Problem-Solving Design: Reasoning about Computational Value, Tradeoffs, and Resources

Eric Horvitz

Medical Computer Science Group
Knowledge Systems Laboratory
Stanford University
Stanford, California 94305

Abstract

Several areas of research on problem-solving tradeoffs in reasoning systems are presented. Issues surrounding the valuation of computation in the context of computational resource constraints are introduced. Areas of research on problem-solving tradeoffs receiving ongoing attention include: (1) strategic control, (2) structural control, and (3) the explanation of computation. In each area, we review the application of utility theory to the task of controlling problem-solving tradeoffs.

1. Introduction

In this paper, we describe research on computational tradeoffs arising in the design and implementation of automated problem solvers. We survey several areas of ongoing research on the application of utility theory and techniques of decision analysis to the design and control of computational problem solving. After a discussion of issues surrounding the assignment of utility to attributes of computation, we describe research on problem-solving tradeoffs in three areas of investigation.

First, we explore the application of utility theory and decision analysis to the problem of *strategic control.* We use the term strategic control to refer to metareasoning techniques that address the selection of a strategy from a well-defined set of discrete reasoning methods. Because numerous strategic control issues are highlighted under conditions of scarce computational resources, we dwell on real-time problem solving. We examine important considerations in the selection of a computational strategy or sequence of strategies to solve particular problem

This work was supported by NASA-Ames under Grant NCC-220-51 and the National Library of Medicine under Grant RO1LM04529. Computing facilities were provided by the SUMEX-AIM Resource under NIH Grant RR-00785.

challenges. We also specify properties desired of strategies for reasoning in situations where there is wide *variation* and *uncertainty* in the amount of computation time available. Finally, the abstract control issues are illustrated with examples of research on value tradeoffs in the realm of strategic reasoning.

Next, we move into a related area of problem-solving tradeoff research focused at a more fundamental level of problem solving: We study the application of utility theory to the detailed configuration of classes of problem-solving strategies. In contrast to choosing from a set of alternative families of reasoning strategies, this area of investigation addresses techniques for reasoning about tradeoffs at the microstructure of problem-solving methods. We refer to metareasoning about the fine details of algorithms as *structural control*.

Finally, we review a class of computational tradeoffs that arise because of human cognitive constraints. Specifically, we examine the tradeoffs that arise at the human-machine interface. Because the value derived from computational problem solving frequently depends on the understandability of computational behavior, automated reasoners must consider the constraints on human comprehension of complex computational inference and results.

2. The Value of Computation

The quantification and formal manipulation of notions of value and preference have been investigated in the field of decision analysis [Howard 84, Howard 70, Raiffa 68]. An axiomatic framework termed *utility theory* [von Neumann 53] lies at the heart of decision theory. Measures of value consistent with the axioms of utility theory are called *utilities*. Von Neumann and Morgenstern, the authors of utility theory, proved that individuals making decisions consistent with a small set of axioms behave as if they associate a measure of utility with alternative outcomes and act to maximize their expected utility [von Neumann 53]. Over the years, decision analysts have developed useful tools for assessing and applying knowledge about the utility structure of complex problems. We have been attempting to turn the power and elegance of these techniques on the process of reasoning and problem solving itself.

2.1. Assigning utility to multiple attributes of problem solving

The value assigned to computational behavior can be directly assessed, or may be described by a qualitative or more detailed function that represents the relationships among costs and benefit associated with alternative outcomes. Such value functions assign a single-value measure to computation based on the status of an *n*-tuple of attributes. For example, the value associated with the use of a medical expert system in a particular context might be a function of a number of attributes, including *speed* of computation, *accuracy* of recommendation, and *clarity* of explanation. We have been working with expert physicians in the intensive-care and tissue-pathology domains to ascertain value models relating measures of utility to multiple attributes of computation.

We are not the first to explore the formal use of utility theory in the control of reasoning. Concurrent research has focused on the usefulness of assigning utilities to alternative strategies in the control of logical reasoning [Smith 86, Treitel 86]. The research presented here differs from the other work in its focus on representing multiple components of value and on the integration of context-specific knowledge concerning human preferences about computational tradeoffs.

