Contents

C ABSTRACT **2**

D INTRODUCTION **3**

E TECHNICAL DISCUSSION **5**
E.1 Problem Statement – 5
E.2 Progress Report (March 1, 1997 – February 29, 2000) 5
E.2.1 Summary of research progress 5
E.2.2 List of publications resulting from the micro award (March 1, 1997 – February 29, 2000) 8
E.2.3 Doctoral dissertations supported by the MICRO award (1997–1999) 10
E.3 Proposed Work 10
E.3.1 Overview 10
E.3.2 Qualitative planning under uncertainty 11
E.3.3 Reasoning With Counterfactuals 12
E.3.4 Evaluating counterfactuals 13
E.3.5 Generating explanations 14
E.3.6 Learning Causal Structures 16

F RELEVANCE TO MICRO **23**

G PERSONNEL **24**

H BUDGET **25**

I JUSTIFICATIONS FOR MAJOR ITEMS **26**

J CURRENT AND PENDING SUPPORT **27**

K LETTER OF INTENT **28**

L LIST OF SUGGESTED REVIEWERS **29**

C ABSTRACT

This proposal requests support for a one year continuation of MICRO Grant 98-118, currently sponsored by Rockwell International. The aim of the proposed research is to develop computer systems capable of operating autonomously in dynamic and uncertain environments. Specifically, we propose to conduct theoretical and experimental studies in the following areas:

1. Information fusion, situation assessment, diagnosis, and planning under uncertainty using causal and counterfactual relationships.

2. Automatic generation of natural language explanations of actions, recommendations, and unexpected eventualities.

3. Learning causal structures from data to facilitate predictions and decisions in e-commerce applications.

D INTRODUCTION

Since the development of belief-network representations in the early 1980's, there has been an upsurge of interest in reviving probabilistic formalisms for AI applications. This interest reflects the recognition that the important aspects of probabilistic knowledge can be expressed in network form (later called Bayesian Belief Networks) and that computations can exploit the topology of these networks. This capability leads to: simplicity of knowledge acquisition, reduction in inference complexity, coherent updating of beliefs, production of meaningful explanations and a reasonable model of cognitive behavior.

The basic technical background of this development is described in several texts [Pearl, 1988; Castillo et al., 1997; Jensen, 1996] which formulate the construction and uses of Belief Networks and demonstrate the feasibility of updating uncertainties and guiding decisions by local computations. A major advantage of basing a reasoning system on a probabilistic foundation is the ability to automatically expand the system's knowledge as more data is obtained. Indeed, the past few years have seen an upsurge of research toward augmenting belief network systems with learning capabilities, i.e., inferring network structures and probability values from empirical data [Pearl and Verma, 1991; Spirtes et al., 1993; Cooper and Hershkovitz, 1990; Heckerman et al., 1994].

A new dimension has been added to Bayesian network research with the introduction of causal interpretation of the network's topology [Pearl and Verma, 1991; Pearl, 1993; Druzdzel and Simon, 1993; Spirtes et al., 1993]. Since the bulk of human knowledge is encoded in the form of causal, rather than statistical relationships, this interpretation clarifies the assumptions embedded in the network, and greatly facilitates the construction Bayesian networks from experts as well as reconfiguring the network to track structural changes in the domain. In addition, the causal interpretation provides an economical encoding of the effect of interventions, thus enabling the analysis of policies [Pearl, 1994] and counterfactuals [Balke and Pearl, 1994, 1995; Breese and Heckerman, 1996; Heckerman and Shachter, 1994].

The research described in this proposal aims at expanding the capabilities of causal Bayesian networks along two avenues.

The first project will aim at developing methods for qualitative planning under uncertainty. Research in this area has shown that many of the features that made probabilistic reasoning powerful can be retained in a symbolic approximations of probabilities [Goldszmidt and Pearl, 1996; Darwiche and Pearl, 1997]. These order-of-magnitude approximations (i.e., associating "q is believed" with "$P(\neg q) = \epsilon$"; ϵ being infinitesimal) facilitate reasoning with qualitative rules, facts, and deductively closed beliefs (as in logic), yet permit us to retract beliefs in response to changing contexts (as in probability). Combining this facility with an order-of-magnitude approximation of utilities [Pearl, 1993; Wilson, 1995; Tan, 1994] has yielded a qualitative version of decision theory, thus forming a basis for symbolic planning and control under uncertainty. We now seek to tie this formalism with the powerful planning method proposed by Kautz and Selman [1996], which treats planning as a propositional satisfiability problem, so as to manage planning in complex domains under conditions of uncertainty. This capability, which has many applications in manufacturing and process control, will be based on a qualitative-probability analysis of counterfactuals, that is, sentences for the form "It is unlikely that Y would have been different had X been enacted".

