770.

Hammond, K.R., McClelland, G.H., and Mumpower, J. *Human Judgment and Decision Making*. New York: Praeger, 1980.

Hoffman, R.R. (1987). The problem of extracting the knowledge of experts from the perspective of experimental psychology. *The AI Magazine, 8*, 53-64.

Hoffman, R.R. (1989). A survey of methods for eliciting the knowledge of experts. *SIGART Newsletter*, Number 108, April, pp. 19-27.

Keeney, R.L. (1977). The art of assessing multiattribute utility functions. *Organizational Behavior and Human Performance, 19*, 267-310.

Keeney, R.L., and Raiffa, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. New York: Wiley.

Keeney, R.L., and von Winterfeldt, D. (1989). On the uses of expert judgment on complex technical problems. *IEEE Transactions on Engineering Management, 36*, 83-86.

Keller, L.R., and Ho, J.L. (1988). Decision problem structuring: Generating options. *IEEE Transactions on Systems, Man, and Cybernetics, 18*, 715-728.

Kleinmuntz, D.N. (1990). Information processing and the misperception of feedback in dynamic decision making. Presented at the 1990 International System Dynamics Conference, Boston, July.

Krantz, D.H., Luce, R.D., Suppes, P., and Tversky, A. (1971). *Foundations of Measurement. Volume I: Additive and Polynomial Representations*. New York: Academic Press.

Larcker, D.F., and Lessig, V.P. (1983). An examination of the linear and retrospective process tracing approaches to judgment modeling. *The Accounting Review, 58*, 58-77.

Leddo, J., Cohen, M.S., O'Connor, M.F., Bresnick, T.A., and Marvin, F.F. (in preparation). Integrated knowledge elicitation and representation framework (Technical

Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Levi, K. (1989). Expert systems should be more accurate than human experts: Evaluation procedures from human judgment and decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics, 19*, 647-657.

Luce, R.D., and Krumhansl, C.L. (1988). Measurement, scaling, and psychophysics. In R.C. Atkinson, R.J. Herrnstein, G. Lindzey, and R.D. Luce (Eds.), *Stevens' Handbook of Experimental Psychology. Second Edition. Volume I: Perceptions and Motivation* (pp. 3-74). New York: Wiley.

Lusk, C.M., and Hammond, K.R. (1991). Judgment in a dynamic task: Microburst forecasting. *J. Behavioral Decision Making, 4*, 55-73.

Lusk, C.M., Stewart, T.R., Hammond, K.R., and Potts, R.J. (1990). Judgment and decision making in dynamic tasks: The case of forecasting the microburst. *Weather and Forecasting, 5*, 627-639.

McLeish, M., and Cecile, M. (1990). Approximate reasoning with an application to medical diagnosis. In I.B. Turksen (Ed.), *Proceedings of NAFIPS'90 Conference, Volume II* (pp. 40-43). Toronto: North American Fuzzy Information Processing Society/University of Toronto.

McNamee, P., and Celona, J. (1987). *Decision Analysis for the Professional with SUPERTREE*. Redwood City, CA: The Scientific Press.

Moninger, W.R., Flueck, J.A., Lusk, C., and Roberts, W.F. (1989). SHOOTOUT-89: A comparative evaluation of knowledge-based systems that forecast severe weather. Preprint Volume, *Uncertainty and AI Workshop 1989*, pp. 265-271.

Moskowitz, H., Wong, R.T., and Chu, P.-Y. (1989). Robust interactive decision-analysis (RID): Behavioral results and implications. *J. Behavioral Decision Making, 2*, 95-111.

Narula, S.C., and Weistroffer, H.R. (1989). A flexible method for nonlinear multicriteria decisionmaking problems. *IEEE Transactions on Systems, Man, and Cybernetics, 19*, 883-887.

Nazareth, D.L. (1989). Issues in the verification of knowledge in rule-based systems. *Int. J. Man-Machine Studies, 30*, 255-271.

Norman, D.A. (1981). Categorization of action slips. *Psychological Review, 88*, 1-15.

Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial Intelligence, 29*, 241-288.

Politser, P.E. (1991). Do medical decision analyses' largest gains grow from the smallest trees? *J. Behavioral Decision Making, 4*, 121-138.

Ravinder, H.V., and Kleinmuntz, D.N. (1991). Random error in additive decompositions of multiattribute utility. *J. Behavioral Decision Making, 4*, 83-97.

Reilly, B.A., and Doherty, M.E. (1989). A note on the assessment of self-insight in judgment research. *Organizational Behavior and Human Decision Processes, 44*, 123-131.

Richmond, B., Peterson, S., and Vescuso, P. (1987). *An Academic User's Guide to STELLA*. Hanover, N.H.: High Performance Systems.

Rohrbaugh, J., and Schuman, S.P. (1988). POLICY PC. Albany, NY: Executive Decision Services, Inc.