To our knowledge, multiattribute utility models were first used in the control of computational inference and explanation in the PATHFINDER expert system [Horvitz 86a, Horvitz 86b]. A component of PATHFINDER research has been focused on the investigation of techniques for the dynamic application of multilinear utility models to control computation and explanation. This aspect of PATHFINDER research evolved into the current PROTOS Project, focused on the development of techniques for reasoning with knowledge about the value of alternative strategies under real-world resource constraints.[2]

2.2. Problem-solving tradeoffs

Computation in a world of bounded resources often is associated with cost/benefit *tradeoffs*. Working with expert physicians on the development of expert systems has highlighted the importance of developing computational techniques that can explicitly control tradeoffs. With a computational tradeoff, the benefit associated with an increase in the quantity of one or more desired attributes of computational value is intrinsically linked to costs incurred through changes imposed on other attributes. More specifically, we define a tradeoff as a relationship among two attributes, such as the *immediacy* and *accuracy* of a computational result, each having a positive influence on the perceived total value of computer performance, such that they are each constrained to be a monotonically decreasing function of the other over some relevant range. In the case of our sample tradeoff,

$$\text{ACCURACY} = \text{F(IMMEDIACY)},\ t_o \leq \text{IMMEDIACY} \leq t_n \qquad (1)$$

where F is some monotonically decreasing function over the range bounded by computational time delays t_o and t_n. This definition can be generalized to the case where the value assigned to tuples of a subset of relevant attributes is a monotonically decreasing function of tuples composed of other attributes. The

[2] PROTOS is an imperfect acronym for Project on computational Resources and TradeOffs.

tradeoff between the immediacy and the accuracy or precision of a solution is particularly explicit in methods that incrementally refine a computational result with time.

Other value tradeoffs encountered in reasoning systems that we have examined include:

- degree of certainty *versus* level of abstraction
- solving a subproblem *versus* solving other subproblems
- metareasoning *versus* object-level reasoning
- inference transparency *versus* inference optimality

Most reasoning systems have been designed with implicit assumptions about the handling of problem-solving tradeoffs. The intent of this research is to develop methods that enable computer scientists to reason explicitly about tradeoffs in the engineering of systems. We are also working to develop tools that will allow reasoning systems to autonomously apply knowledge about the domain and about alternative reasoning methods to tailor inference to a range of problem challenges and contexts. Our research has highlighted the knowledge-intensive nature of reasoning about value tradeoffs in different contexts.

Our research centers on the application of multiple-attribute utility to for the control and analysis of computational tradeoffs. These techniques, in the contexts of engineering and real-time operation of reasoners, provide a means for controlling tradeoffs through the application of knowledge about problem-solving resources and the preferences of system users.

2.3. Components of computational value

We have found it useful to decompose the value associated with computational inference into several components. We assert that the application of an inference strategy is associated with some net benefit or cost to an agent such as a system user, a robot, or a computational subsystem, relying on computation for decision making. We use the term *comprehensive value* (V_c) to refer to the net *expected utility* associated with the application of a computational strategy. We will see that this value is a function of the strategy, of the problem, and of the problem-solving context. We have found it useful, in studying problem-solving tradeoffs under pressing resource limitations, to view the comprehensive value as having two components: the *object-level* value and *inference-related* value.

The *object-level* value (V_o) is the expected utility associated with computation-based increases in information about the *objects* of problem solving. For example, the object-level value associated with the use of an expert system for assistance with a complex medical diagnosis problem refers to the costs and benefits associated solely with the change in information about the entities in the medical problem such as alternative treatments, likelihoods of possible outcomes, and costs of recommended tests.

The *inference-related* value (V_i) is the expected disutility intrinsically associated with *computation*, such as the cost a physician might attribute to the delay of a decision because of the time required by an expert system to generate a recommendation, or the cost associated with his inability to understand the rationale behind a decision recommendation.

The explicit decomposition of problem-solving utility into object- and inference-related value is useful in that it focuses attention on the costs as well as benefits associated with problem-solving activity. In general, we may have to consider important the dependencies between the object- and inference-related value. We assume the existence of a function F that relates V_c to V_o, V_i and additional background information about the problem-specific dependencies that may exist between the two components of value. That is,

$$V_c = F(V_o, V_i, \phi)$$

where ϕ captures problem-specific background information about possible dependencies between object- and inference-related value.

2.4. Optimal, proximal, and heuristic reasoning

We refer to a reasoning strategy that has the highest expected value, in the context of the beliefs of an engineer or automated reasoner as a *rational* computational strategy. In reasoning about problem-solving tradeoffs, it is useful to enumerate a theoretically optimal *frame of reference* for problem solving. This task can be viewed as identifying a theory of inference or computational result that would be desired in a world of unlimited computational resources. Defining such a basis provides a framework for reasoning about inference with the greatest expected value. We refer to non-optimal strategies as either *proximal* or *heuristic.* Proximal strategies yield a computational result in conjunction with a well-defined measure of divergence from an optimal result. In contrast, a computer scientist (or autonomous agent) may often be uncertain about the behavior of a computational strategy. We refer to strategies with uncertain performance strategies as *heuristic*. Thus, we use the term heuristic to reflect a state of incomplete knowledge about computational behavior. It is clear that even when the performance of a strategy is well understood, there may be uncertainty about the utility associated with a strategy. Thus, proximal strategies often have a heuristic value structure.