A second project will focus on developing explanation capabilities based on the action and counterfactual semantics offered by causal networks. This investigation is motivated by the realization that the notion of *explanation*, which so far have been given logical or probabilistic interpretations, is due for drastic reformulation, to take account of causal and counterfactual considerations and to exploit the operationalization of causal and counterfactual inferences using network representation. The immediate beneficiaries would be systems that attempt to explain their actions or recommendations using a natural and friendly discourse such as complex diagnostic systems.

E TECHNICAL DISCUSSION

E.1 Problem Statement –

The long term objectives of this project are to develop a theory of causality, specific enough for machine interpretation, so as to guide the construction of computers program capable of planning, perceiving, and learning in uncertain dynamic environment. We believe this task is realizable and that causal graphs will play a key role in its realization. To that end we propose to develop new theories and techniques of reasoning with causal networks, primarily those that exploit the counterfactual inferencing capabilities that these networks provide. The main application of these techniques will be in the construction of flexible plans the synthesis of reliable models of system users, and the generation of natural explanations of both physical events and inferencing steps. The propose project will focus on developing a theory learning based on spontaneous changes and a theory of explanation based on counterfactuals.

E.2 Progress Report (March 1, 1997 – February 29, 2000)

E.2.1 Summary of research progress

Starting with functional description of physical mechanisms we were able to derive the standard probabilistic properties of Bayesian networks and to show:

- how the effects of unanticipated actions can be predicted from the network topology,
- how qualitative causal judgments can be integrated with statistical data,
- how actions interact with observations,
- how counterfactuals sentences can be interpreted and evaluated,
- how explanations and single-event causation can be defined in a given causal model.

Additionally, we have established an axiomatic characterization of causal dependencies, analogously to the characterization of informational dependencies. Finally, we have demonstrated that network-based identification techniques, in the presence of hidden variables, have a broad scope of new applications, ranging from skill acquisition by autonomous agents, to the analysis of treatment effectiveness in clinical trials.

The following specific results were obtained during the period of performance:

- Computer programs were developed to assist clinicians with assessing the efficacy of treatments in experimental studies for which subject compliance is imperfect [Chickering and Pearl, 1999].

- Axiomatic characterization was given for causal-relevance relationships of the form: "Changing X will not affect Y if we hold Z constant" [Galles and Pearl, 1997]

- The notion of "identification" was extended to non-parametric systems and techniques were developed for non-parametric identification of cause-effect relationships from nonexperimental data [Pearl, 1997].

- Methods were developed for selecting sufficient set of measurements that permit unbiased estimation of causal effects in observational studies [Greenland *et al.*, 1999].

- Polynomial algorithms were developed for finding *minimal separators* in a directed acyclic graphs, namely, finding a set S of nodes that d-separates a given pair nodes, such that no proper subset of S d-separates that pair. Versions of this problem include finding a minimal separator from a restricted set of nodes, finding a minimum-cost separator, and testing whether a given separator is minimal. We have confirmed the intuition that any separator which cannot be reduced by a single node must be minimal [Tian *et al.*, 1998].

- Universal bounds were established for the effectiveness of policies from imperfect experiments [Balke and Pearl, 1997].

- Methods for estimating or bounding counterfactual probabilities from statistical data were developed (e.g., John, who was treated and died, would have had 90% chance of survival had he not been treated) [Balke and Pearl, 1997].

- A formal model has been developed, based on *modifiable structural equations*, which generalizes and unifies the structural and counterfactual approaches to causal inference, explicates their conceptual and mathematical bases and resolves their technical difficulties [Galles and Pearl, 1998].

- It has been proven that the structural and counterfactual formalisms are equivalent in recursive causal models (i.e., systems without feedback) but not when feedback is considered possible. A simple rule was devised for translating a problem back and forth, between the structural and counterfactual representations [Galles and Pearl, 1998].

