Critical Decisions under Uncertainty: Representation and Structure[*]

Benjamin Kuipers[†]
Alan J. Moskowitz[‡]
Jerome P. Kassirer[§]

Original publication: *Cognitive Science* **12**: 177-210, 1988.
Reprinted in G. Shafer and J. Pearl (Eds.), *Readings in Uncertain Reasoning*, San Mateo, CA: Morgan Kaufmann Publishers, 1990.

[*] This research was supported in part by the National Library of Medicine through NIH Grants LM 03603, LM 04125, LM 04374, and LM 04515 to the first author, and by NIH Grants RR 01320, LM 04022, and LM 04493 to the third author.

[†] Department of Computer Sciences, University of Texas at Austin, Austin, Texas 78712.
[‡] Department of Medicine, Tufts University School of Medicine, Boston, Massachusetts 02111.
[§] Department of Medicine, Tufts University School of Medicine, Boston, Massachusetts 02111.

Abstract

How do people make difficult decisions in situations involving substantial risk and uncertainty? In this study, we presented a difficult medical decision to three expert physicians in a combined "thinking aloud" and "cross examination" experiment. Verbatim transcripts were analyzed using *script analysis* to observe the process of constructing and making the decision, and using *referring phrase analysis* to determine the representation of knowledge of likelihoods. These analyses are compared with a formal decision analysis of the same problem to highlight similarities and differences. The process of making the decision resembles an incremental, sequential-refinement planning algorithm, where a complex decision is broken into a sequence of choices to be made with a simplified description of the alternatives. This strategy results in certain kinds of relevant information being underweighted in the final decision. Knowledge of likelihood appears to be represented as symbolic descriptions capturing categorical and ordinal relations with "landmark" likelihoods, only some of which are described numerically. Numerical probabilities, capable of being combined and compared arithmetically, were not observed. These observations suggest an explanation for the heuristics and biases in human decision-making under uncertainty in terms of the processes that manipulate symbolic descriptions of likelihoods and construct plans of action for situations involving risk and uncertainty.

1 Introduction

The critical decision between diagnostic and therapeutic alternatives faced by a physician is a paradigm example of decision-making under uncertainty. Medical decision making is frequently characterized by the need to make decisions in the face of incomplete knowledge of the patient's true condition and the therapeutic effect of a given management strategy.

People clearly have some knowledge of likelihood[1]. They know that the risk of being struck by a meteorite is negligable, while the risk of being struck by an automobile is not. They know that the risk of an invasive test is greater than the risk of a non-invasive one. Many try to make substantial changes in lifestyle to reduce their risk of cancer or heart disease. However, people seldom know, describe, or utilize numerical values for these likelihoods, and their intuitive knowledge of these values and their abilities to compare the risks of widely different hazards are often quite poor (cf. Kahneman, Slovic, and Tversky (1982)). Formal decision analysis uses the mathematical concept of probability as a tool for describing partial knowledge of uncertain situations [Raiffa, 1970; Schwartz, Gorry, Kassirer, and Essig, 1973]. However, in everyday clinical practice, physicians rarely reason using mathematical probabilities nor do they compute expected utilities. When they do use numerical probabilities, they tend to make errors sufficiently important to result in erroneous medical practices [Eddy, 1982].

By carefully studying expert physicians solving a problem involving considerable uncertainty and risk, we can infer properties of the underlying knowledge representation and reasoning strategies they use to make difficult decisions. The analysis focuses on two separate but related aspects of the knowledge representation:

- the strategy for imposing structure on the decision problem, to decompose it into a set of manageable steps;
- the cognitive representation for likelihoods.

Shafer and Tversky [1985] argue that decision-making under uncertainty can be viewed as a thought-experiment, designed and carried out to select a feasible set of elementary judgments, to estimate the required subjective likelihoods, and to combine them to yield the required decision [Shafer and Tversky, 1985]. They illustrate how such experiments would be designed within the formal structures of Bayesian decision analysis [Raiffa, 1970] and

[1] We reserve the term *probability* for mathematical probability: a measure of uncertainty as a real number between 0 and 1. We use the term *likelihood* for "subjective probabilities": knowledge of the fact that some events are more likely than others.

the theory of belief functions [Shafer, 1976]. This thought-experiment metaphor emphasizes that decisions are designed and constructed, and suggests certain empirical questions for study:

1. How are decisions *actually* designed and made in situations of risk and uncertainty.

2. How are individual judgments of subjective likelihoods obtained?

3. What is the cognitive representation of subjective likelihoods?

The second of these questions is the primary focus of studies by Kahneman, Tversky, and their colleagues [Tversky and Kahneman, 1974; Kahneman, Slovic, and Tversky, 1982], which document systematic biases in human decision-making due to the heuristics used to estimate subjective likelihoods. The results presented here address the first and third questions through a protocol analysis experiment in which expert physicians were asked to make important clinical decisions under circumstances of considerable risk and uncertainty. We use a formal decision analysis of the same problem as a contrasting background for our findings.

In the next section, we describe the medical problem which we presented to physician subjects, along with a formal decision analysis that determines and compares the expected utilities of the different courses of action. The third section describes our protocol analysis methodology. The fourth section presents the results of the protocol analysis of two different aspects of the transcript: the structure of the decision-making process, and the representation of knowledge of likelihood. The fifth section discusses the implications of these observations.

2 The Case of the Pulmonary Infiltrates

Three expert pulmonary physicians (specialists in lung disease) were asked to comment on a case selected from the records of the New England Medical Center.

The patient selected for study was a 63 year old man with preleukemia, a disorder of blood cell production that progresses to overt leukemia. During the year following this diagnosis the patient received multiple blood transfusions to treat the anemia associated with this disease. Additionally during that period he developed mucormycosis of the nasal sinuses, a fungal infection that primarily affects patients with compromised immune defenses. Although this infection was successfully controlled with the antifungal drug amphotericin B, the patient developed kidney toxicity from the drug, which resulted in a mild loss of kidney function.

The immediate problem concerns an acute illness that started with several days of fever, which prompted the patient to seek medical attention. A chest x-ray revealed that the patient had pneumonia involving both lungs diffusely. Further testing after admission to the hospital confirmed that the patient had compromised immune defenses and that his respiratory function was impaired by the pneumonia. While diagnostic sputum cultures were incubating, the patient was treated empirically[2] with antibiotics that would effectively treat all likely causative microorganisms except fungi. This therapeutic exception was made to avoid the toxic effects of the antifungal drug (as noted above) and because a fungus was not thought likely to be the cause of his pneumonia at that point. After three days of treatment, however, the patient did not improve and the cultures did not grow a microorganism which could account for the pneumonia.

Faced with the urgency of treating this seriously ill patient, a clear decision-making dilemma emerged. The subjects had to make a choice to either

1. perform a lung biopsy to try to establish whether the patient had fungal pneumonia, in which case he would be treated with antifungal drugs despite the risk of such treatment, or
2. decide not to carry out a biopsy, but add antifungal drugs empirically to treat the possibility that the pneumonia is fungal in origin.

Each option has its own risks and benefits. Performing a lung biopsy, an invasive procedure, can itself result in death (1.5% of the time in a patient this ill), and will fail to identify the fungus 5% of the time when it is present. However, the great majority of the time, the

[2]"Empiric treatment" is therapy directed at a condition whose diagnosis is strongly suspected but not confirmed.

patient will make it through the biopsy, and will be exposed to the toxicity of antifungal drugs only when the fungus is identified in the biopsy specimen. Empiric therapy avoids the risk of the biopsy procedure, but treats the patient whether the fungus is present or not. Of course, this approach may unnecessarily expose the patient to the drug's toxicity, which may result in kidney failure and death.

To complicate matters further, the physicians had two types of biopsy procedures available to them: open lung biopsy and transbronchoscopic lung biopsy. An open lung biopsy requires that the patient be given general anesthesia, the chest opened and a piece of the lung tissue removed. Transbronchoscopic biopsy, which does not require general anesthesia, involves placing a bronchoscope (a flexible tube through which the lung passages are directly visualized) into the patient's airways, passing a probe through the bronchoscope across the surface of the airway, and nipping a fragment of the lung tissue. The advantage of the open lung biopsy is that a larger piece of the lung tissue is taken, increasing the chance that a fungal organism, if present, will be identified. However, the tradeoff is that open lung biopsy is more risky than the transbronchoscopic biopsy, with a mortality risk of 1.5% versus 0.2% with transbronchoscopic biopsy.

