Technical Report CMU-EPP-1990-03-03, Department of Engineering and
Public Policy, Carnegie Mellon University, Pittsburgh, PA, October 1989

Towards Process Models of Judgment Under Uncertainty (Progress Report)

Marek J. Druzdzel*

Department of Engineering and Public Policy
Carnegie Mellon University
Pittsburgh, PA 15213

October 28, 1989

Abstract

Empirical research on biases and heuristics has made a major contribution to the knowledge about human judgment under uncertainty. There is a growing body of data documenting situations in which human judgment violates the axioms of probability theory. Recently, several researchers have postulated the need for a theoretical framework that would explain the findings in terms of problem representation and cognitive processes.

This paper is a progress report of the research done in this direction. Its aim is to give a broad view of the ideas that arose during this research, even if they are speculative in nature and not supported with enough evidence. There are three hypotheses emerging from this work: **(1)** Newell and Simon's Theory of Problem Solving may provide a sufficient theoretical framework for the research on human judgment under uncertainty. **(2)** Process tracing methods, and especially protocol analysis have the power of providing data for such framework. **(3)** Bayesian belief networks are a promising candidate for a symbolic representation of problems involving uncertainty.

A preliminary analysis of verbal protocols of over 20 subjects working on various judgmental tasks was performed. A computational framework for modeling judgmental processes was created in which subjects' relevant knowledge was modeled by a large, unique for each subject belief network. When structuring the problems, a majority of subjects tended to instantiate nodes of that network and create deterministic scenarios of possible events influencing the event in focus. These scenarios resulted in problem spaces resembling incomplete probability trees. Computation of uncertainty seemed to be weighting of subjective likelihood of the considered scenarios. Of the two formal mechanisms for probabilistic reasoning within belief networks: belief propagation and generation of deterministic scenarios, the approach observed clearly resembled the latter one. The paper suggests several possible mechanisms that can account for discrepancies between human judgment and the probability theory.

*This work was supported in part by the National Science Foundation under grant IRI–8807061

1 Introduction

The problem of uncertainty is of particular importance in any situation involving decision making. It is hard to discover any real world decision problem where there is not at least slight uncertainty about the outcomes. In most real problems the key to a good decision is the proper assessment and processing of uncertainty within the analysis [14].

Of the calculi developed for dealing with uncertainty, the most widely used is probability theory. Uncertainty in probability theory is measured by a real number between 0.0 (impossible event) and 1.0 (sure event), called probability. The notion of probability can be derived from a set of axioms of rational decision making. Violating these axioms leads to inconsistency in probabilities and can be demonstrated to cause a sure loss. Following the axioms, on the other hand, guarantees consistency.

Numerous studies done by behavioral psychologists have shown that there are serious discrepancies between the way people estimate and process uncertainty and the way the axioms of probability theory prescribe it. These studies resulted in a discovery and a convincing demonstration of people's propensity for use of empirical, intuitive strategies (*heuristics*) in their estimations of likelihood of uncertain events. The heuristics, of which *representativeness*, *availability*, and *anchoring and adjustment* are the best known, make estimation of probabilities computationally tractable, but often result in systematic violations of the probability axioms (*biases*).

Comparing human judgment against the formal probabilistic methods, and investigating which situations lead to significant discrepancies between these two, can benefit those who desire to make consistent decisions. The large body of research conducted in the last 20 years led to discovery of many such situations and mechanisms that are responsible for human inconsistency ([10] is an anthology providing a thorough review of the main findings). Recently, several researchers expressed the need for a theoretical framework for this research.

Wallsten points out:

> ... the point is that research which focuses on any number of heuristics, simply asking the conditions under which each appears, stops short of some of the most interesting and useful questions. Classification is clearly very important, but there is the danger in such research, frequently succumbed to, that the description of the phenomenon is taken as its explanation, and underlying theoretical principles are not sought. A search for underlying principles rather than a fixation on disparate phenomena will lead to conclusions of greater generality and usefulness [20, page 22].

He suggests further that the theoretical framework for the research on human judgment under uncertainty needs to include human problem representation, and concern both the semantic and the probabilistic aspects of the experimental situations. It should indicate how the problem gets structured, which dimensions of the information provided will be encoded, and how these features will be used. It should furthermore be capable of predicting the range of judgmental effects that are observed [20].

Ginossar and Trope refer to the same issue in the context of studies of the base–rate fallacy, i.e. subjects' insensitivity to base rates in probability updates. Results of several earlier experiments described in the literature show that human judgment is not always dominated by the use of the representativeness heuristic over statistical reasoning, as postulated by the heuristic and biases approach.

> The heuristic judgment approach has convincingly demonstrated that representative case information often dominates statistical base–rate information. However, this approach has not specified the conditions that can weaken and even reverse this dominance. What is needed, then, is a theoretical framework that will not only retain the explanatory power of the heuristic judgment approach but will also delineate the general conditions that determine the application of statistical or nonstatistical rules [6, page 465].

In attempt to achieve this goal, Ginossar and Trope apply a problem solving approach to judgment under uncertainty. They describe six experiments (variants of the Kahneman and Tversky's *lawyer–engineer* [12] and *blue–green cabs* [11] problems) in which subjects' behavior, and in particular application of intuitive statistical rules, is postulated to be related to general problem solving principles. However, this research is limited to pointing out the existence of general problem solving principles and does not attempt to explain issues like problem representation, strategies, processes, etc.

The direction taken by this research is studying processes employed by humans in judgment under uncertainty. The motivation for the research on process level models comes from several directions. One of them is an extension of the work done in behavioral psychology. Progress in understanding of the mechanisms by which people encode and process probabilities, may make it possible to hypothesize when particular heuristics will be used and when biases most probably will occur. This may lead to a more effective design of debiasing procedures or to a better elicitation and communication of uncertainty (e.g. in risk communication).
The process level approach is also motivated by the research on Decision Support Systems (DSS). Bayesian probability theory has been shown to outperform other formalisms for dealing with uncertainty[1] and, as indicated by the findings of the heuristic and biases research, also to outperform humans. Therefore there is a growing interest in building DSS that employ probabilistic inference. One of the criticisms of this approach is that probability theory is not readily understandable by humans. Ability to distinguish correct from erroneous advices of an DSS has been shown to be critical for the combined performance of that DSS and its user [13]. It is therefore essential that either the user has a correct mental model of the DSS or that the DSS is able to explain its reasoning in a way that is intuitive for the user. Through their explicit character, process level models have the capability of demonstrating where the differences between human and Bayesian reasoning are and how they originate. They may allow to express human reasoning (with its imperfections) in probabilistic terms, hence provide the necessary framework for the design of user interfaces.