- Basic causal concepts such as "confounding" and "exogeneity" were given mathematically precise explication. It has been shown that, contrary to folklore, there is no statistical test for confounding. Traditional statistical criteria do not ensure unbiased effect estimates, nor do they follow from the requirement of unbiasedness [Greenland et al., 1999; Pearl, 2000].

- A new semantics for "actual causation" was developed based on a construct named "causal beam," that is, a minimally modified causal model, in reference to which the standard counterfactual criterion is adequate for identifying causes of singular events [Pearl, 1998a, 2000].

- Formal semantics was developed, based on structural models of counterfactuals, for the probabilities that event x is a *necessary* or *sufficient* cause (or both) of another event y [Pearl, 1999].

- Conditions were discovered under which probabilities of necessary and sufficient causation can be learned from data [Pearl, 1999; Tian and Pearl, 2000].

- New methods were developed for eliciting probabilities of causes from a combination of actions and observations. It was found that data from both experimental and nonexperimental studies can be combined to yield information that neither study alone can provide [Pearl, 1999].

- New definition of *causal explanation* was formulated in which explanation is treated as a fragment of knowledge needed to support causation [Halpern and Pearl, 2000].

E.2.2 List of publications resulting from the micro award (March 1, 1997 – February 29, 2000)

Darwiche, A. & Pearl, J., "On the Logic of Iterated Belief Revision," *Artificial Intelligence*, 89(1-2), 1–29, 1997.

Pearl, J. "Causation, Action, and Counterfactuals," In M.L. Dalla Chiara et al. (Eds.), *Logic and Scientific Methods*, Kluwer Academic Publishers, Netherlands, 355–375, 1997.

Pearl, J., "On the Identification of Nonparametric Structural Models," in M. Berkane (Ed.), *Latent Variable Modeling with Application to Causality Conference*, Springer-Verlag, Lecture Notes in Statistics, 29–68, 1997.

Pearl, J., "Graphical Models for Probabilistic and Causal Reasoning," in Allen B. Tucker, Jr. (Ed.), *The Computer Science and Engineering Handbook*, Chapter 31, CRC Press, Inc., 697–714, 1997.

Pearl, J., "The New Challenge: From a Century of Statistics to an Age of Causation," *Computing Science and Statistics*, 29(2), 415–423, 1997.

Galles, D. & Pearl, J., "Axioms of Causal Relevance," *Artificial Intelligence*, 97(1-2), 9–43, 1997.

Balke, A. & Pearl, J., "Bounds on Treatment from Studies with Imperfect Compliance," *Journal of the American Statistical Association (JASA)*, 92(439), 1171–1176, 1997.

Galles, D., "Structural Causal Models: A Formalism for Reasoning About Actions and Counterfactuals," UCLA Cognitive Systems Laboratory, Technical Report (R-258), Ph.D. Thesis, 1997.

Pearl, J., "On the definition of actual cause," UCLA Computer Science Department, Technical Report (R-259), July 1998.

Pearl, J., "TETRAD and SEM," Commentary on "The TETRAD Project: Constraint Based Aids to Causal Model Specification" by R. Scheines, P. Spirtes, C. Glymour, C. Meek, and T. Richardson, in *Multivariate Behavioral Research*, Vol. 33 No. 1, 119–128, 1998.

Galles, D. & Pearl, J., "An Axiomatic Characterization of Causal Counterfactuals," *Foundations of Science*, Vol. 3, Issue 1, 151–182, 1998.

Pearl, J., "Graphs, Causality, and Structural Equation Models," *Sociological Methods and Research*, Vol. 27, No. 2, 226–284, November 1998.

Pearl, J., "Why There Is No Statistical Test For Confounding, Why Many Think There Is, and Why They Are Almost Right," UCLA Computer Science Department, Technical Report R-256, 1998.

Greenland, S., Pearl, J., and Robins, J., "Causal Diagrams for Epidemiological Research." *Epidemiology*, Vol. 1, No. 10, 37–48, January 1999.

Greenland, S., Robins, J., and Pearl, J. "Confounding and collapsibility in causal inference," *Statistical Science*, Vol. 14, No. 1, 29–46, 1999.

Tian, J., Paz, A., and Pearl, J., "Finding Minimal Separating Sets," UCLA Computer Science Department, Technical Report R-254, February 1998.