The definition of heuristic, as dependent on the state of knowledge about the behavior of a computational strategy, implies that new knowledge can lead to the reclassification of a heuristic strategy as a proximal strategy. For example, a heuristic strategy may become a proximal one for solving well analyzed classes of problems. There have been examples of the successful analytic characterization of heuristics [Pearl 84]. For example, work on developing ϵ-approximate algorithms has demonstrated worst-case error bounds for particular heuristic methods [Papadimitriou 82].

A strength of applying the theory of utility to computation is its ability to explicitly handle the consistent assignment of value for *uncertain* as well as *certain* outcomes. Thus, utility values can be assigned to uncertain heuristic strategies as well as to formal inference techniques. The assignment of utility to inference strategies, whether through analysis or direct subjective assignment, establishes a *conceptual continuum* between formal and the more poorly characterized heuristic approaches to reasoning.

3. Strategic Control

One of our chief research focuses has been the study of problem solving under varying limitations in the amount of resource available for reasoning. We are particularly interested in determining useful "optimal," proximal, and heuristic strategies in the context of uncertain and varying constraints on the amount of resources available for representation, inference, and metareasoning about inference. I have been studying techniques for endowing systems with knowledge about the costs, benefits, and uncertainties associated with various computational approaches in alternative contexts. This requires consideration of the expected value of computation and the cost of resources required for reasoning. In the past, numerous value issues central in the design and implementation of problem solvers have remained ill-defined and implicit. Our research centers on identifying and making issues surrounding *informational needs* and *computation availability* explicit in the design, representation and inference in automated reasoners.

3.1. Problems, methods, and reasoning resources

The amount of time available for computation in systems that must perform in complex real-time environments varies greatly depending on the urgency of the situation. Likewise, the computation required for problem solving may vary depending on the task at hand. We are interested in inference strategies that can respond flexibly to wide variations in the availability of resources, and metalevel reasoning techniques that can balance the costs of approximate or incomplete analyses with the benefits of more tractable computational methods.

The importance of tailoring representations and accompanying inference methods to problem demands and resource availability is illustrated by the use of different methods in the physical sciences for reasoning about the same fundamental physical phenomena. The application of quantum mechanics is the most precise method for reasoning about chemistry; however, real-world chemistry problems often do not require the precise results that are theoretically computable from quantum theory. More parsimonious, classical representations of chemical interaction frequently produce useful answers at a fraction of the quantum reasoning cost.

Within automated problem solving, it can be similarly important to fit a problem-solving method to the needs and resources available. Let us now explore approaches to handling computational resource limitations through endowing reasoning systems with formal tools for making decisions about the *value* of computational problem-solving in alternative contexts.

Resource-constraint issues can be especially salient in the context of real-time requirements. In the real world, delaying an action is often costly. Thus, computation about belief and action often incurs inference-related costs. The time required by a reasoning system for inference varies depending on the complexity of the problem at hand. Likewise, the costs associated with delayed action vary depending on the stakes and urgency of the decision context. The real-time problem is additionally complicated by the existence of uncertainty in the functional form of the cost functions associated with delayed action. We are studying reasoning strategies that can respond flexibly to wide variations in the availability of resources. The intent of our research is to develop coherent approaches to generating and selecting the most promising strategy for particular problem-solving challenges.

3.2. Toward a continuum of value: Partial results

Let us now focus on the properties of proximal and heuristic problem-solving methods that would be useful under varying resource constraints. Classical approaches to computational problem solving have focused on the determination of final answers. Complexity theorists have focused almost exclusively on proving results about the time and space resources that must be expended to run algorithms to termination [Garey 79, Aho 83, Papadimitriou 82]. In the real world, strict limitations and variations on the time available for problem solving suggest that the focus on time complexity for algorithmic termination is limited; analyses centering on how *good* a solution can be found in the time available for computation are of importance.

The major rationale for the focus on the time complexity of algorithmic termination seems to reside in the simplifying notion in algorithms research that a computer-genera:ed result can be assigned only one of two measures of utility: either a solution is found and is of value, or a solution is not found and is therefore valueless. However, it is often possible to enumerate representations and inference techniques that can provide partial solutions of varying *degrees* of value. Algorithms that generate partial results incrementally refined with additional computation can be extremely useful because of the great variation and uncertainty in the functions describing inference-related costs.