[1] Wise and Henrion [21] compare several popular mechanisms for treatment of uncertainty and demonstrate that only in the best case will they perform as well as probability theory.

Studying cognitive processes, at last, might give interesting ideas for artificial reasoning schemes.

The remainder of this paper is structured as follows. The first four sections form a concise introduction for readers who are not familiar with the background domains of this research. Sections 2 and 3 describe briefly the main ideas of the Theory of Problem Solving and the assumptions underlying the method of protocol analysis. Sections 4 and 5 present the concept of Bayesian belief networks and explain the principles of two mechanisms of belief updating within the networks: belief propagation and generation of deterministic scenarios. Readers familiar with the above concepts can proceed directly with Section 6.

Section 6 describes the framework for studying human reasoning under uncertainty developed in the course of this research. Section 6.1 characterizes the experiments performed. Sections 6.2 and 6.3 describes modeling the knowledge of the subjects that is relevant to the problems by means of belief networks. Section 6.4 elaborates problem structuring within the available knowledge and Section 6.5 presents hypotheses concerning the computational part of the judgment process (i.e. producing numerical estimations). Section 6.6 describes how heuristics fit in the framework and Section 6.7 how the framework models discrepancies between human and probabilistic reasoning. Section 6.8 discusses possible future directions for this research. Section 7 summarizes the main ideas resulting from this research.

2 Theory of Problem Solving

Newell and Simon's Theory of Problem Solving [15] belongs to the class of information processing approaches to cognition. It involves studying human behavior in solving various tasks and modeling thinking as processing of information. This is based on the view of a human as a processor of information and the assertion that a human brain shares with computers the property of being a physical symbol system.

Problems posed to an individual are solved within an external environment that the problem solver is placed in. This environment, called *task environment*, determines to a large extent the behavior of the problem solver. Problems from that environment are conceptualized in terms of a *problem space* that can be searched selectively for a solution. Information about the general intelligence of the problem solver combined with the information about his or her knowledge in relevant domains is sometimes sufficient to predict what problem space he or she will create. The task environment is the overwhelming determinant of the problem space. The structure of the problem space determines the possible processes that can be used for problem solving. The solution process is effectively modeled as moving through the problem space from the initial to the final state through a series of intermediate states. Reaching the final state is equal to completion of the solution process.

Search through a problem space for a solution, except for a limited number of problems for which a solution method is well known, is guided by empirical, intuitive strategies, called heuristics. Although heuristics do not guarantee a solution, they aid the problem solver in his or her limited cognitive capabilities by making the search problem computationally tractable. A problem solver is capable of acquiring knowledge about the solution process by discovering and learning new strategies to solve the problem.

One of the few invariants over problem solver and task is the serial character of the human information processing, at least at the level of higher cognitive processes. Another important invariant is its goal oriented character. The problem solver attempts to attain a goal and also has a test available to determine whether the goal has been achieved.

Except for a limited number of probabilistic tasks encountered mainly in probability textbooks, most of the problems involving uncertainty are ill structured, in the sense that they are unbounded in their probabilistic dependencies, they lack a clear, widely agreed upon problem space and definite criteria for testing any proposed solution. Studies of cognitive processes in ill structured domains indicate that humans change ill structured problems into well structured problems in the process of preparation (structuring) for solving them. Furthermore, much problem solving effort is directed at structuring problems and only a fraction of it at solving problems once they are structured [18]. Structuring a problem is equivalent to constructing a problem space to work within. Physicians, for example, have been observed to make the process of medical diagnosis a well structured problem from an ill structured one by early hypothesis generation. They find in a relatively early phase of diagnosis a small number of plausible diseases (hypotheses) that explain the symptoms, and then solve the problem: "Which of these hypothesis is right?", which is a well structured problem [7].

3 Protocol Analysis

Studying human cognition at the level of information processes poses methodological problems. A single move between two states of the problem space conceptualized by a human problem solver can take as little as a few seconds. Several process tracing methods have been used in studying human problem solving behavior, for example eye movement tracing, collection of concurrent verbal protocol, denying and registering perceptual access to a stimulus, interrupting ongoing activities with a signal to give a request for the contents of attention. Analyzing the data obtained by these methods aids in the reconstruction of the thinking process.

Of the process tracing methods, the most popular has been recording verbal protocols ("think aloud" protocols). Ericsson and Simon [4] give an overview of the assumptions and available evidence on suitability of protocol analysis to studies of problem solving. It is generally assumed that during the verbalization the subject reports some subset of the information that is held in his short term memory during performance of the task, provided that this information is encoded in a way that makes it easy to verbalize[2]. Thinking aloud has been observed not to interfere with or significantly modify task performance, provided that the subject assigns first priority to performing the task and that the instructions to think aloud are formulated neutrally, i.e. they do not direct the subject to produce specific

[2]Visual stimuli, for example, are not easily verbalized.

kinds of information. Studying verbal protocols has been applied in several domains, from simple, well defined tasks to models of real–life problems, like decision making and medical diagnosis.

4 Bayesian Belief Networks

Bayesian belief networks[3] (BBN) [16] are a mathematical formalism allowing for explicit representation of random variables and probabilistic dependencies among them. Formally, BBN are directed acyclical graphs in which nodes represent random variables. An arc between any two nodes denotes the existence of direct probabilistic influence between them. The strength of this influence is quantified by a conditional probability. Formally, the direction of the arc is arbitrary, but usually it is chosen to denote causal influence between the nodes it connects[4]. This structure allows for a natural form of expressing the relationship of causality as a direct dependency.

BBNs model well the notion of relevance, independence, and conditional dependence, and allow modeling of human qualitative reasoning about probabilities. They also allow one to model such essential aspects of human reasoning as the ability to integrate predictive and diagnostic reasoning, the ability to discount correlated sources of evidence, and inter–causal inference between alternative causes of an event ("explaining away") [8]. Finally, they offer a mechanism for a normative, probabilistic approach to belief updating, while being less rigorous towards inconsistencies than the probabilistic notation.