Pearl, J., "Simpson's paradox: An anatomy," UCLA Computer Science Department, Technical Report (R-264), March 1999.

Pearl, J. and Meshkat, P., "Testing Regression Models With Few Regressors," in D. Heckerman and J. Whittaker (Eds.), *Artificial Intelligence and Statistics 99*, Morgan Kaufmann, San Francisco, CA, 255–259, 1999.

Pearl, J., "Graphical Models for Probabilistic and Causal Reasoning," UCLA Computer Science Department, in D. Gabbay and P. Smets (Eds.), *Handbook on Defeasible Reasoning and Uncertainty Management Systems*, Vol. 1, 367–189, Kluwer Academic Publishers, the Netherlands, 1998.

Pearl, J., "Bayesian Networks," In R.A. Wilson and F. Keil (Eds.), *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge, MA, 72–74, 1999.

Pearl, J., "Graphs, Structural Models and Causality," In C.N. Glymour and G.F. Cooper (Eds.), *Computation, Causation, and Discovery*, AAAI/MIT Press, Cambridge, MA, 95–138, 1999.

Chickering, D.M. and Pearl, J., "A Clinician's Tool for Analyzing Noncompliance," In C.N. Glymour and G.F. Cooper (Eds.), *Computation, Causation, and Discovery*, AAAI/MIT Press, Cambridge, MA, 407–424, 1999.

Pearl, J., "Reasoning with cause and effect," *Proceedings of the International Joint Conference on Artificial Intelligence* (IJCAI-99), Morgan Kaufmann, San Francisco, CA, 1437–1449, 1999.

Pearl, J., "Probabilities of causation: Three counterfactual interpretations and their identification," *Synthese*, Vol. 121, No. 1, 1999.

Pearl, J., *Causality: Models, Reasoning, and Inference*, Cambridge University Press, NY, 2000.

E.2.3 Doctoral dissertations supported by the MICRO award (1997–1999)

Galles, D.J., "Structural Causal Models: A Formalism for Reasoning About Actions and Counterfactuals," June 1997.

E.3 Proposed Work

E.3.1 Overview

The ultimate goal of the proposed project will be the development of computer systems capable of:

1. Integrating sensory inputs into a coherent interpretation of a dynamic, uncertain environment,

2. Selecting actions and goals appropriate for the perceived environment, and

3. Learning to improve performance as more experience is gathered.

As concrete, realizable steps toward achieving these ambitious goals, we propose to undertake the following two projects:

- Developing algorithms for planning and learning under uncertainty, using qualitative approximations of probabilities and causal relationships.

- Developing a theory of explanation for improved reasoning and learning.

The theoretical issues underlying these two projects will be described in the following subsections.

E.3.2 Qualitative planning under uncertainty

The method of qualitative probabilities developed in the past few years [Goldszmidt, 1992] provides ways of combining logic and probabilities so as to achieve the benefits of both. In this method, beliefs are represented propositionally, as in classical logic, and yet are subject to retraction and to varying degrees of firmness, as in probability.

Quantitative probabilities are order-of-magnitude abstractions of numerical probabilities; instead of measuring probabilities on a scale from zero to one, we project probability measures onto a quantized logarithmic scale and then treat beliefs that map onto two different quanta as being of different orders of magnitude and, finally, take the limit and retain only the order-of-magnitude of each probability measure. Thus an integer $k = 0, 1, 2, \ldots$ attached to a proposition p signifies that not-p is believed to a degree k, or that the probability of p is of the order of ϵ^k. The result is a non-standard probability calculus on integers, with min replacing addition, and addition replacing multiplications. Since, in practice, only a few levels of k are of interest, the method above reduces many probabilistic inference to tasks to a finite sequence of logical operations, one per each level of k [Goldszmidt and Pearl, 1996; Darwiche and Pearl, 1997].

A qualitative version of decision theory can likewise be constructed by combining order-of-magnitude approximations of utilities and probabilities, thus providing the basis for qualitative planning under uncertainty [Pearl, 1993; Tan, 1994; Wilson, 1995]. The formalization of actions and persistence in the language of qualitative causal networks [Pearl, 1995a; Goldszmidt and Pearl, 1996; Darwiche and Pearl, 1997; Breese and Heckerman, 1996] has further facilitated the analysis of policies, their consequences, their interaction with observations, and their expected utilities, and, hence, the synthesis of plans and strategies under uncertainty.