2.1 Formal Decision Analysis

Formal decision analysis [Raiffa, 1970; Schwartz, Gorry, Kassirer and Essig, 1973; Kassirer, 1976; Weinstein and Fineberg, 1980] provides a framework for making explicit the many factors in a complex decision problem like this, and allowing the decision to be made by comparing the expected utilities of the alternatives.

2.1.1 Sensitivity to Assumptions

Although one can construct a richly detailed decision tree that captures many of the nuances of a given patient's illness, such models are often difficult to complete because they require highly specific conditional probabilities for which little or no objective data may exist. In such instances the analyst must make simplifying assumptions to tailor the model to the available data. Where objective data do not exist, the subjective estimates of experienced medical specialists are sought and used as proxies. The resultant models are, therefore, merely approximations to clinical reality and the limits of their validity must be established by observing their behavior with sensitivity analysis. Using sensitivity analysis one systematically varies the values assigned to one or several parameters in the decision tree and recalculates expected utility.

To fit the present clinical problem into the decision analysis formalism, we made several assumptions. For the most part, the probabilistic data that we used were obtained from the medical literature, however, these data were not always specific enough for our patient. One assumption that we made, which is representative of others made in generating the model, concerned the patient's prior probability of having fungal pneumonia, a key factor in the decision. We were able to establish, from the medical literature, that fungi cause approximately 20% of the pneumonias in patients with compromised immune defenses.

However, the studies from which this information was derived examined the prevalence of fungal pneumonia among all immunocompromised patients with pneumonia, not just those who had failed to respond to several days of broad spectrum antibiotic drugs aimed at all likely non-fungal causes of pneumonia, as was the case with our patient. Clearly, not responding to several days of such therapy selects out individuals who have a higher likelihood of having fungi as the cause for their pneumonia. In fact, the physicians that originally cared for this patient believed that he had a 30% to 35% chance of having fungi as the cause of his disease.

2.1.2 Structure of the Decision

The problem is structured as a decision tree, a probability model which was constructed to determine the likelihood that this patient will survive his acute illness with each of the therapeutic options under consideration. The tree lays out the temporal sequence of clinical events, which begin at the root of the tree on the far left and end at the terminal branches of the tree on the far right. On the far left is the square decision node representing the choice among three therapeutic options: treating with amphotericin without further diagnostic studies (branch labeled EMPIRIC AMPHOTERICIN), carrying out a lung biopsy and tailoring treatments to the results (branch designated BIOPSY) and neither giving amphotericin empirically nor carrying out further tests (branch labeled NO CHANGE IN RX).

Following the empiric amphotericin strategy (at the top), is a circular node, representing the notion that the patient's pneumonia may or may not be fungal in origin, with branches labeled FUNGAL PNEUMONIA and NONFUNGAL PNEUMONIA. The tree in Figure 1 could be expanded to contain the various causes of pneumonia other than fungi. The term nonfungal pneumonia is an abbreviation for all of the potential causes of pneumonia that were explicitly considered by our subjects. The outcome at this node is under the control of chance. Each of the two events (branches) has an associated likelihood of occurrence, represented here by the probability values given beneath each branch. In the baseline case (the first order approximation), the probability of fungal pneumonia is 20%. This value was obtained from the prevalence of this kind of infection among immunocompromised patients with pneumonia. Because this strategy (empiric amphotericin) dictates that no further diagnostic information be obtained, amphotericin will be given whether or not the

Not in PS file. Sorry.

Figure 1: The decision tree for a formal decision analysis

fungal infection is present. If the patient has the infection (branch designated FUNGAL PNEUMONIA), the maximum therapeutic benefit from this action is achieved, a survival of 44%, which is the utility (value) associated with this outcome. If the lung infiltrate is not caused by a fungus (branch labeled NONFUNGAL PNEUMONIA) the expected survival is 60%. However, because amphotericin is not entirely a benign drug, its inappropriate use here is assumed to reduce the patients survival by 2%, to a value of 58%.

The middle strategy, designated by the label BIOPSY, is the only option in which diagnostic information is obtained before therapy is instituted. Following this branch is another square decision node with two options, open lung biopsy and transbronchial biopsy, labeled OLBx and TBBx, two alternative means of obtaining lung tissue. If open lung biopsy is selected there is a chance that the patient may or may not survive the procedure, represented by the two branches of the circular chance node labeled DIE and SURVIVE. The probability of dying during the open lung biopsy was taken at 1.5% and the probability of surviving is 98.5%. Should the patient die as a complication of biopsy, the utility of this outcome is taken to be 0. Should the patient survive open lung biopsy there remains only a chance that his pneumonia is actually fungal in origin (branches labeled FUNGAL PNEUMONIA and NONFUNGAL PNEUMONIA).

When the pneumonia is fungal in origin, the biopsy specimen may or may not reveal the presence of fungi (branches labeled FUNGI SEEN or NO FUNGI SEEN). Amphotericin is started when the test result is positive and withheld when the test result is negative. Appropriate treatment of fungal pneumonia with amphotericin will achieve survival 44% of the time, which is the utility value assigned to this outcome. Because this test is imperfect, basing therapy on its outcome will occasionally cause us to withhold amphotericin when the pneumonia is actually fungal in origin. This outcome is represented by the path:

$$\text{FUNGAL PNEUMONIA} \Longrightarrow \text{NO FUNGI SEEN} \rightarrow \text{NO AMPHOTERICIN.}$$

Untreated, fungal pneumonia is uniformly fatal, so the utility value attached to this outcome is 0.

When the pneumonia is not fungal in origin (branch labeled NONFUNGAL PNEUMONIA) lung biopsy may be either positive or negative for fungi, designated respectively by the labels FUNGI SEEN and NO FUNGI SEEN. When treatment is withheld (labeled NO AMPHOTERICIN) the chance of survival is 60%. However, if the biopsy reveals fungi (branch labeled FUNGI SEEN), amphotericin will be given inappropriately and the treatment itself will reduce the patients expected survival from nonfungal pneumonia by 2%, to 58%.

If TBBx is selected over open lung biopsy, the same sequence of events ensue as that depicted above for the option labeled OLBx, unless tissue is not obtained by the transbronchoscopic biopsy. Of course, the probabilities of many of the outcomes are different. In particular, the TBBx mortality rate, 0.2%, is considerably lower than the OLBx mortality rate, 1.5%. However, this reduction in procedure risk comes at the expense of diagnostic yield. Transbronchoscopic biopsy will not yield lung tissue every time, and even when lung tissue is obtained, fewer fungal pneumonias are identified, i.e., the test positive rate when fungal pneumonia is present is lower for TBBx than for OLBx, 71.5% vs 95%.

If lung tissue is not obtained by transbronchoscopic biopsy (branch labeled NO TISSUE), the patient is subjected to an open lung biopsy procedure, represented by the branch following the square shaded node. The tree logic following open lung biopsy at this point is identical to that seen in the open lung biopsy strategy above. However, because time is lost in attempting a transbronchoscopic biopsy first, antifungal therapy is delayed and the prognosis of fungal pneumonia is reduced by 4%.

The last therapeutic option is to neither intervene diagnostically nor add amphotericin empirically (branch labeled NO CHANGE IN Rx). In the baseline case, fungal pneumonia is present 20% of the time and will always be fatal. However, 80% of the time, a nonfungal pneumonia is present and by not administering amphotericin survival is uncompromised.

2.1.3 Calculated Expected Utilities

The expected utility of each strategy in the decision tree is calculated using a weighted average technique, referred to as "averaging out and folding back" the decision tree. This process entails averaging together the utility values of each potential outcome for each given strategy, after weighting them by their likelihood of occurrence.

In the present analysis, the choice of open lung biopsy provides the highest expected utility (55.5% survival), followed closely by empiric amphotericin (55.2% survival), and transbronchial biopsy (54.2% survival). The no-change-in-therapy strategy is a distant fourth with a 48% expected survival. The results of this analysis are based on the best available data from the medical literature and on estimates obtained from infectious disease consultants in instances in which such data were not available (Table 1).

Although the expected utilities of the strategies are nearly indistinguishable, the utility of not changing therapy is clearly inferior. Open lung biopsy is favored by a 0.3% survival margin, which according to some analysts, qualifies the decision as a "toss-up" [Pauker and Kassirer 1980]. Indeed, if the analysis were fine-tuned further by including quality adjustments for the short term morbidities of the limited thoracotomy (opening the chest for the lung biopsy), the bronchoscopy, or the prolonged hospitalization for amphotericin therapy (six to eight weeks), these differences might diminish even further.