An example of a BBN is given in Figure 1. This BBN describes a model of the domain of different causes of sneezing of an individual. All variables in this model are binary (*present/absent*). Sneezing can be caused by *Cold* or *Allergy* to *Cats*. Seeing *Paw Marks* on the floor can raise suspicion about presence of a *Cat* in the house, but hearing *Barking* from the room next door can suggest that the *Paw Marks* might have been left by a *Dog* after all. Each node that has no incoming arcs (*Cold*, *Cat*, and *Dog*) is described by a prior distribution of its outcomes (e.g. fraction of houses that keep *Cats*). A directed arc from the node *Cat* to the node *Allergy* contains a description of conditional probabilities of outcomes of *Allergy* given outcomes of *Cat*. If a node has more than one incoming arc (e.g. *Sneezing*), the conditional likelihood of its outcomes is described by a joint probability distribution of all influencing nodes (*Cold* and *Allergy*). Since the probabilities are discrete, the joint conditional probability of *Sneezing* is a three–dimensional matrix, where the three dimensions are outcomes of *Cold*, outcomes of *Allergy*, and outcomes of *Sneezing*. The elements of this matrix will be probabilities of a given outcome of *Sneezing* given the outcomes of *Cold* and *Allergy*.

Notice that it is possible to reason within this BBN in both predictive and diagnostic direction, both qualitatively and quantitatively. If a *Cat* is present, we can easily determine the likelihood of *Paw Marks*, *Allergy*, and *Sneezing*. On the other hand, if we observe *Paw*

[3] Also called *belief networks*, or *Bayesian networks*.

[4] If there is no causal influence between the nodes, the direction of the arc is picked for convenience.

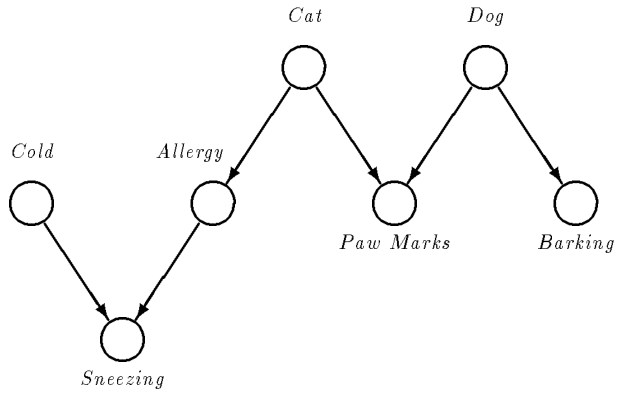

Figure 1: An example of a BBN

Marks, we know that they have been left by either a *Cat* or a *Dog*. It is also possible to compute the probabilities of both. If *Sneezing* has been observed, *Cold* and *Allergy* become correlated — whenever there is more evidence for *Allergy*, *Cold* becomes less likely and opposite ("explaining away"). A similar effect appears when the *Cat* has been observed. *Allergy* and *Paw Marks*, correlated before observing *Cat* become independent once *Cat* has been observed. Reasoning withing a BBN is based on a well defined formalism and is probabilistically sound.

5 Belief Propagation and Scenario Thinking

Consider a random variable X and its distribution function $F(X)$. A *deterministic scenario* is any instance of this variable. For example, deterministic scenarios of a discrete variable "disease" are its instances "disease present" and "disease absent". The probability function of this variable will have two values, p for "disease present" and $1-p$ for "disease absent". A continuous variable "human height" has infinitely many values and each of them is a possible scenario[5].

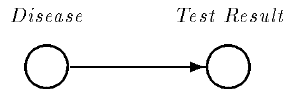

Figure 2: BBN for the disease/test problem

Figure 2 shows two interacting random variables: "disease" and "test result". The variable "disease" influences causally the variable "test result" — the result of the test depends on

[5]It is a common practice to discretize continuous variables. The variable "human height" can be approximated by a discrete variable, for example a variable with three values: "short", "medium", and "tall".

whether the disease is present or not. For simplicity, "test result" will also be a binary variable with two instances: "test positive" and "test negative".

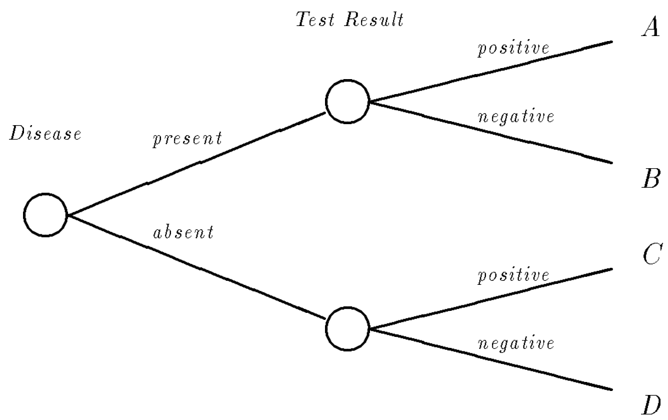

Figure 3: Probability tree for the disease/test problem

A useful way of representing scenarios involving several interacting random variables is a probability tree. An example of a tree for the variables "disease" and "test result" is given in Figure 3. Each node in such a tree represents a random variable and each branch originating from that node, a possible outcome of that variable (scenario). Note, that there are four possible scenarios involving the two variables in Figure 3:

S1 The disease is present and the test is positive.

S2 The disease is present and the test is negative.

S3 The disease is absent and the test is positive.

S4 The disease is absent and the test is negative.

Suppose that we are placed before the problem of determining the likelihood that a patient is ill, given that the prevalence rate of the disease in question among the population of his age is 1:10000, and that both the sensitivity and the specificity of the test are 99% (i.e. given a positive test result, there is a 0.01 chance that it is a false positive and given a negative test result, there is a 0.01 chance that it is a false negative result).