We now propose to investigate the feasibility of tying this formalism with the powerful planning method of Kautz and Selman [1996], which treats planning as a propositional satisfiability problem. Specifically, a planning problem in the qualitative-probability representation will be translated into the problem of satisfying the proposition "an *effective* plan exits" where *effective* is defined as having expected utility of order $O(1)$ (i.e, $k = 0$). If no such plan exists, the computation will be repeated for the proposition: "a *risky* plan exits" where *risky* stands for expected utility of order $O(\epsilon^1)$ (i.e., $k = 1$), and so on.

Another issue to be investigated is the exploitation of the topological features of the causal network within the satisfiability-planning framework. Still another issue, to be discussed in the next subsection, is the computation and utilization of counterfactual information in this planning framework.

E.3.3 Reasoning With Counterfactuals

A counterfactual sentence has the form

If A were true, then C would have been true

where A, the counterfactual antecedent, specifies an event that is contrary to one's actual beliefs. A typical example is "If this voltage were low the system would have failed," which presumes the factual knowledge that the voltage is high, contrary to the antecedent of the sentence.

Counterfactual reasoning is at the heart of every planning activity, especially real-time planning. When a planner discovers that the current state of affairs deviates from the one expected, a "plan repair" activity need be invoked to determine what went wrong and how it could be rectified. This activity amounts to an exercise of counterfactual thinking, as it calls for rolling back the natural course of events and determining, based on the factual observations at hand, whether the culprit lies in previous decisions or in some unexpected, external eventualities. Moreover, in reasoning forward to determine if things would have been different a new model of the world must be consulted, one that embodies hypothetical changes in decisions or eventualities, hence, a breakdown of the old model or theory.

The evaluation of counterfactual sentences is applicable to other tasks as well. For example, determining liability of actions (e.g., "If you had not pushed the table, the glass would not have broken). In diagnostic tasks, counterfactual queries can be used to determine which tests to perform in order to increase the probability that faulty components are identified.

The logic-based planning tools used in AI, such as STRIPS and its variants or those based on the situation calculus, do not readily lend themselves to counterfactual analysis; as they are not geared for coherent integration of abduction with prediction, and they do not readily handle theory changes. Remarkably, the formal system developed in economics and social sciences under the rubric "structural equations models" does offer such capabilities, although these capabilities are not well recognized by current practitioners of structural models.

Recent research on *modifiable structural equations* has resulted in formal semantics, representational schemes, and inference algorithms that facilitate the probabilistic evaluation of counterfactual queries [Balke and Pearl, 1995; Galles and Pearl, 1997, 1998; Halpern, 1998]. World knowledge is represented in the language of causal networks, whose links represent functional mechanisms operating among families of observables. The antecedent of the query is interpreted as a proposition that is established by an external action, thus pruning the corresponding links from the network and facilitating standard Bayesian-network computation to determine the probability of the consequent [Balke and Pearl, 1995]. We propose to explore the computational feasibility of applying this procedure to planning problems, using qualitative probabilities.

E.3.4 Evaluating counterfactuals

A general counterfactual sentence can be written succinctly as

$$a \rightarrow c \mid o \tag{1}$$

read: "Given that we have observed o, if a were true, then c would have been true." The observations o consists of a set of value assignments to variables in a set V, e.g., $V_j = v_j, V_k = v_k$. The counterfactual antecedent a, consists of a conjunction of value assignments to variables in V that are forced to hold true by external intervention. Typically, to justify being called "counterfactual", a conflicts with o. Finally, the counterfactual consequent, c, stands for the proposition of interest, usually the values attained by some variables in the system.

The truth (or probability) of a counterfactual conditional $a \rightarrow c \mid o$ may then be evaluated by the following procedure:

- Use the observations o to update the joint belief[1] for all root nodes in the causal network. This joint belief summarizes the state of the system, because each non-root variable is a deterministic function of the root variables.

- Replace the structural equation for each variable V_k referred to in the antecedent a with the equation $V_k = a_{v_k}$ where a_{v_k} is the value of V_k specified in a. This implements the local intervention that forces the counterfactual antecedent to hold true.

- Compute the belief of the consequent proposition c according to the modified set of structural equations.

