Cognitively Plausible Heuristics to Tackle the Computational Complexity of Abductive Reasoning

Olivier Fischer

The work described in my Ph.D. dissertation (Fischer 1991)[1] merges computational and cognitive investigations of abductive reasoning. It is the outcome of seven years of research focusing on abductive explanation generation and involving the departments of computer and information science, industrial and systems engineering, pathology, and allied medical professions at The Ohio State University. In the first phase of my work, I characterized abductive problem solving and performed a comparative analysis of two abductive problem solvers (Smith and Fischer 1990). I then summa-

> *The work… merges computational and cognitive investigations of abductive reasoning.*

rized three lines of work on abduction: a knowledge-based approach (Smith et al. 1985), the formal computational complexity analysis of abductive reasoning (Bylander et al. 1989), and the gathering and study of protocols of human experts performing an abductive task (Smith et al. 1989). The unification of these three approaches shows the relevance of computational complexity to understanding human behavior and brought me to enhance the existing knowledge-based system in the Department of Computer and Information Science by integrating cogni-

tive and computational efficiency issues into its design. Thus, I implemented two cognitively plausible heuristics to tackle the complexity of abductive reasoning and successfully experimented with them. This work, originally applied to the domain of alloantibody identification, was generalized to domain-independent abductive problem solving.

Abduction

Abduction, that is, inference to a hypothesis that best explains a set of data, appears to be ubiquitous in cognition. For example, diagnosis (Pople 1977), plan recognition (Charniak 1988), and hypothesis revision (Pearl 1987) can all be viewed as instances of the more general abduction problem.

Given a set of data D to be explained, if hypotheses that can individually explain all of D are directly available, then the hypothesis that explains D most plausibly is the best explanation for it. In general, however, a hypothesis explaining all of D is not directly available. Instead, a best composite hypothesis has to be constructed from a set of elementary hypotheses, where each selected elementary hypothesis can plausibly account for some of D.

Although there is no general agreement on the characterization of the best composite explanation, it appears reasonable (Josephson et al. 1987) that one composite hypothesis $H1$ is a better explanation than another hypothesis $H2$ if $H1$ explains more of D than $H2$, $H1$ is more plausible than $H2$, and $H1$ is less redundant an explanation of D than $H2$.

A Computational Complexity Analysis of Abduction

Chapter 5 of my dissertation describes an analysis of the computational complexity of abductive problems (Bylander et al. 1989). This complexity analysis defines abduction as inference to the best explanation given a set of data to explain. The complexity of reaching such an explanation varies according to the properties of the elementary hypotheses available. The study (Bylander et al. 1989) centers on the influence that some of these properties have on the computational complexity of assembling a composite hypothesis. The properties analyzed are hypothesis independence, monotonicity of the hypotheses' explanatory power, incompatibility between hypotheses, cancellation effect between causal processes associated with different hypotheses, and existence of an ordering of the plausibilities of the hypotheses to be true.

The Protocols

Although the computational complexity of abductive reasoning varies depending on the previously defined properties, abduction is, in general, NP-complete relative to the complexity of computing the data explained by hypotheses. However, protocols show that humans successfully perform this task. Thus, I was led to hypothesize the existence of heuristics that make abduction tractable and that would be related to the properties cited earlier. I studied two of these heuristics and how they are used in the abductive task of alloantibody identification.

A patient in need of a transfusion not only needs blood but requires blood compatible with his/her immune system. Blood bankers analyze the patient's blood to determine which blood in the bank can safely be transfused. Part of this analysis centers on the determination of what antibodies are present or can be produced in the patient's blood and what antigens are present in the donor's cells. The presence of one (or more) antibody in the patient's blood is inferred from reactions between samples of the recipient's blood and a set of carefully selected cells usually bought from specialized laboratories. The task of alloantibody identification was modeled as an instance of

abductive reasoning in the RED system (Smith et al. 1985).

To increase the cognitive content of the mostly performance-oriented RED system, protocol analyses of blood bankers were performed (Smith et al. 1989). This study, described in chapter 4, yielded the following results:

First, the experts use patterns in the data to generate simplifying assumptions, generally decomposing a multiple-antibody problem into single-antibody subproblems.