Recall that a key assumption was the 20% likelihood of fungal pneumonia, where some

Cost (in survival) of amphotericin in nonfungal pneumonia	2%
Cost (in survival) of delaying definitive therapy	4%
Mortality of open biopsy	1.5%
Mortality of transbronchoscopic biopsy (TBBX)	0.2%
Prior probability of fungal pneumonia	20%
Probability of getting tissue with TBBX	93.5%
Sensitivity of open biopsy	95%
Sensitivity of TBBX	71.5%
Specificity of open biopsy and TBBX	99%
Utility of procedure death	0
Utility of untreated fungal pneumonia	0
Utility of treated fungal pneumonia	44%
Utility of nonfungal pneumonia	60%

Table 1: Numerical data for the decision tree, derived from the medical literature and estimates from expert consultants.

experts estimated the value as high as 30% to 35%. Sensitivity analysis determines the change to an estimated value that would be required to change the recommended strategy. It reveals that empiric amphotericin B dominates the other three strategies when the prior probability of fungal pneumonia is greater than 0.21. Above this value, the casualties from diagnostic error (missing the fungi when present) for either OLBx or TBBx, is greater than the mortality from the overuse of amphotericin B. Other sensitivity analyses that we performed substantiated the "close call" nature of this decision.

The decision tree provides a useful framework for comparison of the cognitive decision-making strategies with formal decision analysis. As we shall see from analysis of the transcripts, the expert physicians took quite a different view of this difficult decision. Empiric treatment with amphotericin B, which is at least a close call and possibly the dominant treatment strategy according to the formal decision analysis, is hardly considered at all by the experts.

3 Data Collection and Analysis Methods

In a previous survey, we reviewed the experimental methods available for the study of clinical cognition [Kassirer, Kuipers, and Gorry, 1982]. We concluded that the complex phenomena of knowledge representation and problem-solving strategies are best derived from the type of data that appear in verbatim transcripts (or *protocols*) of interview studies. A more statistically tractable source of data such as a questionnaire would obscure the details we need to observe, by aggregating across subjects to eliminate the effect of individual variation. In previous studies of expert physicians, interviews producing verbatim transcripts appeared to provide sufficient fidelity to the actual task environment, and the analysis of these transcripts is feasible, though laborious. The additional effort and methodological difficulties required for collection and analysis of videotaped patient encounters, for example, does not, in our opinion, pay off in better or more revealing data about clinical cognition [Kassirer, et al, 1982].

In this study, we used two different types of verbatim protocol collection methods on each subject, thinking aloud protocols and cross examination protocols.

- *Thinking aloud protocols* are obtained by recording subjects' responses to clinical data without interruptions for questions. This technique is sensitive to the natural control flow of the subject's problem-solving, but it is insensitive to the limits of the knowledge stored. Care must be taken to distinguish the trace of on-going problem-solving from the "canned" explanation of a previously produced result.

- *Cross examination protocols* are obtained after the thinking aloud protocol is completed, by asking the subject to respond to specific questions about the clinical problem at hand. These questions probe the limits of a subject's knowledge directly, even resembling oral examinations, but they are insensitive to the subject's natural control structure.

A mix of the two methods makes it possible to examine different aspects of the subjects' knowledge representation.

The analysis of the protocols can also be of several different types, looking at different aspects of the reasoning in the transcript.

- *Referring phrase analysis* identifies the set of referring noun phrases in a protocol excerpt and defines a small natural universe of underlying conceptual objects which can be the referents of those phrases. This universe constitutes an ontology for the domain being discussed.

- *Assertional analysis* identifies the set of assertions being made in the excerpt about the objects identified by referring phrase analysis. A set of relations on objects and connectives and operators on sentences are then defined to express the content of the assertions. This constitutes an epistemology for the domain being discussed.

- *Script analysis* identifies the overall structure of the reasoning process, argument, or explanation being given in the excerpt. The analysis is intended to reveal the goal structure of the problem-solving process or the explanation strategy.

The first two methods were developed as part of our study of causal reasoning in medicine [Kuipers and Kassirer, 1983, 1984], and the third was developed for the needs of the study reported here. The methodological principles underlying protocol analysis are examined in depth by Ericsson and Simon (1984) using an information processing model of the verbalization process.

4 Results of Protocol Analysis

4.1 The Structure of the Decision

We analyzed verbatim protocols of three expert physicians as they were deciding how to manage the patient. We excerpted for detailed analysis the fragment of each interview, after all the data had been presented, in which each subject considered whether to apply an invasive test (a lung biopsy) and which type of test to apply (a transbronchoscopic biopsy (TBBx) or an open-lung biopsy (OLBx)). This decision is a major portion of the decision whether to administer amphotericin.

We believe that the transcript represents a trace of an ongoing problem-solving process in all three protocols, particularly because we observed our subjects making statements and reaching conclusions which they modified after further consideration. The subjects usually did not remark on the change of opinion.

Script analysis allowed us to identify the steps of reasoning and decision-making in the different protocols. Appendix A presents a complete script analysis of the "thinking aloud" section of one of the transcripts.

There was considerable variation among the different subjects in their explanation strategies. In explaining their reasoning, some subjects were considerably more verbose than others. For example, in some cases a subject would describe the current decision problem, discuss a set of relevant factors, and finally select an action. In other cases, the problem was left unstated but the factors that were discussed, and the statement of the final choice, clearly presupposed the problem formulation. In yet other cases, the problem was stated, the relevant factors were discussed, but no decision was stated explicitly except as a presupposition to the next problem. In such cases, we considered it justified to insert the implicit reasoning step into the coding sequence, suitably marked.

There were also significant idiosyncratic elements in the reasoning strategies used by the three different physicians. One subject built a management plan in a straight-forward manner. Another did a considerable amount of reasoning in hypothetical contexts both before and after the point in the case where the problem was posed. The third paused in the middle of a decision to consider a new aspect of a prior decision.

In spite of these differences, there are important similarities among the reasoning protocols of the physician subjects, and notable distinctions between the protocols and formal decision analysis. To clarify this common structure, we extract from the script analyses of the three transcripts those operators which deal directly with formulation of a decision problem (PROBLEM), the selection of an alternative (CHOOSE), or insertion of a new action into a partially-completed plan (INSERT). Tables 2, 3, and 4 present this skeleton of

PROBLEM: Biopsy vs. No change in therapy. (implicit)
CHOOSE: Biopsy.
In hypothetical context (time of admission, 48 hours previously):
  PROBLEM: Immediate biopsy vs. Wait 48 hours on current therapy.
  CHOOSE: Wait 48 hours on current therapy.
PROBLEM: TBBx vs. OLBx. (implicit)
CHOOSE: TBBx.
In hypothetical context (after negative result from TBBx):
  PROBLEM: Accept negative finding vs. Go on to OLBx.
  CHOOSE: If condition improves, accept negative finding.
  CHOOSE: If condition deteriorates, go on to OLBx.
PROBLEM: Consider pericardial tap.
PROBLEM: Consider pleural effusion tap.
PROBLEM: Consider repeated X-rays. Reject.
REVIEW: TBBx in context of possible diagnoses.

In addition to the common sequence, this subject formulated additional decision problems in hypothetical contexts.

Table 2: Script analysis for VP-18.

each of the three script analyses.

The three physician subjects share a common sequence of operations in their formulation of the decision:

PROBLEM: Biopsy vs. No change in therapy.
CHOOSE: Biopsy.
PROBLEM: TBBx vs. OLBx.
CHOOSE: TBBx.
REVIEW: TBBx in context of possible diagnoses.

By way of contrast, if we were to construct a formal decision analysis, including the same factors discussed by the physicians, and with the probabilities and utilities assigned so that it would reach the same conclusion, its reasoning would follow a quite different sequence:

PROBLEM: Biopsy vs. No change in therapy.
PROBLEM: TBBx vs. OLBx.
REVIEW: all choices (TBBx and OLBx) in context of possible diagnoses.
CHOOSE: TBBx.

PROBLEM: Biopsy vs. No change in therapy.
CHOOSE: Biopsy. (implicit)
INSERT: Sputum culture, before biopsy.
PROBLEM: TBBx vs. OLBx.
    Reject percutaneous needle biopsy and transtracheal aspiration.
CHOOSE: TBBx.
INSERT: Lavage, with TBBx.
INSERT: OLBx, if no results from TBBx and patient deteriorates.
REVIEW: TBBx in context of possible diagnoses.

In addition to making the decisions in the common sequence, this subject inserted non-critical steps into the plan as opportunity presented itself.