3.3. Example: Sorting and searching under resource constraints

We have been working to adapt classic searching and sorting algorithms to a partial-result paradigm through identifying intermediate states and applying value functions that capture preferences about partial results. Our research in this area is focused on investigating fundamental issues and useful control architectures for real-time reasoning. In addition to enumerating tradeoffs, we are studying the relevance of classes of *immediacy versus goodness-of-answer* tradeoffs being uncovered in recent computational theory research. In addition to our theoretical work, we have been exploring the empirical behavior of alternative sorting and searching strategies under plausible valuation models within the PROTOS system testbed. One insight gained from this work is that an interleaving of strategies--where one sorting or searching routine is handed the partial result of previously applied procedures--may have a higher expected value than any one strategy.

3.4. Example: Diagnosis under varying resource constraints

The example of sorting or searching under resource constraints yields useful insights about bounded-resource problem solving within more sophisticated applications. Much of our research effort has dwelled on diagnostic reasoning under varying resource bounds. Diagnosis refers to the problem of finding the most likely explanation for a set of findings. More comprehensively, diagnostic problem solving describes the process of making a sequence of decisions about the best tests to perform and revising the belief in hypotheses as test results become available. As reasoning under uncertainty is unavoidable for most realworld diagnosis problems, we seek a base theory for diagnosis under uncertainty. Investigators have been exploring a number of formal and informal approaches to reasoning and decision making under uncertainty [Zadeh 83, Shortliffe 75, Shafer 76, Cheeseman 85, Horvitz 86c].

A *frame of reference* for optimal diagnostic reasoning is the classical normative basis for inference. The basis, enumerated in middle-twentieth century [von Neumann 53, Howard 84, Pratt 65, Raiffa 68], is commonly referred to as *decision theory* and is based on two sets of axioms. The first set of axioms defines the theory of probability. The second set defines the theory of utility introduced above. The use of probability theory for assigning belief to alternative hypotheses and the use of utility theory for making decisions about the best tests to undertake has been viewed as a *normative basis* for diagnosis. That is, the properties of probability theory and utility theory have been accepted in several disciplines as defining rational inference [Horvitz 86c, Heckerman86]. Our research on diagnosis, has thus centered on applying a rational framework for metareasoning about the task of controlling a rational framework for diagnosis.

3.4.1. Diagnosis in the real world

The theoretically optimal basis for diagnostic reasoning has performed well on small problems but there have been difficulties associated with its use in problems of realistic complexity. More so than for any other reason, researchers in artificial intelligence have looked beyond the normative basis for diagnosis because of its associated computational complexity and demand for the representation of large amounts of knowledge [Gorry 73, Davis 82, Szolovits 82].

While the normative basis may provide a gold standard for diagnostic inference with sufficient resources, it is clear that engineering and computation costs have impact on its optimality. For example, the inference-related cost associated with the delay required for computation, often renders the comprehensive value of decision-theoretic inference worthless. Clearly, a physician seeking expert advice on a life threatening problem may not find a theoretically optimal recommendation to be "optimal" for his needs when he must wait several hours for the result.

Most importantly, straightforward implementations of the base strategy have been fragile to swings in the amount of computation time; they have been of little value if the amount of available computational resource dips below the amount required for solving an entire problem. Classical implementations of the base strategy can only respond with the maximum object-level value, given sufficient resources. If the allocated computation time falls short by even a few milliseconds, the value of computation is zero for the decision at hand.

3.4.2. Bounded-resource inference for diagnosis

Classic normative diagnosis has centered on the precise calculation of probabilities. That is, the traditional approach to queries of the form "P(X | Y) = ?" has been to calculate point-probability values. The intractability of general probabilistic inference can make the calculation of point probabilities impractical within the time available for decision making. This intractability has been seen as supporting the view that probability is an impractical theory for reasoning under uncertainty in the real world. This perspective is based on the assumption that computational strategies for probabilistic inference are of no value until a final point-probability assignment is calculated. The view is countered by the enumeration of inference strategies that can incrementally refine probabilistic assignments. Such strategies can generate partial solutions with increasing degrees of value with computation.

Examples of strategies that generate partial results in diagnostic reasoning are stochastic simulation [Pearl 86], probabilistic bounding [Cooper 84], completeness modulation [Horvitz 87a], and abstraction modulation [Horvitz 87a]. These approaches can frequently generate partial solutions through trading off a gain in efficiency for a decrease in the resulting precision or accuracy of inference. More details of research on these strategies and on applying decision theory to controlling decision-theoretic inference under ranging resource constraints is discussed in [Horvitz 87a].