Second, the experts use symbolic reasoning processes to focus on a candidate solution based on the previous assumption.

Third, the experts evaluate the plausibility of their solution (and, indirectly, the validity of their assumptions) by assessing the plausibility that all the observed data could have been generated by the candidate solution.

The processes used in focusing on a solution can be divided into two categories: processes for generating and testing hypotheses regarding a specific antibody or group of antibodies and processes that generate an abstract model of the solution free of any reference to particular antibodies.

Analyzing the Protocols Using Complexity Theory

In chapter 5, I merge the two lines of work described here. The behavioral description of the blood banker leads to the characterization of a computational model that accounts for the observed data and that fits the computational complexity analysis of abduction. First, the problem solving in the task of alloantibody identification is viewed as an instance of abductive explanation generation (Smith et al. 1985): Given a set of reactions (data), the task is to produce a list of hypothesized antibodies (hypotheses) whose presence could be used to explain the observed reactions. The simplifying assumptions made by the blood banker to solve alloantibody identification problems can be mapped to computationally relevant properties of the problem. The blood banker's classification of the problem into a problem with a single- versus multiple-antibody solution corresponds to the distinction between single- versus multiple-fault diagnosis. Also, blood bankers usually attempt to structure the abductive

…alloantibody identification was modeled as an instance of abductive reasoning in the RED system.

problem by dividing the data into a number of noninteracting sets of reactions, each set corresponding to one of the antibodies present (Smith et al. 1989). This division corresponds to the distinction between the multiple-fault diagnostic problems in which faults interact and those in which faults are independent of one another. Of the two types, the latter is computationally the least expensive to solve (Bylander et al. 1989).

This analysis shows some of the experts' behaviors can be rationalized on computational grounds. The combination of conclusions concerning the abstract problem type (single antibody, multiple antibody, interacting antibody, and so on) and of knowledge strictly related to the domain of blood banking, such as general classes of antibodies, constrains the set of possible solutions to a subset of solutions that satisfy these multiple constraints. The use of information concerning the abstract problem type provides a way to select a more computationally attractive problem-solving method. These steps correspond to the assumption and confirmation of three major characteristics that make an abductive problem more computationally feasible. These characteristics are the existence of only a small number of plausible candidate elementary hypotheses, the absence of any interaction between causal processes associated with different hypotheses, and the existence of a single-hypothesis solution.

In chapter 7, I describe the addition of two heuristics that match some of the blood banker's behavior to the RED system (Fischer et al. 1990). I also report on experiments with the resulting systems. With 42 alloantibody identification cases solved, the new methods did better than the old RED system in both the efficiency of problem solving and the quality of the final answer. In chapter 8, I compare my work to previous

efforts to study the relevance of computational complexity to the study of human cognition (Newell 1982; Marr 1982; Chandrasekaran 1988; Anderson 1990) and to work in abductive reasoning (Mill 1855; Pople 1977).

The Contributions

There are two novel aspects of this work. The relevance of computational complexity theory to the study and understanding of human reasoning is shown. In problem solving, meta-abduction is introduced, abductive reasoning is driven by multiple types of knowledge, a strictly layered abductive strategy is developed and implemented, and two types of single-fault assumption are distinguished.

I described how human behavior can be rationalized and better understood by comparing it to the computational complexity analysis of the task performed by the problem solver. Other research has had a similar intuition concerning the relevance of computational complexity in the understanding of human reasoning (Chandrasekaran 1988; Anderson 1990; Newell 1982), but none had attempted to rationalize empirical data on human reasoning using complexity theory.

Based on the behavior of human immunohematologists, a more abstract level of abductive reasoning was added to RED. This new level of reasoning produces domain-independent hypotheses concerning the abstract nature (single antibody, multiple antibody, interacting antibody, and so on) of the specific problem that is being solved. Abductive problem solving can now be seen as having two levels. At one level, abductive hypotheses are made that concern the abstract domain-independent type of problem instance at hand. At the second level, hypotheses are proposed with respect to what domain-specific hypotheses (in my case antibodies) are present. Both levels produce abductive hypotheses. By analogy with MOLGEN's metaplanning (Stefik 1980), the abstract level of abductive reasoning is performing meta-abduction. The hypotheses produced by the first, domain-independent level of abductive reasoning are used as input in the second, domain-dependent level. Thus, this work can be seen as a model and implementation of a truly layered abductive system (Fischer et al. 1991).