This procedure will yield a definite value for *Belief* (c) whenever we have the functional form of the mechanisms involved. In cases where the functional forms are not known, only bounds may be calculated for the belief of a counterfactual consequent. These considerations apply in both the probabilistic and qualitative formulations of beliefs [Balke, 1995].

We propose to investigate whether computational advantages could be achieved by casting the counterfactual evaluation problem as a problem in propositional satisfiability.

E.3.5 Generating explanations

It is a commonplace wisdom that explanation improve understanding, and that he who understands more, can reason and learn more effectively. The notion of explanation, on the other hand, is strongly associated with causal relationships, for which we currently have a computational theory. We therefore seek to apply this theory to the task of automatic generation of explanations for a given set of observations, and, subsequently, to investigate how reasoning and learning can be improved when such explanations are adopted.

The following list, taken from [Galles and Pearl, 1997], provides brief examples of concepts used in explanatory discourse and their associate semantics in the modifiable structural model. The notation used in this list is based on the counterfactual variable $Y_x(u)$ which reads: The value that Y would

[1]Here we use the generic term "belief" to refer to either truth assignments or probabilities.

take on in state $U = u$, had X been equal to x. The equation-deletion procedure described in Section E.3.4 permits us to calculate this variable from a set of structural equations. Likewise, if u is not known, the probabilities of counterfactuals, such as $P(Y_x = y \;\&\; Y_{x'} = y')$, can be computed given $P(u)$.

- "X is a cause of Y", if there exist two values x and x' of X and a value u of U such that $Y_x(u) \neq Y_{x'}(u)$.
- "X is a cause of Y in context $Z = z$", if there exist two values x and x' of X and a value u of U such that $Y_{xz} \neq Y_{x'z}(u)$.
- "X is a direct cause of Y", if there exist two values x and x' of X, and a value u of U such that $Y_{xr}(u) \neq Y_{x'r}(u)$ where r is some realization of $V \backslash X$.
- "X is an indirect cause of Y", if X is a cause of Y, and X is not a direct cause of Y.
- "Event $X = x$ may have caused $Y = y$" if
  - **(i)** $X = x$ and $Y = y$ are true, and
  - **(ii)** There exists a value u of U such that $X(u) = x$, $Y(u) = y, Y_x(u) = y$ and $Y_{x'}(u) \neq y$ for some $x' \neq x$.
- "The unobserved event $X = x$ is a likely cause of $Y = y$" if
  - **(i)** $Y = y$ is true, and
  - **(ii)** $P(Y_x = y, Y_{x'} \neq y | Y = y)$ is high for some $x' \neq x$
- "Event $Y = y$ occurred despite $X = x$", if
  - **(i)** $X = x$ and $Y = y$ are true, and
  - **(ii)** $P(Y_x = y)$ is low.

The preceding list demonstrates the flexibility of modifiable structural models in formalizing nuances of causal expressions. Additional nuances, invoking notions such as *enabling*, *preventing*, *maintaining*, and *producing*, etc. should be formalized as well. We propose to implement this semantics in a system that automatically selects the appropriate explanatory expression, for a given context and for a given query.

Additionally, we will construct semantical and axiomatic characterization of explanatory sentences, for example: "Event A explains the occurrence of event B", or "A would explain B if C were the case", or "B occurred despite of A, because C was true". Such explanatory sentences should be generated automatically by a reasoning program, and used to guide future information-gathering actions, in the pursuit of causal understanding of the environment. Additionally, the ability to generate such explanatory sentences, or to select the expression most appropriate for the context will improve the effectiveness of man-machine conversation.

E.3.6 Learning Causal Structures

The possibility of learning causal relationships from raw data has been on philosophers' dream lists since the time of Hume (1711–1776). That possibility entered the realm of formal treatment and feasible computation in the mid-1980s, when the mathematical relationships between graphs and probabilistic dependencies came into light [Pearl, 1988]. Several systems have been developed for this purpose [Pearl and Verma, 1991, Spirtes *et al.*, 1993], which systematically search and identify causal structures (with hidden variables) from empirical data. Technically, because these algorithms rely merely on conditional independence relationships, the structures found are valid only if one is willing to accept weaker forms of guarantees than those obtained through controlled randomized experiments: minimality and stability [Pearl and Verma, 1991]. Minimality guarantees that any other structure compatible with the data is necessarily less specific, and hence less testable and less trustworthy, than the one(s) inferred. Stability ensures that any alternative structure compatible with the data must be less stable than the one(s) inferred; namely, slight fluctuations in experimental conditions will render that structure no longer compatible with the data. With these forms of guarantees, the theory provides criteria for identifying genuine and spurious causes, with or without temporal information.