3.4.3. Application: Decision-support in the intensive-care unit

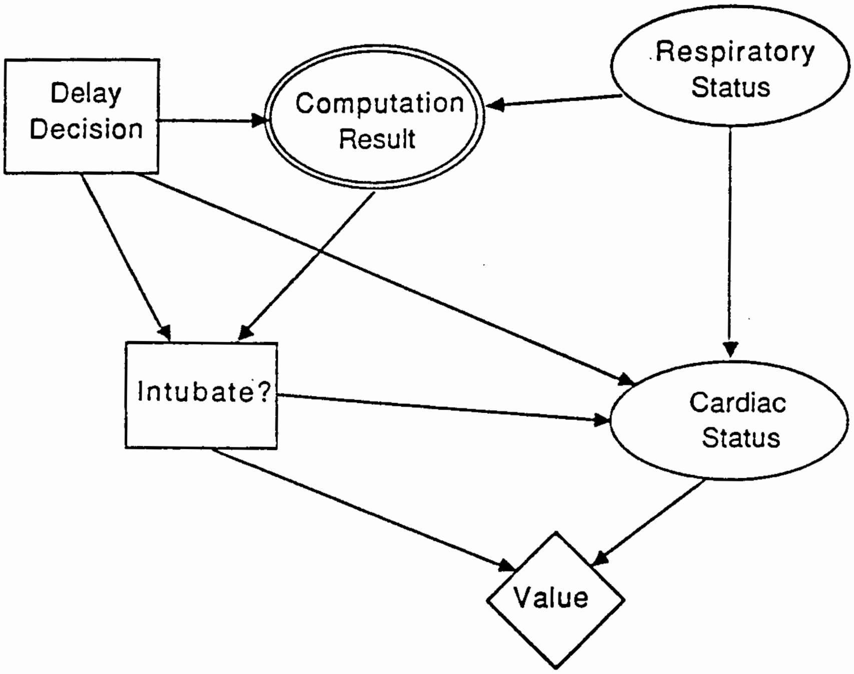

Figure 1: Diagnosis under resource constraints in the intensive-care unit.

We have been exploring diagnostic decision-support under resource bounds within the intensive-care-unit application area. Figure 1 shows an influence diagram representation of the problem facing a physician who must decide whether or not to put a patient showing signs of respiratory distress on an artificial respirator (intubation). Arcs and nodes in an influence diagram have a well-defined decision-theoretic semantics [Heckerman 87]. For the sake of this brief overview, it is important only to note that the arcs indicate dependency between propositions represented by nodes. The figure relays that the value of an outcome depends, in part, on a decision about whether or not action should be delayed to gain computed diagnostic support about a patient's status. While waiting may give a physician useful information about whether a patient should be intubated, the delay can be costly: If the patient is in respiratory distress, a delay raises the probability of cardiac failure. Thus, the benefits of waiting for computed advice must be balanced with the costs of delay. Depending on the specific value structure of the situation, different bounded-resource diagnostic methods and delays will optimize the expected value of the situation for the patient. This example is discussed in greater detail in [Horvitz 87a].

4. Structural Control

In addition to studying methods for choosing the best strategy from a number of alternative strategies, we are using decision analysis for determining optimal configurations of the structure of problem-solving methods. The goal of structural control is to customize algorithmic problem-solving approaches to specific value functions and resource availability.

4.1. Reasoning about the structure of algorithms

As an example, consider the issues surrounding the discretization of variables manipulated by a reasoning system. To optimize the usefulness of a strategy under resource constraints, it is often necessary to reason about the granularity with which knowledge is represented and processed. The size of a computational problem can vary depending on the coarseness with which important variables are represented. The coarser the variables are, the fewer discrete objects must be handled by a system under resource pressures. It is often important to study the value tradeoffs that arise from balancing the benefits of tractability associated with increases in the granularity of inference (or representation) with the losses based in a decreased precision or accuracy of a problem solution.

With an explicit value model, a system designer may vary the granularity of the intervals considered over such metrics as distance or time in an algorithm to optimize the value of a reasoning strategy. Different variables may be assigned different optimal levels of granularity. In cases of computational complexity and severe resource constraints, such problem-solving optimization can mean the difference between valuable and worthless inference.

4.2. Example: Configuration of divide-and-conquer algorithms

More sophisticated manipulation of the structure of problem solving may be carried out at the fine details of algorithms. As an example, we consider the general form of divide-and-conquer algorithms. A divide-and-conquer algorithm can be divided into a set of inference-related costs associated with

1. Decomposition of comprehensive problem statement into subproblems

2. Solution of subproblems

3. Recomposition of solved subproblems into the complete solution

Divide-and-conquer algorithms can be converted from traditional invariant approaches to methods with flexible problem-solving dimensions. We have been experimenting with techniques for reasoning about the configuration of a class of divide-and-conquer algorithms to optimize the expected value of reasoning given a problem and resource context [Horvitz 86d].