The behavior identified in the pro-

tocols and their implementation in RED shed new light on single-fault assumption. This term is generally understood as a limitation of a system's capabilities to solving only single-fault problems. The single-fault assumption made in this work is now dynamically performed at problem-solving time to optimize performance. The problem solver can still handle multiple-fault problems. In these cases, based on the features of the particular problem, it cannot make the single-fault assumption and, therefore, must work harder at finding a solution.

Acknowledgments

Note

1. Copies of this dissertation are available from Sandra D. Thomas, Department of Computer and Information Science, The Ohio State University, 2036 Neil Avenue, Columbus, OH 43210.

References

Anderson, J. 1990. *The Adaptive Character of Thought*. Hillsdale, N. J.: Lawrence Erlbaum.

Bylander, T.; Allemang, D.; Tanner, M.; and Josephson, J. 1989. Some Results Concerning the Computational Complexity of Abduction. In *KR-89: Principles of Knowledge Representation and Reasoning—Proceedings of the First International Conference*, 45–54. San Mateo, Calif.: Morgan Kaufmann.

Chandrasekaran, B. 1988. What Kind of Information Processing Is Intelligence? A Perspective on AI Paradigms and a Proposal. In *Source Book on the Foundations of AI*, eds. D. Partridge and Y. Wilks, 14–46. Cambridge, U.K.: Cambridge University Press.

Charniak, E. 1988. Motivation Analysis, Abductive Unification, and Non-Monotonic Equality. *Artificial Intelligence* 34(3): 275–295.

Fischer, O. 1991. Cognitively Plausible Heuristics to Tackle the Complexity of Abductive Reasoning. Ph.D. diss., Dept. of Computer and Information Science, The Ohio State Univ.

Fischer, O.; Smith, J.; Svirbely, J.; and Smith, P. 1991. Meta-Abduction: A Cognitively Plausible Domain-Independent Computational Model of Abductive Reasoning, Technical Report, 91-OF-META, Laboratory for Artificial Intelligence, The Ohio State Univ.

Fischer, O.; Smith, J.; Svirbely, J.; and Smith, P. 1990. Integrating Problem and Knowledge Search in Abductive Reasoning, Technical Report, 90-OR-ABDREAS, Laboratory for Artificial Intelligence, The Ohio State Univ.

Josephson, J.; Chandrasekaran, B.; Smith, J.; and Tanner, M. 1987. A Mechanism for Forming Composite Explanatory Hypotheses. *IEEE Transactions on Systems, Man, and Cybernetics* 17(3): 445–454.

Marr, D. 1982. *Vision*. San Francisco: Freeman.

Mill, J. 1855. *Systems of Logic, Ratiocinative and Inductive*. New York: Harper & Row.

Newell, A. 1982. The Knowledge Level. *Artificial Intelligence* 18(1): 87–127.

Pearl, J. 1987. Distributed Revision of Composite Beliefs. *Artificial Intelligence* 33(2): 173–215.

Pople, H. 1977. The Formation of Composite Hypotheses in Diagnostic Problem Solving: An Exercise in Synthetic Reasoning. In Proceedings of the Fifth International Joint Conference on Artificial Intelligence, 1030–1037. Menlo Park, Calif.: International Joint Conferences on Artificial Intelligence.

Smith, J., and Fischer, O. 1990. Towards Taxonomies of Abductive Systems at the Knowledge and Symbol Level. Presented at 1990 AAAI Spring Symposium on Automated Abduction, Stanford, Calif., March 27–29.

Smith, J.; Svirbely, J.; Evans, C.; Strohm, P.; Josephson, J.; and Tanner, M. 1985. RED: A Red Cell Antibody Identification Expert Module. *Journal of Medical Systems* 9(3): 121–138.

Smith, P.; Smith, J.; Fraser, J.; Svirbely, J.; Krawczak, D.; Rudmann, S.; Miller, T.; and Blazina, J. 1989. Coping with the Complexities of Multiple-Solution Problems: A Case Study, Technical Report, 89-PS-COPING, Laboratory for Artificial Intelligence Research, The Ohio State Univ.