Roose, J.E., and Doherty, M.E. (1976). Judgment theory applied to the selection of life insurance salesmen. *Organizational Behavior and Human Performance, 16*, 231-249.

Rouse, W.B., Hammer, John M., and Lewis, C.M. (1989). On capturing human skills and knowledge: Algorithmic approaches to model identification. *IEEE Transactions on Systems, Man, and Cybernetics, 19*, 558-572.

Schon, D. (1988). From technical rationality to reflection-in-action. In J. Dowie and A. Elstein (Eds.), *Professional Judgment: A Reader in Clinical Decision Making* (pp. 60-77). Cambridge: Cambridge University Press.

Slovic, P., Lichtenstein, S., and Edwards, W., (1965). Boredom-induced changes in preferences among bets. *American J. of Psychology, 78*, 208-217.

SPSS, Inc. (1990). *SPSS Categories*. Chicago: SPSS, Inc.

Steels, L. (1987). Second generation expert systems. In M. A. Bramer (Ed.), *Research and Development in Expert Systems III*. Cambridge: Cambridge University Press.

Sterman, J.D. (1989). Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Processes, 43*, 301-335.

Stewart, T.R. (1988). Judgment analysis. In B. Brehmer and C.R.B. Joyce (Eds.), *Human Judgment: The Social Judgment Theory View*. Amsterdam: North-Holland.

Stewart, T.R. (1991). Seven components of forecasting skill. Albany, NY: Center for Policy Research, SUNY Albany.

Stewart, T.R., and Ely, D.W. (1984). Range sensitivity: A necessary condition and a test for the validity of weights (Report No. 270). Boulder, CO: Center for Research on Judgment and Policy, University of Colorado.

Stewart, T.R., and McMillan, C., Jr. (1987). Descriptive and prescriptive models for judgment and decision making: Implications for knowledge engineering. In J. Mumpower et al (Eds.), *Expert Judgment and Expert Systems* (NATO ASI Series, Vol. F35, pp 305-320). Berlin: Springer-Verlag.

Stewart, T.R., Moninger, W.R., Grassia, J., Brady, R.H., and Merrem, F.H. (1988). Analysis of expert judgment and skill in a hail forecasting experiment. Boulder, CO: Center for Research on Judgment and Policy, University of Colorado.

Stillwell, W.G., Barron, H.F., and Edwards, W. (1983). Evaluating credit applications: A validation of multiattribute utility weight elicitation techniques. *Organizational Behavior and Human Performance, 32*, 87-108.

Stillwell, W.G., von Winterfeldt, D., and John, R.S. (1987). Comparing hierarchical and nonhierarchical weighting methods for eliciting multiattribute value models. *Management Science, 33*, 442-450.

Tatman, J.A., and Shachter, R.D. (1990). Dynamic programming and influence diagrams. *IEEE Transactions on Systems, Man, and Cybernetics, 20,* 365-379.

Weber, M., Eisenfuhr, F., and von Winterfeldt, D. (1988). The effects of splitting attributes on weights in multiattribute utility measurement. *Management Science, 34,* 431-445.

White, C.C., III. (1990). A survey on the integration of decision analysis and expert systems for decision support. *IEEE Transactions on Systems, Man, and Cybernetics, 20,* 358-364.

White, C.C., III, Sage, A.P., and Dozono, S. (1984). A model of multiattribute decisionmaking and trade-off weight determination under uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics, 14,* 223-229.

von Winterfeldt, D., and Edwards, W. (1986). *Decision Analysis and Behavioral Research.* New York: Cambridge University Press.

Zalinski, J., and Anderson, N.H. (1986). AVERAGE: A user-friendly FORTRAN-77 program for parameter estimation for the averaging model of information integration theory. La Jolla, CA: Department of Psychology, UCSD.

Zhu, S.-H., and Anderson, N.H. (1991). Self-estimation of weight parameter in multiattribute analysis. *Organizational Behavior and Human Decision Processes, 48,* 36-54.

9 Acknowledgements.

Table 1.
Steps involved in guided model building procedure (from Hamm, 1990).
----
1. Define measurement scale of output dimension and input dimensions. Establish scale type for each.

2. Graphically express relation between each input dimension and the output dimension.

3. Summarize information about all input dimension in one table.

4. Identify interactions between dimensions. Is each input dimension's impact on the output dimension moderated or influenced by any other dimensions?

5. Group input dimensions, according to similarity or existence of interactions, and then arrange groups in a hierarchical tree. Influence flows from twigs to trunk in this upside-down tree. The output dimension is the top node in the tree.

6. Give each group of variables a name, define its measurement scale, graph its effect on the output dimension, and consider whether it has interactions with other groups or individual variables.

7. For each internal node in the tree, specify how the (local) input variables are to be combined into the (local) output variable. This involves choosing whether the variables are to be combined by averaging, by multiplying, or by defining a table that names the output for every possible combination of values of the input variables.

8. When the mathematical form of every node in the tree has been specified, these specifications jointly define the model.