Let d stand for "disease present", $\overline{d}$ for "disease absent", t for "test positive", and $\overline{t}$ for "test negative". This problem can be solved in two different ways. The first is by an algorithm called belief propagation. We know that the test is quite accurate, we also know that it came out positive, so effectively it increased our belief in patient's disease. In other words, the test result supplied diagnostic support for the presence of the disease. The new probability that the patient has the disease, computed by means of the Bayes theorem is:

$$p(d \mid t) = \frac{p(t \mid d) \cdot p(d)}{p(t \mid d) \cdot p(d) + p(t \mid \overline{d}) \cdot p(\overline{d})} = \frac{0.99 \cdot 0.0001}{0.99 \cdot 0.0001 + 0.01 \cdot 0.9999} = 0.0098$$

An alternative algorithm, using the probability tree representation of the problem of Figure 3, approaches the problem by considering probabilities of deterministic scenarios of possible events and then aggregating them in order to find the probability of the disease. Given the fact that the test turned out to be positive, scenarios *S2* and *S4* became impossible. Only two scenarios remained: *S1* and *S3*. The disease is present and the test is positive (*S1*) or the patient is healthy and the test is positive (*S3*). The probability of scenario *S1* is equal to the product of the original probability that the disease is present and the conditional probability that the test will be positive given that the disease is present.

$$p(S1) = p(d) \cdot p(t \mid d) = 0.0001 \cdot 0.99 = 0.000099$$

The probability of scenario *S3* is equal to the product of the original probability that the disease is absent and the conditional probability that the test will be positive given that the disease is absent.

$$p(S3) = p(\overline{d}) \cdot p(t \mid \overline{d}) = 0.9999 \cdot 0.01 = 0.009999$$

We weight these two scenarios against each other and find out that scenario *S1* is 101 times less likely than scenario *S3*. This process, equal to normalization of the probabilities of the remaining scenarios (since the remaining two scenarios exhaust all possibilities, their sum has to be 1.0), yields $p(d) = 0.0098$, the same answer as when using the belief propagation scheme.

Readers accustomed to the decision analytic tools, will notice that these two representations of interacting random variables and the two methods of propagating updates are analogical to influence diagrams and decision trees. Influence diagrams are essentially a generalization of BBN which includes nodes for decisions and values (utilities). Decision trees are essentially probability trees extended with decision branches and values of outcomes. Any given decision problem can be represented by a decision tree and an isomorphic influence diagram. Both are equivalent in the sense that they produce the same normative, Bayesian result, yet using different methods. Update propagation algorithms that were recently developed for BBN and influence diagrams are based on methods that are analogous to belief propagation [16, 17]. On the other hand, the algorithms for decision trees are based on deterministic scenarios. Each branch of a decision tree corresponds to a deterministic scenario of events and has an associated value of the outcome. The value of the root of the tree is found by taking a weighted average of the values associated with all scenarios in the tree.

6 Framework for Modeling Human Inference

This section describes a preliminary framework for modeling human inference in the presence of uncertainty. The development of this framework has been approached from the perspective of Newell and Simon's Theory of Problem Solving. It is suggested that this theory is capable of addressing the main questions posed to a theoretical framework for judgment under uncertainty, and is also capable of providing a computational model of judgment, useful for applications in DSSs, such as user interfaces.

Section 6.1 describes briefly the experimental work that was conducted in the course of this study. Sections 6.2 and 6.3 discuss modeling the knowledge of the subjects that is relevant to the problems by means of BBN. Section 6.4 elaborates problem structuring and Section 6.5 presents hypotheses concerning the computational part of the judgment process (i.e. producing numerical estimations). Section 6.6 discusses how findings concerning judgmental heuristics fit in the framework and Section 6.7 how the framework models discrepancies between human and probabilistic reasoning.

6.1 Allergic Sneezes and Disabled Presidents

A possible reason for the difficulty with developing a process model of human judgment is in the methodology that has typically been employed. The majority of experiments that have been conducted provided data that reflected the end product of the probability estimation process such as likelihood ranking between two or more uncertain events or the probability estimate of an uncertain event. These data give little insight about how the problem was solved or even whether the question was well understood. Generally, the likelihood of an event can be judged by means of some heuristic mechanism, but the subject may also happen to know the answer (he or she had solved the same or a similar problem before or had previously heard or read about it), may decide to guess it (lack of time or lack of motivation), or may decide to decompose it into even more elementary events, hoping to be able to aggregate their likelihoods back. It is hard to determine from the final answer only which of the above four cases took place and in many cases, determining what really happened is rather impossible. Applying process tracing methods and in particular collecting and analyzing verbal protocols of judgmental tasks has the power of providing a better insight into human judgment.

To allow tracing of judgmental processes simple real–world problems involving probability estimations and updating belief in the light of new evidence have been presented to fellow graduate students at Carnegie Mellon University. The problems were chosen in such a way that they could naturally be decomposed in more elementary problems. In this way, decomposition and aggregation of events could be observed.

The first problem involved Henrion's *Sneeze Example* [8] in which a person suffering from frequent colds and allergy to cats, both resulting in sneezing, finds himself in a new house. The questions posed to subjects involved the probability of that person suffering from an incipient cold after he indeed starts sneezing, after he observes paw marks on the floor (is there a cat in the house?), and after he hears barking from the room next door (were the paw marks left by the dog?) [2]. A simplified probabilistic model of this problem was presented in Section 4 (Figure 1).

The second problem involved estimating the probability of the current Vice–President of

U.S.A. (Dan Quayle) becoming President before the end of the term of the current President (George Bush). Of course, this is implicitly referring to the probability that the current President becomes unable to fulfill his function.

Other (informal and not recorded) experiments included estimations of the probability of economic recession, subjects of upcoming examination, and others.

Several protocols collected during these experiments were typed and analyzed, so far in a rather informal manner, to allow for making initial observations. Samples of the protocols from the above experiments will be used as an illustration throughout the remainder of this section.

6.2 BBNs as a Tool for Modeling Subjects' Knowledge

In the traditional approach to probability theory, the existence of cause and effect relationships between random variables is immaterial. Human thinking, on the other hand, has been observed to rely on causal schemas — perceiving sequences of events in terms of causal relations. Several experiments performed by Kahneman and Tversky [19] indicated that subjects heavily relied upon previously formed causal schemas in determining what information they would use to make judgments. Similarly Einhorn and Hogarth [3] pointed out that causality plays an important role in judgment under uncertainty. An important conclusion emerging from these findings is that any theoretical framework for modeling human behavior should contain mechanisms allowing for a natural representation of causal relationships.