Table 3: Script analysis for VP-19.

PROBLEM: Biopsy vs. No change in therapy. (implicit)
CHOOSE: Biopsy.
PROBLEM: TBBx vs. OLBx.
    PROBLEM: Pericardial tap vs. Biopsy.
    CHOOSE: Biopsy. CHOOSE: TBBx.
REVIEW: TBBx in context of possible diagnoses.

After formulating the more detailed problem, this subject returned to consider an eliminate a new alternative to a previous decision.

Table 4: Script analysis for VP-20.

CHOOSE: Biopsy.
REVIEW: sensitivity to variations in probabilities and utilities.

Thus, the common structure shared by the physicians reflects the structure of the knowledge representation and the inference methods used, and cannot be merely a mathematical consequence of the structure of the problem.

As noted below, we also observed that none of the physician subjects explicitly considered the possibility of empiric treatment with amphotericin during the "thinking aloud" portion of the interview. When the topic was probed during the "cross examination" portion, empiric amphotericin was considered obviously inappropriate unless the patient were too sick for any biopsy at all.

Other steps in the reasoning process, such as insertion of non-critical plan steps, review of risk factors and prerequisites to biopsies, and examination of less important alternative treatments, were not consistently ordered with respect to the core sequence.

The implications of these findings will be discussed in more detail in section 5.

4.2 Representation of Likelihood

The second question we address is how knowledge of likelihood is represented. That is, how are subjective probabilities described? We used referring phrase analysis [Kuipers and Kassirer, 1984] to identify and classify each phrase in the transcripts that referred to likelihood. The assumption behind referring phrase analysis is that any concept or relation that is expressed verbally in an explanation must be expressible in the cognitive knowledge representation.

The selection criteria for phrases to be analyzed are designed to be inclusive, rather than making a closely defined distinction. A phrase is included for analysis either because it contains a relevant word like *chance*, *risk*, *yield*, etc., or because the meaning of the phrase in context clearly relates to likelihoods.

Once a set of referring phrases has been identified, each phrase is classified into one (occasionally two) of a small set of categories, according to the type of description of a likelihood that is being asserted. An analysis is considered successful if a small set of categories can account for the complete set of identified phrases.

Table 5 shows a protocol extract with the phrases referring to likelihoods in boldface. The righthand column in that table gives the classification of each phrase.

We accounted for all relevant referring phrases in the transcripts by classifying them into three major categories, with sub-classifications.

- **Descriptions of Likelihood.** The likelihoods of events are generally described in symbolic, rather than numerical, terms.

L183	But if this man were severely thrombocytopenic,	
L184	and if his bleeding time was abnormal,	
L185	I would put him in **a much higher risk category**	ORD,CAT
L186	for undergoing an open procedure	
L187	or a transbronchial biopsy for that matter.	
L188	**He could probably still undergo a lavage**.	CAT
L189	You know, **I would still probably consider** doing that	HEDGE
L190	because **the yield**,	REF
L191	even in fungal infection	
L192	is **pretty good** with a lavage.	CAT
L193	But I would be **very reluctant** to do any invasive procedure on him,	CAT
L194	because I think **the chances of killing him** . . . due to hemorrhage,	REF
L195	would be **in the range of the ten-fifteen percent**	NUM
L196	and I think **the chances of killing him with amphotericin**	REF
L197	would probably be **a bit less than that**.	ORD
L198	So if he were deteriorating and	
L199	if I felt	
L200	based on these parameters	
L201	that doing an invasive procedure would be **too risky**,	ORD
L202	then I would certainly cover him empirically.	

The boldfaced phrases are those phrases considered to make some direct or indirect reference to likelihood. The annotations at the right of certain lines classify the referring phrases into the categories described in the text.

Table 5: A fragment from the cross-examination portion of a transcript.

CAT: categorical values without precise boundaries: low, moderate, very high, etc.

ORD: ordinal assertions of relative likelihoods: greater-than, less-than, equal-to.

NUM: numerical values or ranges

- **Possibility of Events.** When establishing the *structure* of a decision, possible events are roughly classified according to whether they will be treated as alternative outcomes in the decision. These references are a special case of categorical descriptions, distinguished by their role in formulating the decision.

  ALWAYS: This event, though perhaps not absolutely certain, may be assumed to occur.

  SOMETIMES: This event is one of a set of possibilities; its likelihood will need to be considered.

  NEVER: This event, though perhaps not absolutely impossible, may be excluded from consideration.

- **Other References.** Surface constructions referring to likelihoods without providing information about the underlying knowledge representations.

  REF: A construction referring to a likelihood without describing it, e.g. "The chances of killing him with amphotericin . . . "

  HEDGE: A rhetorical usage indicating lack of commitment, e.g. "I would still probably consider doing that."

This classification accounts for every referring phrase in the three protocols meeting our selection criteria, leading us to believe that it captures real properties of the representation of knowledge of likelihood and probability. Table 6 shows the total number of phrases in each category found in the different transcripts.

Two conclusions may be drawn from these data. First, the descriptions of likelihood used in the physicians' explanations are almost exclusively qualitative. Second, there is substantial individual variation in the mix of phrases used. Appendix B includes the complete sets of CAT, ORD, and NUM referring phrases from the most verbose of the transcripts (VP-18).

The main results of the referring phrase analysis are, first, that references to likelihoods consist primarily of two distinct kinds of symbolic description of likelihood — categorical and ordinal — and, second, that references to numerical values or ranges are treated in the same way as qualitative descriptions, without taking advantage of their numerical properties.

	VP-18	VP-19	VP-20
CAT	20	9	10
ORD	19	14	5
NUM	12	1	3
SOMETIMES	7	7	6
ALWAYS	5	0	0
NEVER	2	2	0
REF	13	8	1
HEDGE	0	4	1

Table 6: Counts of Referring Phrases by Category in Three Transcripts

Numerical terms (coded as NUM in the referring phrase analysis) appear in the explanation as descriptions of likelihoods, but in a way that is quite different from the mathematical theory of probability. In some cases, numerical descriptions act as categorical descriptions by providing focal values for a region without precise boundaries, e.g. "... you have maybe a 70 - 80% chance (of) making the diagnosis." In other cases, the numerical value is used as an anchoring value for an ordinal description, e.g. "... that occurs you know less than a half percent of the time." In both cases, rather than functioning as coordinates on the real number line, capable of supporting arithmetic operations, the numerical phases serve as *names* for focal or bounding points defining qualitative regions. Appendix B presents the complete set of categorical, ordinal, and numerical referring phrases from one transcript.

These observations suggest that knowledge of likelihood is represented as symbolic descriptions of categorical and ordinal relations, and that numerical descriptions function as easily communicable names for important landmarks rather than as real numbers. The concept of "landmark values" as qualitatively important distinctions is inspired by research on visual similarity and cognitive maps [Goldmeier, 1972; Lynch, 1960; Kuipers, 1978], and contributes importantly to research on qualitative reasoning about physical mechanisms [Bobrow, 1985].

5 Discussion

5.1 The Structure of the Decision

Our analysis of the reasoning stages of three expert physicians as they formulated the structure of the decision revealed a consistent pattern. They made an initial decision at an abstract level, and then went on to specify it more precisely. In a rough comparison with a decision tree, their reasoning steps could be seen as moving from the root to the terminal branches, but with various elaborations considered at unpredictable times (i.e. "opportunistically"). Their approach contrasts with formal decision analysis, which lays out a tree of choice and chance nodes, determines the utilities of a large number of highly specific final outcomes, then backs up through the tree to calculate the expected utilities of the branches at the fundamental choice point.

The decision process we observed in the protocols is well described as one of planning by successive refinement of an abstract plan [Sacerdoti, 1975], combined with opportunistic insertion of plan steps [Hayes-Roth and Hayes-Roth, 1979].

- The decision is first approached at its most abstract level: "Test or wait?"
- Two factors argue in the same direction: there is no clear response from the current therapy, and there is considerable uncertainty about the identity of the infection.
- This formulation of the problem presupposes a decision on a closely related question: "Treat or test?"
- This implicit decision is dominated by a single factor: the risk of harming the patient with unnecessary therapy is viewed as large compared with the risk of the invasive test.
- Once the decision to test has been made, the next decision is formulated: "Which test?"
- This decision is dominated by the trade-off between relative invasiveness and relative yield.
- As opportunity arises, additional diagnostic and therapeutic procedures which are not part of the critical decision are inserted into the plan.
- Finally, the plan is reviewed under various hypothetical circumstances, in case revision should be indicated.