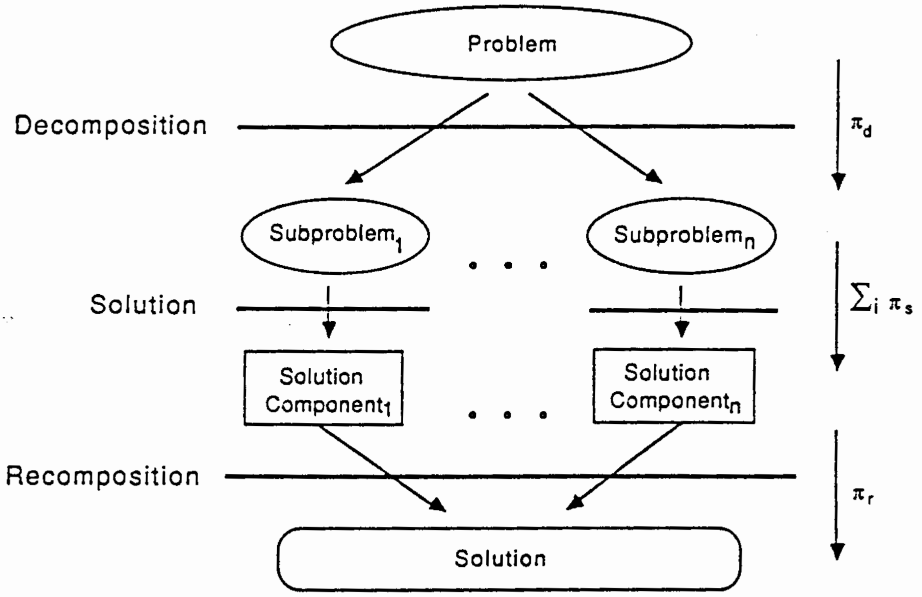

Figure 2: Reasoning about the structure of divide-and-conquer algorithms through considering costs of decomposition, subproblem solution, and recomposition.

Figure 2 depicts fundamental components of divide-and-conquer problem solving. With such algorithms, a problem is decomposed into a set of subproblems. After the subproblems are solved, component solutions are recomposed into a final answer. The decomposition typically entails some computational cost (Π_d) that is dependent on the number of subproblems created. The cost of solving all of the subproblems is the sum over the separate subproblem solution costs (Π_s). The cost of solving a particular subproblem is some function of the size of subproblems. Subproblem size is dependent on the decomposition. Finally, the cost of assembling the solution (Π_r) is a function of the number of subproblems. These costs comprise the inference-related cost associated with the application of a divide-and-conquer algorithm.

The inference-related cost can be minimized by varying independent parameters of a divide-and-conquer approach. We have begun to analyze prototypical descriptions of divide-and-conquer problems in which the number and average size of subproblems can be varied. In many cases, functions can be enumerated that relate the number of subproblems, the size of subproblems, and the expected costs associated with problem decomposition, solution, and recomposition. In such cases, we have used the expected inference-related costs to determine the optimal decomposition configuration given a specific problem. Similar techniques can be applied to more complicated problem-solving methods. For example, a set of problem solving parameters can be considered simultaneously.

5. Explanation

A third area of our research on computational tradeoffs focuses on problems with the *comprehension* of inference strategies and results. The transparency of inference has been considered a definitive component of expert systems, distinguishing them from numerical programs and other kinds of reasoning systems in artificial intelligence [Buchanan 82]. The comprehension of strategies and results of automated reasoning has also been identified as an important factor in the acceptance of expert systems [Teach 81]. Most explanation research has centered on the refinement of knowledge used by expert systems [Swartout 81, Wallis 82, Patil 81]. In contrast, we have been studying the enumeration and control of explanation tradeoffs under constraints in cognition and time.

Research in cognitive psychology on the limited ability of humans to comprehend complex information in the short term [Bruner 56, Miller 56, Waugh 65] underscores the need for managing the complexity of explanation in expert systems. Our experiences and several psychological studies[3] have confirmed the existence of significant tradeoffs between the benefits of completeness in transmitting all possible relevant information and the costs incurred in the time, effort, and losses in comprehension. We have investigated explanation tradeoffs in the context of explaining formal diagnostic inference and complex biological simulations. The tradeoffs and reasoning issues that arise with the imposition of cognitive constraints on an unconstrained problem solver, are similar to those based in resource constraints.