Graphs are probably the most common representation of conceptual dependencies, expressing directly and qualitatively the dependence relationships, meeting the requirements of explicitness, saliency, and stability. They have been used for portraying human memory, for example in semantic networks. BBNs are a form of graphs modeling probabilistic dependencies between the nodes. It seems, at least in the first approximation, that human subject's beliefs about causal and probabilistic dependencies within a certain domain can be modeled by a BBN.

A BBN modeling subject's relevant knowledge will generally be unique to each person, although due to similarities in people's environment, large parts of individual networks may be similar to each other. For practical purposes (modeling), I will assume that it is possible to draw a boundary between domains by considering only relevant beliefs and excluding those events and dependencies that are remote in nature. An example of an event that I find remote in nature to the domain of American Presidency is the current position of Venus in her trajectory around the Sun. I should stress here that there may exist a person who finds the position of Venus relevant to this domain. In such a case, this event does belong to his or her relevant network of beliefs. A network of beliefs in a certain domain is interconnected with networks modeling beliefs in other domains. The total network of beliefs is large but probably not infinite.

Modeling human relevant knowledge by means of a BBN is only a certain abstraction, a convenient approximation. A BBN neither resides in human memory, nor it is fully conceptualized or constructed as such in the thinking process. It is a set of individual beliefs about the surrounding world glued together to form a network. It is a convenient portrait of all relevant pieces of knowledge about the environment, the picture of what could possibly be constructed by the subject given what he or she knows.

A person does not have all his or her relevant knowledge readily accessible. Thinking about a problem, a person will typically recall only a small number of relevant events. Neutral prompting will generally not result in enumerating all elements of subject's beliefs. It is, however, theoretically possible to elicit large parts of his or her network by directed questioning over a longer period of time. Again, the beliefs are generally dynamic and some of them may change even during the elicitation process, for example when the subject notices an inconsistency in his or her beliefs.

6.3 Individual BBN as a Universe of Problem Spaces

A subject's beliefs modeled by a BBN can be looked at as a universe of all possible problem spaces he or she can create when attempting to solve a problem in that domain. A problem space will always be a subset of the relevant beliefs and will consist of all the elements of the network (events) and their interconnections (probabilistic dependencies) that the subject considers when solving the problem. A subject can modify his or her problem space by adding or dropping events or dependencies he or she finds relevant or irrelevant to the problem.

For example, the universe of problem spaces of a subject estimating the likelihood of a premature termination of the current U.S. President's term in office consists of all his or her relevant knowledge about the matter, such as his age, health condition, knowledge about assassinations of former American Presidents in office, the political enemies of the current president, his foreign enemies, effectiveness of the CIA, effectiveness of the body guards, effectiveness of the secret police of the foreign countries that he visits, the President's political affiliation, the frequency of his trips, his habits of entering the crowd during his trips, etc.

For each elementary event there is some idea about its prevalence, and for each two events there is some idea about the strength of influence between them, in whatever way it is cognitively encoded.

6.4 Giving the Problem a Structure: Creation of a Problem Space

I observed that when reasoning about the likelihood of an uncertain target event, subjects first tried to fit the problem into a structure they have previously seen or learned about. Subjects who had some but not enough training in probability theory, often tried to apply known mathematical techniques, for example, multiplication of numbers or probability updating by means of the Bayes theorem. In the protocol below, the subject attempts to apply his knowledge about statistics in the Sneeze Example:

E347: **What's the probability that he has a cold?** (pause)
E348: **It's just** ... (pause)
E349: **What's this** ... (pause) **ahm** ...
E350: **It's an exponential distribution?**
E351: **Time to first pro ... time to first** ...
E352: **Is it a Poisson?** (pause)
E353: **This likelihood till he gets ... to a thing is** ... (pause)
E354: **grows and grows and grows?**

E362: **I'm not sure how to guess that probability** ...
E363: **Cause I don't know what's the likelihood of** ... (pause)
E364: **How is the distribution of Harry's getting colds?**

When there was no "ready–made" structure that could fit the current problem, they proceeded with constructing a structure for the problem.

Most of the subjects attempted to investigate what other events possibly influenced that target event. This search for possible causes was aimed at selecting an event or a set of events that would be further used to estimate the likelihood of the target event (*event selection* processes). Event selection is assumed to be performed by following diagnostic and causal links starting from the target. Once a subject's attention had been focused on a candidate event, he or she then investigated how this event could influence the target event (*path building* processes). Event selection and path building processes were used complementarily. These two processes gave the problem a structure. As an illustration, we will look at the following protocol piece from the Sneeze Example.

C288: **Aha! Paw marks on the floor.**
C289: **Now, here we go again, because paw marks do not mean cats, right?**
C290: **Are these cat's paw marks, are these dog's paw marks?**
C291: **I mean, the guy could keep a squirrel, you know, or it might be a rat, who knows?**
C292: *How and in what way does seeing the paw marks affect the probability of Harry's incipient ... incipient cold you estimated in the question of the previous page?*
C293: **Well, if the paw marks are cat paw** ...
C294: **It really doesn't affect it. It mean** ...
C295: **It stays the same question.**
C296: **Is ... are these ... these are cat paw marks, then the probability is close to zero.**
C297: **If it's not a cat paw marks, then the probability of the cold causing it is close to one, right?**

Subject reads the information about paw marks being observed on the floor. She quickly finds events that are possible causes of *Paw Marks*: *Cat*, *Dog*, *Squirrel*, *Rat*, and *Any Small Animal*. She further analyses which of these events and how are influencing the target event (*Cold*).

When creating a problem space for the President problem, a subject can think about him, for instance, as being assassinated, getting seriously ill or causing a scandal. Each of the three is causally related to the target event: premature termination of the presidency. Another problem space may include a nuclear war, a hunting or an airplane accident, etc. A person who believes that the President is miraculously protected by an extra–terrestrial power just because of the fact that he is a Republican ("Republicans don't die in office"), creates a problem space consisting of just two random events: the *divine protection* with a very strong causal influence on the event *President's death*. There is also a problem space possible with one single event, namely premature termination of presidency. A subject may decide that the only information relevant to the problem are the base rates of American Presidents terminating their presidency prematurely[6].