The cognitive methods for formulating a decision thus appear to be incremental and local, breaking the overall decision into manageable pieces that can be decided on the basis of only a few attributes. This method is cognitively useful because it allows sophisticated reasoning with limited computational resources, but it underweights factors whose significance is not local to an individual point in the decision.

5.1.1 The First Choice: Invasive Biopsy vs. Continue Current Therapy

This was not a difficult choice for our subjects, nor should it have been. The formal decision analysis also shows that the choice of continuing the current therapy in hopes of improvement clearly has the lowest expected utility of all possibilities considered.

In the protocols, we observe two influences on the decision, both in the direction of the invasive test. First, the patient is clearly very sick, and there has been no improvement after a reasonable time period on the previous therapy. While it is still possible that the current therapy is correct and the patient's state will soon improve, the patient's poor condition adds urgency and argues against waiting. Second, the identity of the patient's infection is not known. Because there are specific, though risky, treatments for fungal infections, there is pressure for testing to reduce the physician's uncertainty and to select more appropriate treatment.

In formal decision analysis, the complete set of alternatives for this decision would be: "Treat, Test, or No Treat?" Here we are considering only the last two. The first possibility, of treating without testing, has already been discarded.

5.1.2 The Implicit Choice: Empiric Amphotericin Therapy vs. Invasive Biopsy

Our subjects did not explicitly discuss the alternative of empiric treatment with amphotericin except under direct cross-examination. The dominant value in the subjects' decision-making was the need to *know* that a fungal infection is present before administering amphotericin. The unacceptable scenario of administering this putatively dangerous drug when the fungal infection is not present appeared to dominate the reasoning of our subjects. This is a specific instance of one of the oldest principles in medicine: "As to disease, make a habit of two things – to help, or at least to do no harm." [Hippocrates, Epidemics, B.I, Sect. XI, translated by W. H. S. Jones] Only if a patient were so sick that a biopsy would probably be fatal would the physicians be willing to administer amphotericin empirically.

However, the formal decision analysis shows that the decision between empiric administration of amphotericin without a biopsy, versus an invasive biopsy, may be more difficult than our subjects recognize. The invasive biopsy has two significant drawbacks. First, there is a small but significant chance that this very ill patient may die as a direct result of the biopsy procedure. Second, he may die from untreated fungal pneumonia if the biopsy result is falsely negative. Each of these factors decreases the utility of invasive biopsy relative to empiric treatment with amphotericin. The individual and combined impact of these factors appears to be under-weighted by our subjects.

Viewing the decision as a planning process, we can suspect the existence of a rule of the form:

```
if a patient is likely but not certain to have disease D,
   and the therapy for disease D is risky,
   and the patient is able to tolerate a definitive test for D,
then test for disease D before administering the therapy.
```

We find an explanation stated in almost this form in the cross-examination portion of one of the protocols.

L198 So if he were deteriorating and
L199 if I felt
L200 based on these parameters
L201 that doing an invasive procedure would be too risky,
L202 then I would certainly cover him empirically.
L203 But, as long as I thought he could undergo an invasive procedure,
L204 I would do it. [VP-19]

This rule regards the test abstractly as an information source. The risks of the test are only minimally considered, and the possibility of false negative results are not considered at all. These more detailed characteristics are considered only after the decision to test has been made, when choosing between biopsy alternatives.

5.1.3 The Second Choice: Open Lung vs. Transbronchoscopic Biopsy

Once the decision had been made to test with an invasive biopsy, the decision between methods involved a trade-off. The primary consideration, and the one that dominated the final decision, was that TBBx is a less invasive procedure (i.e. less risky) than open lung biopsy. The secondary consideration is that TBBx may yield an insufficient amount of lung tissue, resulting either in an increased chance of a false negative result or the need to do the open lung biopsy anyway, as a second procedure, after some delay.

As with the first choice, a single consideration appeared to dominate the subjects' reasoning and they consistently chose the transbronchoscopic biopsy. In two of the protocols, contingency plans were formulated for proceeding on to OLBx in case TBBx produced insufficient tissue. Factors such as risk of the test and possibility of insufficient tissue or false negative results were considered only in the local context of the choice between biopsy alternatives. Only the relative risks of TBBx and OLBx are relevant to this decision, not the absolute level of risk. However, the absolute level of biopsy risk, and other factors that might affect the previous "Test or Treat" decision, were not propagated globally and reconsidered.

5.1.4 Comparison of Decisions

The final decisions made by all the subjects are plans of approximately the following form, with justifications and caveats included.

- Perform a lung biopsy, because the current therapy has not produced an improvement and the diagnosis is still unknown.
- Perform a lung biopsy before treating with amphotericin, to avoid the possibility of giving a dangerous drug when the infection is not present.
  - If the patient is too sick for a biopsy, however, treat without one.
- Do the transbronchoscopic biopsy, because it is less invasive.
  - If results are inconclusive, do the open lung biopsy next.
  - If the patient is too sick for two diagnostic procedures, do only the open lung biopsy.

Each choice is well justified, based on the factors taken into account when it was made. However, it is clear that information that was potentially relevant to the first decision was only considered within the local context of the second decision.

This pattern is in marked contrast with the formal decision analysis in section 2, where the detailed risks and possible errors of the tests are combined into expected utility values and propagated globally. As a result, three alternatives — the two biopsies and empiric therapy — end up with very similar expected utilities, rather than the clearcut distinctions seen by the human subjects.

It might be argued that human decision-makers take into account subtle factors that are inexpressible in the decision analysis formalism. This did not appear to be reflected in our protocol analysis. The subjects considered a variety of factors in making their choices. However, the use of information in a local context to make a series of incremental decisions was the dominant characteristic we observed in the protocols.

There is a clear reason why human decision-making should follow the approach of incremental planning rather than formal decision analysis: computational resources are limited. Human cognitive processing is relatively slow, has a very limited supply of working memory (though effectively unlimited long-term memory), and is subject to frequent interruptions [Miller, 1956; Newell and Simon, 1972; Sternberg, 1966].

An incremental planning strategy allows a large and difficult decision problem to be broken into a sequence of small choices, each of which can be made on the basis of a relatively limited amount of information. If the planning process is interrupted, the partially refined plan can be stored and reexamined later, so relatively little work is lost. A formal decision analysis, on the other hand, requires a large decision tree with many numerical values to be found, computed, and stored before the decision can be made. If working memory is not large enough for the tree, or if part of the tree is lost during an interruption, a great deal of work is lost. Thus, formal decision analysis is only feasible with significant amounts of external memory and computational resources. The incremental planning approach is far better suited to the cognitive resources of human decision-makers. Further general implications of these computational limitations are discussed in [Kuipers, 1979].

5.1.5 Decision-Making as Planning

Our observations of this planning process are consistent with the findings of Hayes-Roth and Hayes-Roth (1979), who have performed protocol analysis studies of planning processes, and propose an "opportunistic" model of the planning process. In their model, a number of local specialists communicate through a shared "blackboard" data structure, and assemble a plan in a sequence of loosely-coordinated steps. Their model predicts that the different steps of the planning process will not be strictly ordered: their sequence will be determined partly by the structure of the problem and partly by which information happens to be available, or which inferences happen to be activated, at which times. We observe this, in that there is considerable individual variation in the overall reasoning process, especially in the points when hypotheses are considered and risks are assessed.

The consistent sequence of decision steps we observed may be explained by two factors. First, the logical structure of the problem determines that one must decide *whether* to test before deciding *which* test. Second, the sequential refinement character of the process, where an abstract decision is completed before a more specific one is formulated, conserves the working memory that would otherwise be needed to store large numbers of open problems. In Sacerdoti's NOAH planner (1977) the sequential refinement sequence is specified by the intrinsic control structure of the planner. Hayes-Roth and Hayes-Roth (1979) observe that the sequential refinement sequence can be derived, in an opportunistic model, from constraints within the problem. We can extend their point to argue that resource limits in the planner may force decisions at an abstract level to be made before more detailed problems can even be considered.

We also observe a substantial difference between what plans look like in the AI planning literature (e.g. [Sacerdoti, 1977]) and what they look like to our physician subjects. In particular, patient management planning raises a number of important issues not considered in previous AI planning research. The state of the patient and the effects of actions are both uncertain. Diagnosis and therapy are interleaved rather than sequential. Even when the patient's disease is known, a given therapy may have several possible outcomes. As a result, decision-makers construct contingency plans, including branches on possible future events. There are also embedded planning operators: a physician may state that at a certain point in the future there will be enough information to formulate an appropriate plan. Such observations provide good evidence that the medical domain is a fruitful area for future research in planning.