Alternative methods of identifying structure in data assign prior probabilities to the parameters of the network and use Bayesian updating to score the degree to which a given network fits the data [Cooper and Herskovits, 1990, Heckerman *et al.*, 1994]. These methods have the advantage of operating well under small sample conditions, but encounter difficulties coping with hidden variables.

Causal structures in e-commerce applications

The central aim of data analysis in e-commerce is to form a reliable model of consumer behavior, so as to predict interests in future transactions. Many statistical routines are being developed for this purpose, often under the enterprises of "data mining" or "knowledge mining," but only few are designed to build models on causal relationships that can be inferred from the data. The general attitude is that statistical associations alone would be sufficient, since the task is one of prediction, rather than manipulation.

We believe this attitude to be erroneous. First, pure black-box predictions are not as useful as those that are accompanied with causal understanding of the underlying processes. When a statistical package predicts that a group of users will be likely to demand a certain product in the future, the question always arises whether the association discovered is long-lived, and whether it is transportable across contexts. For example, if one product is functionally supplementary to another, the association between the two demands is stable. If, on the other hand, demands for products A and B are correlated merely because the two were advertised simultaneously in the same medium, the association is short lived, and will disappear as soon as advertising strategies change.

Second, consumer behavior models are not used exclusively for passive predictions. Vendors constantly try new techniques of presentation, and new methods of capturing users' attention. These changes are the commercial analogue of scientific experimentation, and only causal models can capture the results of these experiments so as to predict response to future changes.

Finally, there is an additional advantage to basing even purely predictive decisions on causal, rather than associational models. The advantage involves considerations of "locality." When some conditions in the environment undergo change, it is usually only a few causal mechanisms that are affected by the change; the rest remain unaltered. It is simpler and more effective, then, to reassess (judgmentally) or reestimate (statistically) the model parameters knowing that the corresponding change in the model is also local, involving just a few parameters, than to reestimate the entire model from scratch. In non-causal systems, such as neural nets or those based on regression equations, a local change in mechanism space would spread its effect over all model parameters; in causal systems the change remains local.

Learning from spontaneous changes

The innovative contribution that this proposal makes to knowledge discovery involves a new method of discovering causal relations in data, based on the detection and interpretation of local spontaneous changes in the environment. While all systems of causal discovery are static, that is, assuming a time-invariant distribution and a time-invariant data-generating model, our proposal aims at exploiting dynamic changes in the environment. Such changes are always present in an environment of consumers that is embedded in a larger context such as the general economy. Whereas static analysis views these changes as nuisance, and (attempts) to adjust and compensate for them, we view these changes as an invaluable source of information about the causal structure of the data-generating process.

The basic idea has its roots in the economic literature of the 1980's. The economist Kevin Hoover (1990) inferred the direction of causal influences among economic variables (e.g., employment and money supply) by observing the changes that sudden modifications in the economy (e.g., tax reform, labor dispute) induced in the statistics of these variables. Hoover assumed that the conditional probabilities of an effect given its causes remains invariant to structural changes in the mechanism that generates the cause, while the conditional probability of a cause given the effect would not remain invariant under such changes. Indeed, today we understand more precisely the conditions under which such asymmetries would prevail and how to interpret such asymmetries in the context of large, multi variariate systems. Whenever we obtain reliable information (e.g., from historical or institutional knowledge) that an abrupt local change has taken place in a specific mechanism f_i that constrains a given family $(X_1, \ldots, X_n)$ of variables, we can use the observed changes in the marginal and conditional probabilities surrounding those variables to determine whether X_i is indeed the dependent variable in that family, thus determining the direction of causal influences in the domain. The statistical features that remain invariant under such changes, as well as the causal assumptions underlying this invariance, are displayed vividly and formally in the causal diagram at hand, and can be used therefore for testing the validity of a given structure, and for automatic restructuring of its topology.