Example of structuring the President problem by one of the subjects is given underneath.

> A2: **Well, the first thing that comes to my mind is: what could be the reasons of him stepping down.**
> A3: **And the picture of ... has ... has always been his health ...**
> A4: **I mean, there's nothing wrong with his health, but the picture is: his health,**
>
> A8: **Ah ... So now the second question is:**
> A9: **What about ah ... scandal?**
> A10: **He is certainly surrounded by a lot of scandal.**
>
> A19: **So, I've just eliminated two reasons ...**
> A20: **And now I'm just searching ... for what other reasons could someone step down.**
> A21: **He could be killed.**

The structure created can be thought of as the subject's problems space, selected paths through the subject's knowledge (modeled by a BBN). These loose paths or scenarios, all rooted in the target event, show how the likelihood of the target is influenced[7].

Human thinking at the level of higher cognitive processes has been observed to be predominantly serial [15]. The capacity of Short Term Memory poses another limitation and determines the maximum number of items that the subject can focus on at the same time. This limitation has been observed to lead in problem solving to search strategies resembling Depth First Search rather than Breadth First Search (e.g. Progressive Deepening in Chess [15]). Serial Depth First Search in a graph inevitably leads to scenario–like reasoning.

Whatever structure will be given to the problem, it's details are to a certain extent predictable, in the sense that they will be rather consistent with the beliefs of the subject and correlated with the widely accepted common sense beliefs. Creation of the problem space is driven by various constraints. In the beginning of structuring the President problem, the subject is constrained by what he or she knows of the American law. A President will

[6] It is arguably the most reliable strategy to solve this problem.

[7] Note that in the simplest case in which the BBN consists of only one node, scenarios are just the possible outcomes of that node.

finish his term prematurely if he is unable to fulfill his function, is impeached, or resigns. Being unable to fulfill his function can be caused by a finite set of events, namely a disease, accidental or intended injury or death, or imprisonment. So, the problem is ill structured in its entirety, but quite well structured and defined when divided into smaller modules (see also [18]).

6.5 Computation of Uncertainty

Once the problem is structured, the tree of paths constituting the subject's problem space is used for the computation of uncertainty. This phase takes a very short time and yields scanty protocols. It seems like the computation of uncertainty is almost instantaneous.

Just as the task of structuring the problem, computation of uncertainty within a newly created problem space is subjected to various constraints. The main constraints are those rules of the probability calculus that the subject knows and finds applicable. An example is the restriction that the probability can never be lower than 0.0 and larger than 1.0 and that probability 0.0 means an impossible event and 1.0 means a sure event. Other constraints are combinations of probability calculus and domain knowledge. Most of the subjects realize that having two causes of possible termination of presidency makes this event more probable than having only one cause. Also, most subjects realize that if the probability of the event "President is shot at" increases, then with other things unchanged, the probability of the event "President killed" also inevitably increases.

In the initial state the problem space has no likelihoods associated with different elementary events, In the final state the target event is in a stable state, i.e. the subject believes that the current problem space includes all significant events, that the likelihood of the target event is estimated correctly, and that no belief update inside the problem space will change it significantly. The criterion for reaching the final state is vague, even to the subjects themselves. Questions in the final stage of the experiment: "Is this enough?", "Shall I go on?", "Can I stop now?" directed towards the experimenter are quite common.

It is likely that uncertainty estimation is done by assessing strength of influence of each path on the target event and aggregating these strengths somehow (e.g. weighting). This is suggested by the tendency observed in the majority of subjects to consider deterministic scenarios of the possible events[8].

> C318: *How probable is it that Harry suffers from an incipient cold?*
> C319: **Still stays the same.**
> C320: **Eh ... If there happens to be a cat ...**
> C321: **You know, the presence of dog does not ... does not exclude the presence of a cat.**
> C322: **And the paw marks ... We never say anywhere that the paw marks were actually caused by the dog.**
> C323: **So, no conclusion.**
> C324: **The same conclusion as before.**

[8] And often even refusing to aggregate them in a somewhat unfamiliar, artificial domain (like the Sneeze Example).

C325: **If there is a cat, close to zero,**
C326: **if there's no cat, close to one.**

A plausible way of computing, consistent with the anchoring and adjustment heuristic, would be concentrating on the strength of the most plausible scenario and adjusting its estimated likelihood for all the other possible scenarios. Consider the following protocol as an example:

A44: **I'd say 10 percent.**
A45: **Now, stop.**
A46: (E: Of being shot or of stepping down?)
A47: **Well, 10 percent of being shot and**
A48: **I just don't see ... I mean the ... No probability is zero ...**
A49: **I don't believe any, you know, probability is zero, but ...**
A50: **And the 10 percent was completely arbitrary.**
A51: **It's just that, you know, I looked back and, you know, being shot is not at all incredible and on the other hand very likely.**
A52: **Ahm ... And I went through the other two things that I could think of that would be a problem**
A53: **and neither of those seemed to me ...**
A54: **Which were they?**
A55: **Oh, his health!**
A56: **Now, they're not zero,**
A57: **so why don't I add a couple of percentages with them,**
A58: **you know I'll add 5 percent.**
A59: **OK. So, I'll say 15 percent.**
A60: **And now I say: do I really believe 15 percent?**
A61: **15 percent strikes me as high.**
A62: **I'll go back to 10 percent total.** (laughing)

This way of computing beliefs seems to be consistent with the finding that linear models give a good approximation on human judgment [1]. While belief in an event computed by means of probabilistic methods is a complex function of beliefs in relevant events influencing that event, weighting selected scenarios against each other can easily be approximated by a linear function of outcomes that determine the choice of those scenarios (i.e. outcomes that make them plausible). In the example demonstrated above, it could be represented as the linear function: $Bel(Terminates) = C_1 \cdot Bel(Killed) + C_2 \cdot Bel(Scandal) + C_3 \cdot Bel(Ill)$, where Bel denotes subjective probability of the event[9], and C_1, C_2, and C_3 some weights.