5.1.6 Decisions, Explanations, and Policy

There may be additional factors, distinct from the direct effect of computational resource limits on the decision-maker, responsible for the local nature of the observed decision-making process. The physician is under several constraints that make his task different from that of minimizing the patient's probability of death. First, Hippocrates' command, "Do no harm," may cause the physician to give additional weight to negative effects the patient suffers as a (potentially preventable) result of direct action, e.g. administration of a dangerous drug when the disease is not present. Second, the physician knows that a decision must be explainable as reasonable and prudent under the circumstances.

Different decision-making processes lend themselves to different types of explanations. The explanation of a decision analysis requires justifying the choice and chance nodes included in the tree, and demonstrating through sensitivity analysis (section 2.1.1) the effects of the estimated probabilities and utilities on the outcome [Pauker and Kassirer, 1987]. Calculation of expected utilities at a choice node presupposes the relatively sophisticated concepts of probability and weighted average.

The local process of constructing a plan through sequential refinement is explainable in terms of rules that transform the plan from its initial to its final form. For example, consider a rule that says, "A dangerous drug may only be given after establishing the presence of the disease it treats." Such a rule represents a *policy*: a universally quantified imperative specifying actions appropriate to situations matching the rule antecedents. In this case, the rule provides an ordering constraint on actions in the plan, and effectively prohibits empiric treatment with a dangerous drug.

In order to be useful and comprehensible, the antecedents of a policy must be relatively simple. Considered across situations matching its relatively simple antecedent, a rule such as the one above appears to be (and very likely is) the best possible policy. A sequence of such rule applications (or syllogisms), appropriately linked, provides a convincing and defensible *syllogistic* explanation of the overall decision.

The explanation of a formal decision analysis may be seen by a physician as more sophisticated, less familiar to patients, and more difficult to construct, convey, and defend than a local, step-by-step, syllogistic explanation. Thus, even though a decision analysis can weigh global considerations that are not seen by the rule-based approach, the more explainable method may be preferred.

Another factor is the representation of knowledge learned from previous experience, and its role in both decision-making and explanation. If previous experience is compiled into rules or policies, it is natural for both the decision-making process and the explanation to draw on that form of existing knowledge, rather than constructing each decision from scratch.

Although these additional factors must be considered in interpreting our results, they are not entirely independent of the computational resource limits we have discussed. Constructing an explanation is a process that requires computational resources of various kinds, including previously acquired knowledge, and it must allow for the resources available to a listener attempting to understand the explanation. Further development of a computational theory of critical decision-making must take these factors into account.

5.2 Representation of Likelihood

Our referring phrase analysis provides constraints which must be satisfied by the representation of knowledge of likelihood used by expert physicians. Thus, a cognitive model of knowledge of likelihood must have the following characteristics.

- All likelihood descriptions are symbolic. There was no evidence of high-resolution storage of numerical values, or the cognitive use of arithmetic operators. Numerical values are used as recognizable landmark values.
- There are two distinct symbolic descriptions of likelihood:
  - categorical descriptions: “fuzzy” values on an absolute scale;
  - ordinal descriptions: sharply-defined ordinal relations (greater-than, less-than, equal) between pairs of values.

- Reference or "landmark" values, specified either numerically or in terms of a well-known example situations, serve as focal or boundary values for symbolic descriptions of likelihoods.

The important role of landmark values, as focal values for categorical descriptions or as boundaries for ordinal descriptions, leads us to speculate about the derivation of these descriptions. This provides a possible explanation for two of the phenomena documented by Tversky and Kahneman (1974): the *representativeness* heuristic and the *anchoring* heuristic.

If categories are defined in terms of one or more focal examples, the appropriate categorical description is determined by the closest match between an observation and the focal examples of the different categories. A risk category such as "very high" would be associated with a small set of domain-specific, focal examples of very high risk situations. The risk of a particular situation is assessed by determining its similarity to the focal examples. This is exactly the *representativeness* heuristic. Although it is a reasonable heuristic in many cases, it can lead to substantial biases in cases where the prior probability of the most similar category is very much smaller than that of a less similar one.

Ordinal descriptions are defined by boundary values, which may be well-known landmarks or may be recently derived values. These descriptions can be derived, or at least filtered, by adjusting a value away from a reference point. For example, suppose we know that the mortality risk of an open lung biopsy (OLBx) is "low." Transbronchoscopic biopsy (TBBx) differs from OLBx on two features relevant to risk of mortality: TBBx is less risky than OLBx in only requiring local rather than general anesthesia, and it is less risky in being less invasive. Thus, both of the relevant differences between TBBx and OLBx provide evidence that the risk of TBBx is less than that of OLBx. Even without knowing the magnitudes of the differences, we can conclude that the overall mortality risk from TBBx is less than that of OLBx. This type of inference is essentially the *anchoring heuristic*, and is thus subject to biases in the direction of the anchoring value.

Computationally, categorical and ordinal representations have different advantages and disadvantages (figure 2). Categorical descriptions provide a qualitative classification of values that is useful in rule antecedents (cf. Clancey (1985)). For example, we observe one subject describing what appears to be a rule application:

L0183	"But if this man were severely thrombocytopenic
L0184	and if his bleeding time was abnormal
L0185	I would put him in a much higher risk category
L0186	for undergoing an open procedure
L0187	or a transbronchial biopsy for that matter." (VP-19)

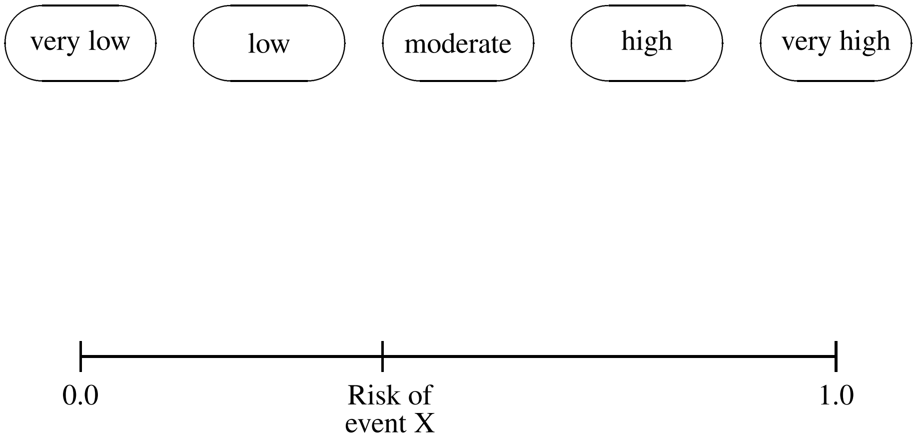

Figure 2: Categorical and ordinal descriptions of risk.
A categorical description is one of a small set of regions without definite endpoints. An ordinal description is a point or interval defined by specified landmark values.

By contrast, ordinal relations support high-resolution distinctions in the neighborhood of a boundary, but highly aggregated, qualitative descriptions in the regions between landmarks. Some have argued that ordinal relations with landmark values are the basis of human similarity judgments. In a dramatic series of studies, Goldmeier (1972) observed that similarity judgments among visually observed figures are strongly affected by ordinal comparisons with certain landmark values. Two values which are quite close but separated by a landmark are judged to be "less similar" than two values which have no intervening landmark, even though they are more widely separated.

In general, categorical and ordinal value descriptions are not mutually translatable. A categorical description such as *very high* does not correspond to the set of values greater than some definable landmark value. Conversely, one might know that $x < y$ without having x and y belong to different categories. There are, of course, certain inferences that one *can* make, to find new ordinal or categorical descriptions, or to rule out certain combinations.

- if x is very high, and $x < y$, then y is very high;
- if x is low, and y is high, then $x < y$.
- if x is normal, and $x < y$, then y cannot be low.

Our observations provide empirical evidence that is generally consistent with Cohen's proposal (1985) for symbolic rather than numerical representation of probabilities. However, Cohen's specific idea of *endorsements* requires creating, storing, and manipulating an extensive structure of justifications as part of each description of a likelihood. Furthermore, his proposal allows an open-ended and perhaps domain-specific language for endorsements. Ordinal and categorical descriptions of likelihood are not as domain-independent as numerical probabilities, since the set of landmarks or categories is domain-specific. However, they do not carry an audit-trail of justifications from the prior stages of a problem. Our observations of the local nature of the decision-making process suggest that there is normally very little provision for reexamining an earlier conclusion when additional relevant evidence is encountered. Thus, while we agree with Cohen's belief that knowledge of probabilities should be represented symbolically, the evidence from our study does not require an explanation as complex as his theory of endorsements.