The explicit control of explanation tradeoffs in diagnosis has been performed within the PATHFINDER Project [Horvitz 86a, Horvitz 86b]. PATHFINDER explanation

[3] For example, classic studies have demonstrated that the decision-making performance of humans begins to degrade as the quantity of relevant pieces of information presented increases beyond a relatively small number of items.

research has focused on the enumeration and control of a tradeoff between the transparency of diagnostic inference and the optimality of inference. We discovered that, although recommendations were "optimal" within the limited scope of the formal theory used for recommending tests, users frequently found recommendations to be unnatural and opaque. Applying human-oriented problem-solving hierarchies to simplify the system's reasoning produced recommendations that were more transparent, but informationally suboptimal. Thus, cognitive constraints can be viewed as imposing inference-related costs in the context of complex and unnatural artificial reasoning strategies. Further work has focused on the automated selection of a strategy from a set of alternative simplification strategies with value functions that represent knowledge about a user's preferences about trading-off optimality for clarity. We refer readers interested in explanation under bounded cognitive resources to [Horvitz 87b].

6. Toward a Science of Limited Rationality

Our research on application of decision-analytic techniques to metareasoning about problem solving under realistic constraints is motivated by an attempt to integrate and apply problem-solving methods developed over the last three decades within the disciplines of operations research, decision science, and artificial intelligence. The explicit consideration of knowledge about value considerations and resource availability is essential in relating alternative reasoning methods to one another and to the endeavor of constructing problem solvers that perform under real-world constraints.

We have been pursuing the development of a science of problem solving under resource constraints. It is not clear yet whether there is enough regularity in relations among components of problem-solving methods to warrant belief that such a science is possible. Whether reasoning under well-specified constraints evolves into a science or remains a focus of engineering design, it is clear that the identification and control of common, inescapable, problem-solving tradeoffs will play an important role in the pursuit of automated intelligence.

Acknowledgements

I am grateful to Bruce Buchanan, Greg Cooper, George Dantzig, Edward Feigenbaum, Ron Howard, Nils Nilsson, Christos Papadimitriou, and Edward Shortliffe for providing insightful feedback on this research. Jacques Suermondt and Lynne Hollander provided useful comments on an earlier version of this manuscript.

References

[Aho 83] Aho, A.V., Hopcroft, J.E., and Ullman, J.D.
Data Structures and Algorithms.
Addison-Wesley, Menlo Park, California, 1983.

[Bruner 56] J.S. Bruner, J.J. Goodnow, G.A. Austin.
A study of thinking.
Wiley, 1956.

[Buchanan 82] Buchanan, B. G.
Research on Expert Systems.
In J. Hayes, D. Michie, Y. H. Pao (editors), *Machine Intelligence*, pages 269-299. Ellis Howard Ltd., Chichester, England, 1982.

[Cheeseman 85] Cheeseman, P.
In defense of probability.
In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence.* IJCAI-85, 1985.

[Cooper 84] Cooper, G.F.
NESTOR: A Computer-Based Medical Diagnostic Aid that Integrates Causal and Probabilistic Knowledge.
PhD thesis, Medical Information Sciences, Stanford University, Stanford, California, 1984.
CS report no. STAN-CS-84-1031.

[Davis 82] Davis, R.
Consultation, knowledge acquisition, and instruction.
in P. Szolovits (editor), *Artificial Intelligence In Medicine*, pages 57-78. Westview Press, 1982.

[Garey 79] Garey, M.R., and Johnson, D.S.
Computers and Intractability: A Guide to the Theory of NP-Completeness.
W.H. Freeman and Company, New York, 1979.

[Gorry 73] Gorry, G. A., Kassirer, J. P., Essig, A., and Schwartz, W. B.
Decision analysis as the basis for computer-aided management of acute renal failure.
American Journal of Medicine 55:473-484, 1973.

[Heckerman 86] Heckerman, D.E.
Probabilistic interpretations for MYCIN's certainty factors.
In Kanal, L.N., and Lemmer, J.F. (editors), *Uncertainty in Artificial Intelligence*, pages 167-196. North Holland, New York, 1986.

[Heckerman 87] Heckerman, D.E., and Horvitz, E.J.
On the expressiveness of rule-based systems for reasoning under uncertainty.
In *Proceedings of AAAI.* American Association for Artificial Intelligence, Seattle, Washington, July, 1987.

[Horvitz 86a] Horvitz, E.J., Heckerman, D.E., Nathwani, B.N., and Fagan, L.M.
The use of a heuristic problem-solving hierarchy to facilitate the explanation of hypothesis-directed reasoning.
In *Proceedings of Medinfo.* Medinfo, October, 1986.

[Horvitz 86b] Horvitz, E. J.
Toward a science of expert systems.
In *Proceedings of the 18th Symposium on the Interface of Computer Science and Statistics.* Ft. Collins, Colorado, March, 1986.