6.6 Heuristics

The framework just presented is compatible with the findings of behavioral psychologists concerning judgmental heuristics. Heuristics in this framework are intuitive procedures

[9] We use Bel instead of p to stress that the former one does not necessarily conform to the axioms of probability.

aiding in searches through the BBN, likelihood judgments of elementary events, and estimating the strength of dependencies. Choice of the events forming the problem space, estimating the prior probability of elementary events and their conditional dependence on other events, estimating the likelihood of an entire scenario or parts thereof, can very plausibly be determined by their availability to memory and their representativeness for the problem described.

The following piece of protocol shows that the subject created a mental image of the President and matched it against her image of a President who could possibly cause a scandal (*representativeness*), in order to estimate the likelihood of the event: "current President will cause a scandal".

> A9: **What about ah ...scandal?**
> A10: **He is certainly surrounded by a lot of scandal.**
> A11: **Ah ...Scandal** (pause) **is not a possibility either, seems to me ...**
> A12: **Eh ...because he ...ah ...**
> A13: **I mean, I sort of wanna say he's sort of preppy ...**
> A14: **I mean, he ...he comes from a proper family,**
> A15: **he's very wealthy ...**
> A16: **ah ...and he ...he strikes me as someone who's above board,**
> A17: **I mean who'd not, I mean ...That's certainly his image.**
> A18: **He's not someone who would stoop to do anything that was not completely above board.**

The following piece shows recalling of past events that are analogous with the target event. The ease with which such events come to mind seems to be crucial in likelihood judgment (*availability*). Note also that the subject attempts to decompose this event (line A30), and then gives up to estimate its probability heuristically.

> A26: **Ah ...Now I have to sort of say how likely and I have to figure out ...ah ...probability for ...for his ...his getting killed.**
> A27: **And ...Now I'm sort of searching back to other Presidents who've been wounded or killed in the past ..**
> A28: **or ...or public leaders and it's sure not unlikely.**

> A30: **So, I'm thinking about the kinds of questions: will ...will Bush raise taxes and the fact that you look back behind this what other Presidents have raised taxes.**
> A31: **But anyway ...OK**
> A32: **So, Kennedy was killed,**
> A33: **Raegan was shot,**
> A34: **Ford somebody trying to shoot ...**
> A35: **ah ...the other Kennedy brothers ...**
> A36: **Robert was killed ...**
> A37: **I would say, the chances are not that bad at all.** (laughing)

The following protocol fragment also shows the use of *availability* heuristic:

E381: **What's the probability that a person has both a cat plus a dog?**
E382: **I'd say the probability is low that a person has a cat plus a dog.**
E383: **Ahm ... No, it isn't!**
E384: **How many people do I know that have both cats and dogs?**

Use of *anchoring and adjustment* heuristic was demonstrated in the protocol fragment shown on page 16.

Another heuristic worth mentioning is *simulation* heuristic, observed by Kahneman and Tversky [10]. In the framework, it is an instance of a scenario, a path through a branch of the probability tree of the problem space.

6.7 Discrepancies with the Bayesian Probability Theory

Generally, there are four sources of observed discrepancies between judgmental reasoning and the probabilistic reasoning. The first is related to misunderstanding of the problem posed. The second is related to the data that the judgment is based on, i.e. subjects' qualitative and quantitative beliefs about probabilistic relationships among relevant variables. The third type is related to the process of structuring the problem, i.e. determining what approach will be taken in the solution process and what elements and properties of the domain are relevant. The fourth type results from errors in processing the data.

The cognitive processes observed in subjects' protocols were very similar to propagating updates by generating deterministic scenarios. This suggests that the general approach is not inherently different from the probabilistic methods. The example that was used for illustration of the method of deterministic scenarios in Section 5 consisted of only four scenarios. Generally, in a model with n discrete binary random variables there are 2^n possible scenarios. This is, for sufficiently large n, incomprehensible for humans. Due to cognitive limitations subjects are able to consider only a limited number of them, hence create an incomplete probability probability tree. If the tree were complete, the algorithm might be correct in Bayesian sense[10].

The framework proposed in this paper models the discrepancies between the human judgment and probabilistic reasoning in the following ways:

1. The subject may partly *misunderstand* the problem or the intentions of the experimenter. Subject's *beliefs may differ* from what can generally be assumed based on common reason. Therefore, the universe of problem spaces, the BBN, that he or she works within, is different from what the experimenter expects.

2. The *strength of connections*, the elementary probabilistic influences and prior likelihoods can be different across subjects.

[10] Actually, *probabilistic logic sampling* proposed by Henrion [9] is based exactly on this scheme and is an example of a belief update algorithm for BBN.

3. Due to cognitive limitations, the subject may *fail to include relevant factors* (that he or she would normally consider relevant) that have an impact on the problem space. This seems to be an important source of discrepancies, because it is effectively equal to omitting entire subtrees of a normative probability tree.
4. Computational imperfections can cause *wrong assessment of strength* (probabilities) of an entire scenario.
5. *Aggregation of strengths* of the scenarios considered can be imperfect and have impact on the result.

It would be interesting to investigate, which of these five have the biggest impact on the result in this framework and which of them could be held responsible for causing the systematic errors that have been postulated in the literature. Computer simulation of human reasoning within this model might yield an answer.

6.8 Further Work on the Model

My preliminary findings have led to a framework for comparing human reasoning under uncertainty to Bayesian inference and to formulation of preliminary hypotheses concerning cognitive processes. However, the studies conducted were limited in scope and number of subjects.

In order to give this work a solid foundation, I would like to investigate whether BBNs are indeed a sufficient formalism for modeling subjects' knowledge. To do this it is necessary to determine what elementary building blocks, rules of uncertain reasoning are used by humans. Two such building blocks underlying the BBNs are conditional dependence between two random variables and inter–causal dependence between alternative causes of an event ("explaining away"). It is essential to find out whether there are other reasoning rules specific to human reasoning, that are not modeled by BBN. If such rules are found, human uncertain knowledge would need to be modeled by a superset of BBN or another formalism. Isolating and testing the rules of uncertain inference would consist of experiments with human subjects. The experiments will test subjects' abilities to apply the rules across various domains and in qualitative and quantitative settings. The method that seems to be the most suitable for this goal is protocol analysis, although more conventional experiments testing validity of well formulated hypotheses are also possible.