6 Conclusions

Our analysis suggests ways to account for some of the heuristics, biases, and distortions observed when people reason about likelihoods [Tversky and Kahneman, 1974; Kahneman, Slovic, and Tversky, 1982]. We have observed two important characteristics of human decision-making under uncertainty that are quite different from formal mathematical models of probability:

- Decisions are not *made* after gathering all the facts, but rather *constructed* through an incremental process of planning by successive refinement. This process has the cognitive advantage of allowing a complex problem to be solved with limited computational resources. Its disadvantage, when compared with formal decision analysis, is that an early decision is made with simplified, abstracted information about the alternatives; more detailed information that may be relevant to the decision is not considered until later. As a result, in our medical example, we observe that the risk of biopsy and the risk of false negative test results were underweighted in the decision between biopsy versus empiric amphotericin.

- Subjective probabilities, i.e. knowledge of likelihoods, are not *numbers*, subject to multiplication and addition operations, with errors and biases measurable as distances from the correct values. Rather, they are symbolic *descriptions* of numbers, expressing categorical and ordinal relations with certain focal or bounding landmark values. Even likelihoods that are verbally described as numbers are treated in the explanation as names of landmarks. If this view is correct, then the representativeness and anchoring heuristics are consequences of natural classification strategies for determining categorical and ordinal descriptions, respectively.

7 Acknowledgments

We have received helpful comments from Mike Wellman, Elisha Sacks, Peter Szolovits, Herbert Simon, and Bill Clancey.

8 References

1. Bobrow, D. G. (Ed.). (1985). *Qualitative Reasoning about Physical Systems*. Cambridge, Massachusetts: Bradford Books/MIT Press.

2. Clancey, W. J. (1985). Heuristic classification. *Artificial Intelligence* **27**, 289-350.

3. Cohen, P. R. (1985). *Heuristic Reasoning about Uncertainty: An Artificial Intelligence Approach*. Boston: Pitman Advanced Publishing Program.

4. Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

5. Ericsson, A. & Simon, H. A. (1984). *Protocol Analysis*. Cambridge, MA: MIT Press.

6. Feinstein, A. R. (1967). *Clinical Judgment*. Huntington, NY: Krieger Publishing Co.

7. Goldmeier, E. (1972). *Similarity in Visually Perceived Forms*. *Psychological Issues* **8**: 1; monograph 29. New York: International Universities Press.

8. Hayes-Roth, B. & Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive Science* **3**, 275-310.

9. Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

10. Kassirer, J. P. (1976). The principles of clinical decision making: An introduction to decision analysis. *The Yale Journal of Biology and Medicine* **49**, 149-164.

11. Kassirer, J. P., Kuipers, B. J., & Gorry, G. A. (1982). Toward a theory of clinical expertise. *The American Journal of Medicine* **73**, 251-259.

12. Kuipers, B. J. (1978). Modeling spatial knowledge. *Cognitive Science* **2**: 129-153.

13. Kuipers, B. J. (1979). On representing commonsense knowledge. In N. V. Findler (Ed.), *Associative networks: The representation and use of knowledge by computers.* New York: Academic Press.

14. Kuipers, B. J. & Kassirer, J. P. (1983). How to discover a knowledge representation for causal reasoning by studying an expert physician. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*. Palo Alto: William Kaufman.

15. Kuipers, B. J. & Kassirer, J. P. (1984). Causal reasoning in medicine: analysis of a protocol. *Cognitive Science* **8**, 363-385.

16. Lynch, K. (1960). *The Image of the City*. Cambridge, MA: MIT Press.

17. Miller, G. A. (1956). The magical number seven, plus or minus two. *Psychology Review* **63**, 81-97.

18. Pauker, S. G. & Kassirer, J. P. (1980). The threshold approach to clinical decision making. *New England Journal of Medicine* **302**, 1109-1117.

19. Pauker, S. G. & Kassirer, J. P. (1987). Medical progress: Decision analysis. *New England Journal of Medicine* **316**, 250-258.

20. Sacerdoti, E. D. (1977). *A Structure for Plans and Behavior.* New York: Elsevier North-Holland.

21. Raiffa, H. (1970). *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. Reading, MA: Addison-Wesley.

22. Schwartz, W. B, Gorry, G. A., Kassirer, J. P, & Essig, A. (1973). Decision analysis and clinical judgment. *American Journal of Medicine 55*, 459.

23. Shafer, G. (1976). *A Mathematical Theory of Evidence.* Princeton: Princeton University Press.

24. Shafer, G. & Tversky, A. (1985). Languages and designs for probability judgments. *Cognitive Science* **9**, 309-339.

25. Sternberg, S. (1966). High speed scanning in human memory. *Science* **153**, 652-654.

26. Szolovitz, P. and Pauker, S. G. (1978). Categorical and probabilistic reasoning in medical diagnosis. *Artificial Intelligence 11*, 115-144.

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

28. Weinstein, M. C. & Fineberg, H. V. (1980). *Clinical Decision Analysis*. Philadelphia: W. B. Saunders.

A Script Analysis of VP-20

To clarify the protocol analysis method, this appendix contains the “thinking aloud” portion of transcript VP-20 in which the biopsy decision was made. Line and paragraph breaks divide the transcript into meaningful small units. Each line is numbered so the analysis can refer back to it.

After examination of the transcript, a set of operators is devised to account for the major reasoning processes observed in the explanation. From the analysis of this excerpt, the operators and the objects they apply to are the following:

- REVIEW: list of facts, hypotheses, prerequisites or plan-steps.
- CONSIDER: hypothesis.
- COMPARE: quantities.
- PROBLEM: set of possible actions.
- CHOOSE: action

These operators are created from examination of the transcript prior to a formal theory of knowledge representation and inference for the observed reasoning. Thus, we can code a segment of the transcript as REVIEW if we observe the subject systematically considering all or most of the current disease hypotheses, even though we lack a complete theory of how such an iterative process takes place.

For the purpose of this study, the critical parts of the analysis are the identification of the PROBLEM and CHOOSE operators, where decision problems are identified and an alternative is selected. Our analysis rests on correct identification of those operators, and their distinction from other actions in the transcript.

A.1 Verbatim Transcript with Script Analysis

The script analysis is presented in italics, interleaved with the lines of the transcript.

L001 But he did apparently get along alright
L002 for a period of between 5 months before admission
L003 and 2 months before admission
L004 without any other information given.
L001-L004 REVIEW: facts

L005 So, I would assume that he might have had a good response.
L006 1.8 grams cumulative dose is not all that big of a dose,
L007 but maybe he had a response,
L008 because there is no further information.
L005-L008 REVIEW: facts

L009 So then he turns up just 3-4 days before admission with fever,
L010 diffuse process,
L011 again pancytopenia on peripheral count.
L009-L011 REVIEW: facts

L012 And so my thinking . . .
L012 CONSIDER: ?

L013 and was treated with broad spectrum drugs
L014 without an effect in 48 hours.
L013-L014 REVIEW: facts

L015 Now that still may be a somewhat short time.
L015 CONSIDER: Current treatment OK

L016 Even if those were going . . .
L017 even that combination of drugs was going to have an effect,
L018 but I think that what is important here
L019 is to make a specific diagnosis,
L020 and the approach is a lung biopsy.
L016-L020 CHOOSE: Invasive biopsy over No change in therapy.

L021 I assume that they have already x-rayed his sinuses.
L022 (I) question whether there is something still present in the sinuses.
L021-L022 CONSIDER: infection in sinuses. Ruled out.

L023 So, I think the first approach would be a lung biopsy
L024 and then the question comes up
L025 should this be an open biopsy
L026 or should it be a close biopsy.
L023-L026 PROBLEM: TBBx vs. OLBx.

L027 And then the question comes up
L028 how much of a risk he'll be for biopsy
L029 whichever way it's done.
L030 There's no information given about parameters
L031 such as platelet count,
L032 bleeding times,
L033 things of that sort which would be important to know
L034 in face of any plan biopsy.
L035 But I assume that those will be done and
L036 then if he needed platelets, he would receive them prior to biopsy,
L037 and if he . . .
L038 he would probably be transfused prior to biopsy because his
L039 hematocrit and hemoglobin . . . again with.
L040 So, all of that could be taken care of
L041 and that doesn't sound like that's going to be a major sort of a problem.
L027-L041 REVIEW: risks of biopsy.