Stefik, M. 1980. Planning with Constraints, Ph.D. diss., Dept. of Computer Science, Stanford Univ.

Olivier Fischer is an instructor in the Department of Computer and Information Science at The Ohio State University, Columbus, Ohio. He obtained his Ph.D. in computer and information science from The Ohio State University in 1991. His current interests center on task-specific architectures, diagnostic reasoning, and decision support systems in management.

Cognitively Plausible Heuristics to Tackle the Computational Complexity of Abductive Reasoning

Olivier Fischer

The work described in my Ph.D. dissertation (Fischer 1991)[1] merges computational and cognitive investigations of abductive reasoning. It is the outcome of seven years of research focusing on abductive explanation generation and involving the departments of computer and information science, industrial and systems engineering, pathology, and allied medical professions at The Ohio State University. In the first phase of my work, I characterized abductive problem solving and performed a comparative analysis of two abductive problem solvers (Smith and Fischer 1990). I then summa-

> *The work… merges computational and cognitive investigations of abductive reasoning.*

rized three lines of work on abduction: a knowledge-based approach (Smith et al. 1985), the formal computational complexity analysis of abductive reasoning (Bylander et al. 1989), and the gathering and study of protocols of human experts performing an abductive task (Smith et al. 1989). The unification of these three approaches shows the relevance of computational complexity to understanding human behavior and brought me to enhance the existing knowledge-based system in the Department of Computer and Information Science by integrating cogni-

tive and computational efficiency issues into its design. Thus, I implemented two cognitively plausible heuristics to tackle the complexity of abductive reasoning and successfully experimented with them. This work, originally applied to the domain of alloantibody identification, was generalized to domain-independent abductive problem solving.

Abduction

Abduction, that is, inference to a hypothesis that best explains a set of data, appears to be ubiquitous in cognition. For example, diagnosis (Pople 1977), plan recognition (Charniak 1988), and hypothesis revision (Pearl 1987) can all be viewed as instances of the more general abduction problem.

Given a set of data D to be explained, if hypotheses that can individually explain all of D are directly available, then the hypothesis that explains D most plausibly is the best explanation for it. In general, however, a hypothesis explaining all of D is not directly available. Instead, a best composite hypothesis has to be constructed from a set of elementary hypotheses, where each selected elementary hypothesis can plausibly account for some of D.

Although there is no general agreement on the characterization of the best composite explanation, it appears reasonable (Josephson et al. 1987) that one composite hypothesis $H1$ is a better explanation than another hypothesis $H2$ if $H1$ explains more of D than $H2$, $H1$ is more plausible than $H2$, and $H1$ is less redundant an explanation of D than $H2$.

A Computational Complexity Analysis of Abduction

Chapter 5 of my dissertation describes an analysis of the computational complexity of abductive problems (Bylander et al. 1989). This complexity analysis defines abduction as inference to the best explanation given a set of data to explain. The complexity of reaching such an explanation varies according to the properties of the elementary hypotheses available. The study (Bylander et al. 1989) centers on the influence that some of these properties have on the computational complexity of assembling a composite hypothesis. The properties analyzed are hypothesis independence, monotonicity of the hypotheses' explanatory power, incompatibility between hypotheses, cancellation effect between causal processes associated with different hypotheses, and existence of an ordering of the plausibilities of the hypotheses to be true.

The Protocols

Although the computational complexity of abductive reasoning varies depending on the previously defined properties, abduction is, in general, NP-complete relative to the complexity of computing the data explained by hypotheses. However, protocols show that humans successfully perform this task. Thus, I was led to hypothesize the existence of heuristics that make abduction tractable and that would be related to the properties cited earlier. I studied two of these heuristics and how they are used in the abductive task of alloantibody identification.

A patient in need of a transfusion not only needs blood but requires blood compatible with his/her immune system. Blood bankers analyze the patient's blood to determine which blood in the bank can safely be transfused. Part of this analysis centers on the determination of what antibodies are present or can be produced in the patient's blood and what antigens are present in the donor's cells. The presence of one (or more) antibody in the patient's blood is inferred from reactions between samples of the recipient's blood and a set of carefully selected cells usually bought from specialized laboratories. The task of alloantibody identification was modeled as an instance of