In order to maximize the information available from the protocols and insure uniformity of interpretation, the protocol analysis will have to be formalized. Thus, a suitable coding scheme has to be developed to code the protocols at a small–grain or at least an intermediate–grain level.

The way a judgmental problem is approached by humans has been demonstrated to be highly dependent on the form in which the information crucial for the problem is presented and also on context–dependent experience of the subjects [5]. An interesting problem that may be addressed by this research is decomposability of problems in the structuring phase. How do subjects decide how far they go in decomposing a problem? When do they decide that a problem is sufficiently decomposed? Among the subjects solving the President problem, there were those who considered only the target event and those who built elaborated problem spaces.

Simulation of the proposed framework would consist of collecting protocols of subjects performing judgmental tasks in a simple domain, analyzing these protocols and creating a BBN model of that domain, and implementing it by means of a computer program written in a production oriented language (e.g. OPS-5). This program could then simulate various possible ways in which the task could be performed within the framework, starting from correct Bayesian algorithms, through single computational distortions (enumerated in Section 6.7) towards complete processes compatible with those observed in the protocols. The next stage in construction of such a model would involve tuning it up and checking if the results it produces are compatible with the existing literature on heuristics and biases in probability estimations. The final stage would involve automatic generation of outlines of verbal protocols that would be similar to those observed in the experiments.

Observations made in my experiments indicate that out of the two methods of belief updating used by humans, scenario thinking was more common than belief propagation. An interesting research topic is finding out whether scenario thinking can be used for automatic generation of explanations of uncertain inference.

7 Conclusion

The research described in this paper applied the Theory of Problem Solving to human judgment under uncertainty. In order to learn more about the problem representation and the cognitive processes, it used analysis of concurrent verbal protocols. As the first approach, Bayesian belief networks were used as a mechanism for symbolic modeling of the problem representation.

Using these three methodological approaches led to development of a computational framework in which human judgment can be modeled and compared to probability theory. A strength of this framework is its explicitness in representing the elementary events constituting the problem space and its ability to point out explicitly where human processes stop short of following the probability calculus.

Initial findings from an informal analysis of protocols show that structuring the problem plays an important part in the judgment. Non–linear, network–like knowledge that is relevant to the problem, seems to be structured in form of scenarios, chains of events connected by causal links. All scenarios lead to the same point: the target event. The problem space created by the subject often resembles an incomplete probability tree. Difference between this tree and the tree constructed conform the probability theory possibly accounts for the imperfections in human judgment. It is suggested that current observations could be verified by a formal study of protocols and by computer simulation.

References

[1] Robyn M. Dawes. The robust beauty of improper linear models in decision making. *American Psychologist*, 34(7):571–582, July 1979.

[2] Marek J. Druzdzel. "What caused Harry's sneeze?": Cognitive Processes and Problem Solving – term project report. Department of Engineering and Public Policy, Carnegie Mellon University, December 1988.

[3] Hillel J. Einhorn and Robin M. Hogarth. Judging probable cause. *Psychological Bulletin*, 99(1):3–19, January 1986.

[4] K. Anders Ericsson and Herbert A. Simon. *Protocol Analysis: Verbal Reports as Data.* MIT Press, 1984.

[5] Gerd Gigerenzer, Wolfgang Hell, and Hartmut Blank. Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3):513–525, August 1988.

[6] Zvi Ginossar and Yaacov Trope. Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology*, 52(3):464–474, March 1987.

[7] Jr. Harry E. Pople. Heuristic methods for imposing structure on ill structured problems: The structuring of medical diagnosis. In Peter Szolovits, editor, *Artificial Intelligence in Medicine*, chapter 5, pages 119–189. Westview Press, Boulder, CO, 1979.

[8] Max Henrion. Uncertainty in artificial intelligence: Is probability epistemologically and heuristically adequate? In Jeryl Mumpower, Ortwin Renn, Lawrence D. Phillips, and V.R.R. Uppuluri, editors, *Expert Judgment and Expert Systems, Proceedings of the NATO Advanced Research Workshop on Expert Judgment and Expert Systems*, pages 105–129, Porto, Portugal, August 1986. Berlin, Germany: Springer Verlag.

[9] Max Henrion. Propagating uncertainty in Bayesian networks by probabilistic logic sampling. In J.F. Lemmer and L.N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*, pages 149–163. Elsevier Science Publishers B.V. (North Holland), 1988.

[10] Daniel Kahneman, Paul Slovic, and Amos Tversky, editors. *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge University Press, 1982.

[11] Daniel Kahneman and Amos Tversky. On prediction and judgment. *ORI Research Monograph*, 12(4), 1972.

[12] Daniel Kahneman and Amos Tversky. On the psychology of prediction. *Psychological Review*, 80(4):237–251, July 1973.

[13] Paul E. Lehner. When should a decision maker ignore the advice of a decision aid? In *Proceedings of the 5th Workshop on Uncertainty in Artificial Intelligence*, pages 216–223, Windsor, Ontario, August 1989.

[14] M. Granger Morgan, Max Henrion, and Mitchell J. Small. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* Cambridge University Press, 1989.

[15] Allen Newell and Herbert A. Simon. *Human Problem Solving.* Prentice–Hall, Inc., Englewood Cliffs, NJ, 1972.

[16] Judea Pearl. Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, 29(3):241–288, September 1986.

[17] Ros D. Shachter. Evaluating influence diagrams. *Operations Research*, 34(6):871–882, November–December 1986.

[18] Herbert A. Simon. The structure of ill structured problems. *Artificial Intelligence*, 4(3,4):181–201, Winter 1973.

[19] Amos Tversky and Daniel Kahneman. Causal schemata in judgments under uncertainty. In M. Fishbein, editor, *Progress in Social Psychology.* Lawrence Erlbaum Associates, Hillsdale, NJ, 1977.

[20] Thomas S. Wallsten. The theoretical status of judgmental heuristics. In Roland W. Scholz, editor, *Decision Making Under Uncertainty*, pages 21–37. Elsevier Science Publishers B.V. (North Holland), 1983.

[21] Ben P. Wise and Max Henrion. A framework for comparing uncertain inference systems to probability. In L.N. Kanal and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 69–83. Elsevier Science Publishers B.V. (North Holland), 1986.