[Horvitz 86c] Horvitz, E.J., Heckerman, D.E., and Langlotz, C.P.
A framework for comparing alternative formalisms for plausible reasoning.
In *Proceedings of AAAI.* American Association for Artificial Intelligence, Philadelphia, Pennsylvania, August, 1986.

[Horvitz 86d] Horvitz, E.J.
Reasoning about inference tradeoffs in a world of bounded resources.
Technical Report, Stanford University, 1986.
Knowledge Systems Lab Technical Report KSL-55-86, Stanford University, September, 1986.

[Horvitz 87a] Horvitz, E.J.
Reasoning about beliefs and actions under computational resource constraints.
In *Proceedings of AAAI Workshop on Uncertainty in Artificial Intelligence.* American Assoiciation for Artificial Intelligence, Seattle, Washington, July, 1987.
Also available as Knowledge Systems Laboratory Technical Report KSL-29-87, March, 1987.

[Horvitz 87b] Horvitz, E.
A Multiattribute Approach to Inference Understandability and Explanation.
Technical Report KSL-28-87, Stanford University, Knowledge Systems Laboratory, Stanford, California, March, 1987.

[Howard 70] Howard, R.A.
Decision analysis: perspectives on inference, decision, and experimentation.
Proceedings of the IEEE 58(5):632-643, May, 1970.

[Howard 84] Howard, R. A., and Matheson, J. E.
Readings on the Principles and Applications of Decision Analysis.
Strategic Decisions Group, Menlo Park, California, 1984.
2nd Edition.

[Miller 56] Miller, G.A.
The magical number seven, plus or minus two.
Psychological Review 63:81-97, 1956.

[Papadimitriou 82] Papadimitriou, C.H., and Steiglitz, K.
Combinatorial Optimization: Algorithms and Complexity.
Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1982.

[Patil 81] R.S. Patil.
Causal Representation of Patient Illness for Electrolyte and Acid-Base Diagnosis.
PhD thesis, M.I.T., 1981.

[Pearl 84] Pearl, J.
Heuristics.
Addison-Wesley, Reading, Massachusetts, 1984.

[Pearl 86] Pearl, J.
Evidential Reasoning Using Stochastic Simulation of Causal models.
Technical Report R-68, CSD-8600##, Cognitive Systems Laboratory, UCLA Computer Science Department, September, 1986.

[Pratt 65] Pratt, J. W., Raiffa, H., and Schlaifer, R.
Introduction to Statistical Decision Theory (Preliminary Edition).
McGraw-Hill, New York, 1965.

[Raiffa 68] Raiffa, H.
Decision Analysis: Introductory Lectures on Choice Under Uncertainty.
Addison-Wesley, Reading, Mass., 1968.

[Shafer 76] Shafer, G.
A Mathematical Theory of Evidence.
Princeton University Press, Princeton, NJ, 1976.

[Shortliffe 75] Shortliffe, E. H. and Buchanan, B. G.
A model of inexact reasoning in medicine.
Mathematical Biosciences 23:351-379, 1975.

[Smith 86] Smith, D.E.
Controlling inference.
Technical Report STAN-CS-86-1107, Stanford University, April, 1986.

[Swartout 81] Swartout, W.R.
Producing Explanations and Justifications of Expert Consulting Programs.
PhD thesis, Department of Computer Science, M.I.T., 1981.
Report no. LCS-TR-251.

[Szolovits 82] Szolovits, P.
Artificial intelligence and medicine.
In Szolovits, P. (editors), *Artificial Intelligence in Medicine*, pages 1-19. Westview Press, Boulder, Colorado, 1982.

[Teach 81] Teach, R. L., and Shortliffe, E. H.
An analysis of physician attitudes regarding computer-based clinical consultation systems.
Computers and Biomedical Research 14:542-558, 1981.

[Treitel 86] Treitel, R. and Genesereth, M. R.
Choosing Directions for Rules.
In *Proceedings of the AAAI.* AAAI, Morgan Kaufman, Palo Alto, California, August, 1986.

[von Neumann 53] von Neumann, J., and Morgenstern, O.
Theory of Games and Economic Behavior.
Wiley, New York, 1953.
(3rd edition).

[Wallis 82] Wallis, J.W., and Shortliffe, E.H.
Explanatory power for medical expert systems: Studies in the representation of causal relationships for clinical consultation.
Methods of Information in Medicine 21:127-136, 1982.

[Waugh 65] N.C. Waugh, D.A. Norman.
Primary Memory.
Psychological Review 72:89-104, 1965.

[Zadeh 83] Zadeh, L.A.
The role of fuzzy logic in the management of uncertainty in expert systems.
Fuzzy Sets and Systems (11):199-227, 1983.