L042 The other bit of information here is that he has a pericarditis
L043 both by EKG and physical examination
L044 and so the presumption here will be that this is part of the
L045 same process that's causing . . .
L046 the that's causing the the pulmonary infiltrates.
L042-L046 CONSIDER: pericarditis

L047 I suppose that one question might come up
L048 whether one might want to consider a pericardial tap
L049 as an initial diagnostic procedure,
L050 and that would be something that might be considered.
L047-L050 PROBLEM: pericardial tap vs. biopsy

L051 I probably would have gone directly for the lung biopsy
L052 because I see no reason he couldn't tolerate that
L053 from the information that we have here.
L054 He has a pO2 of 50.
L055 That can probably be brought up
L056 with delivery of oxygen.

L057 So, the general feeling from information given
L058 is that he can tolerate a lung biopsy with some appropriate therapy
L059 such as blood transfusions,
L060 prior to the time that it's done.
L061 So I would opt first for a lung biopsy
L062 and then keep the idea of a pericardial tap in mind.

L063 The situation about the tap is
L064 one would have to rely upon culture.
L065 We're not dealing basically with a high suspicion tumor candidate.
L066 We're considering infection is the primary event,
L067 and one's not going to be able to see anything right away,
L068 but would have to wait for culture.
L069 With lung biopsy one might see fungi,
L070 or whatever else there would be.
L051-L070 CHOOSE: biopsy over pericardial tap
L052-L060 REVIEW: risks of biopsy
L063-L070 COMPARE: time for culture vs. time for biopsy result

L071 So, I would elect probably
L072 to do this as a closed biopsy . . .
L073 transbronchial.
L074 The choice of closed versus open here
L075 is probably somewhat of a toss-up.
L071-L075 CHOOSE: TBBx over OLBx.

L076 However, with the types of things that one is considering here
L077 like fungal infection
L078 and I don't think
L079 that one has to think very hard about possibilities
L080 like protozoa
L081 because he's basically not immunosuppressed.

L082 White count . . .let me just get back to that for a moment.
L083 (I) question whether he is . . .
L084 white count was 1200

L085 with a total neutrophil count of 150.
L086 So he does have a pretty low count
L087 and in a sense he is suppressed
L088 for overcoming any type of infectious organism.

L089 Just quickly going down the lists of possible infections
L090 that would be considered here:
L091 fungal infection;
L092 bacterial;
L093 protozoa,
L094 that I probably put down lower on the list;
L095 TB,
L096 not really a strong consideration;
L097 viral infection,
L098 always a possibility;
L099 but no reason to think that here in particular.
L076-L099 REVIEW: disease hypotheses.

L100 So I would do the lung biopsy
L101 and would probably get a . . .
L102 an immediate wet prep from the biopsy . . .
L103 touch prep and see what that shows.
L104 We'd put some tissue up for culture
L105 then we'd put some tissue through the usual fixation process.
L100-L105 REVIEW: plan steps.

A.2 Final Script Analysis

Thus, the final result of the script analysis is a sequence of operators applied to various aspects of the description of the case.

L001-L004	REVIEW: facts
L005-L008	REVIEW: facts
L009-L011	REVIEW: facts
L012	CONSIDER: ?
L013-L014	REVIEW: facts
L015	CONSIDER: Current treatment OK
L016-L020	CHOOSE: Invasive biopsy over No change in therapy.
L021-L022	CONSIDER: infection in sinuses. Ruled out.
L023-L026	PROBLEM: TBBx vs. OLBx.
L027-L041	REVIEW: risks of biopsy.
L042-L046	CONSIDER: pericarditis
L047-L050	PROBLEM: pericardial tap vs. biopsy
L051-L070	CHOOSE: biopsy over pericardial tap
L052-L060	REVIEW: risks of biopsy
L063-L070	COMPARE: time for culture vs. time for biopsy result
L071-L075	CHOOSE: TBBx over OLBx.
L076-L099	REVIEW: disease hypotheses.
L100-L105	REVIEW: plan steps.

B Referring Phrase Analysis of VP-18

This appendix presents the complete sets of categorical, ordinal, and numerical range descriptions appearing in one of the transcripts (VP-18).

B.1 Categorical Descriptions

Categorical descriptions classify a quantity into one of a small set of categories without sharply defined boundaries.

. . . is notoriously a procedure that	**does not give accurate diagnoses**
. . . has had	**very little success** overall,
. . . which you would have	**a very high yield** of finding . . .
. . . would	**pretty much eliminate** it from the cause . . .
. . . considering	**how good we've gotten** in diagnosing that.
. . . which a transbronchial biopsy is also	**good for**,
. . . washings	**aren't any good**,
. . . would also be	**pretty well** diagnosed.
. . . I think he would not be	**a high risk**,
. . . but	**(a) satisfactory candidate** for bronchoscopy
. . . you('d) have	**a fairly high yield**,
However, he would still be	**a high risk person**,
. . . and	**not a very safe path to take**.
. . . the chances of . . .	**is actually small**.
. . . because	**it would be unsafe to proceed**.
I think his chance of mortality . . .	**is still very small**.
So,	**the mortality isn't very high**.
. . . which	**doesn't occur all that frequently**,
. . . the prognosis . . .	**is extremely poor**.
. . . your chance of survival is	**practically nil**.

B.2 Ordinal Descriptions

Ordinal descriptions describe a quantity in terms of its ordinal relation — greater than, less than, equal to — with another quantity.

although,	**not quite as good as** pneumocystis.
And since it's	**a safer procedure** then general anesthesia, . . .
the chances of . . . is	**probably not all that much greater**.
. . .	**more safely**
or have	**a better chance of getting a result**.
Because the patient can	**tolerate the procedure better**,
Pneumothoraces . . . actually are	**even less than that**.
So,	**that would have remained the same**.
His risk of bleeding	**is probably higher**.
and (?death from) bleeding	**less than that**.
that occurs you know	**less than a half percent of the time**.
I think his chance of . . .	**would have been increased**,
	more than the general population,
(it)	**would still be a safer procedure** than general anesthesia.
I think it still would be	**a safer procedure**.
I'd say	**the risk clearly outweighs** the risk of amphotericin.
. . . to find a specific answer . . .	**would clearly outweigh that risk**.
the risk of the bronchoscopy	**would outweigh** empiric treatment
I	**would rather find out** what I have **than** . . .

B.3 Numerical Range Descriptions

Numerical range descriptions specify a quantity in terms of a numerical measure. It is important to note, however, that these "numbers" function as *names* for values, rather than real numbers subject to arithmetic operations. The actual content and use of these descriptions resembles the categorical and ordinal descriptions.

... you should be	**well over 90%** of diagnosing pneumocystis.
... you have	**maybe a 70 - 80% chance** (of) making the diagnosis,
...	**between 60 and 70% chance** of making that diagnosis
... a fairly high yield,	**at least 60%** ...
...	**up to 90% chance** of getting that answer ...
... whether that	**small means 10%, 15%**, ...
It's still less than	**a half or tenth of a percent**.
His risk ... would still be	**about 5 to 7%**, ...
... but his risk ... would be	**probably somewhere between 5 and 10%**, ...
... that occurs you know	**less than a half percent of the time**.
...	**usually about 2 or 3%**.
... maybe	**somewhere between 5 and 10%**.

B.4 Possible Event Descriptions

The possible event descriptions are a low-resolution variant of categorical descriptions, used to determine whether an event should be included as an alternative in a decision or not.

ALWAYS: This event, though perhaps not absolutely certain, may be assumed to occur.

SOMETIMES: This event is one of a set of possibilities; its likelihood will need to be considered.

NEVER: This event, though perhaps not absolutely impossible, may be excluded from consideration.[3]

This is a selected excerpt from the transcript illustrating the usage, not a complete list.

L242	the potential bleeding complications scare me with invasive procedures.	SOMETIMES
L243	I think those are clearly the things that kill people.	
L244	you know,	
L245	these people generally don’t die from procedures	NEVER
L246	of . . .	
L247	generally don’t die from respiratory failure	NEVER
L248	following these procedures	
L249	‘cause we can ——	
L250	we can ventilate them,	
L251	we can support them.	
L252	But they do die from hemorrhagic complications,	SOMETIMES
L253	and that’s something that would frighten me,	
L254	but cardiopulmonary things I would just try to control	
L255	and do the biopsy.	

[3]
Chorus: “Never?”
Captain Corcoran: “Well, hardly ever.”
[H.M.S. Pinafore, Gilbert and Sullivan.]