We propose to initiate a theoretical and experimental program to exploit these new possibilities in adaptive systems that learn causal structures and causal parameters. We propose to test these systems in e-commerce applications and to compare their performance to static learning systems.

References

[Balke and Pearl, 1994] A. Balke and J. Pearl. Probabilistic evaluation of counterfactual queries. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, volume I, pages 230–237. MIT Press, Menlo Park, CA, 1994.

[Balke and Pearl, 1995] A. Balke and J. Pearl. Counterfactuals and policy analysis in structural models. In P. Besnard and S. Hanks, editors, *Uncertainty in Artificial Intelligence 11*, pages 11–18. Morgan Kaufmann, San Francisco, 1995.

[Balke and Pearl, 1997] A. Balke and J. Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1172–1176, September 1997.

[Balke, 1995] A. Balke. *Probabilistic Counterfactuals: Semantics, Computation, and Applications.* PhD thesis, Computer Science Department, University of California, Los Angeles, CA, November 1995.

[Breese and Heckerman, 1996] J.S. Breese and D. Heckerman. Decision-theoretic troubleshooting: A framework for repair and experiment. In E. Horvitz and F. Jensen, editors, *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, pages 124–132. Morgan Kaufmann, San Francisco, CA, 1996.

[Castillo *et al.*, 1997] E. Castillo, J.M. Gutierrez, and A.S. Hadi. *Expert Systems and Probabilistic Network Models.* Springer-Verlag, New York, 1997.

[Chickering and Pearl, 1999] D.M. Chickering and J. Pearl. A clinician's tool for analyzing noncompliance. In C.N. Glymour and G.F. Cooper, editors, *Computation, Causation, and Discovery*, pages 407–424. AAAI/MIT Press, Cambridge, MA, 1999.

[Cooper and Herskovits, 1990] G.F. Cooper and E. Herskovits. A Bayesian method for constructing Bayesian belief networks from databases. *Proceedings of the Conference on Uncertainty in AI*, pages 86–94, 1990.

[Darwiche and Pearl, 1997] A. Darwiche and J. Pearl. On the logic of iterated belief revision. *Artificial Intelligence*, 89(1–2):1–29, 1997.

[Druzdzel and Simon, 1993] M.J. Druzdzel and H.A. Simon. Causality in Bayesian belief networks. In D. Heckerman and A. Mamdani, editors, *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, pages 3–11, San Mateo, CA, 1993. Morgan Kaufmann.

[Galles and Pearl, 1997] D. Galles and J. Pearl. Axioms of causal relevance. *Artificial Intelligence*, 97(1-2):9–43, 1997.

[Galles and Pearl, 1998] D. Galles and J. Pearl. An axiomatic characterization of causal counterfactuals. *Foundation of Science*, 3(1):151–182, 1998.

[Goldszmidt and Pearl, 1996] M. Goldszmidt and J. Pearl. Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artificial Intelligence*, 84(1–2):57–112, July 1996.

[Goldszmidt, 1992] M. Goldszmidt. Qualitative probabilities: A normative framework for commonsense reasoning. Technical Report R-190, Computer Science Department, UCLA, 1992.

[Greenland *et al.*, 1999] S. Greenland, J.M. Robins, and J. Pearl. Confounding and collapsibility in causal inference. *Statistical Science*, 14(1):29–46, February 1999.

[Halpern and Pearl, 2000] J.Y. Halpern and J. Pearl. Actual causality. Technical Report R-266, Department of Computer Science, University of California, Los Angeles, CA, March 2000.

[Halpern, 1998] J.Y. Halpern. Axiomatizing causal reasoning. In G.F. Cooper and S. Moral, editors, *Uncertainty in Artificial Intelligence*, pages 202–210. Morgan Kaufmann, San Francisco, CA, 1998.

[Heckerman and Shachter, 1994] D. Heckerman and R. Shachter. A decision-based view of causality. In *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, pages 302–310, San Francisco, CA, 1994. Morgan Kaufmann.

[Heckerman *et al.*, 1994] D. Heckerman, D. Geiger, and D. Chickering. Learning Bayesian networks: The combination of knowledge and statistical data. In R. Lopez de Mantaras and D. Poole, editors, *Uncertainty in Artificial Intelligence 10*, pages 293–301. Morgan Kaufmann, San Mateo, CA, 1994.

[Hoover, 1990] K.D. Hoover. The logic of causal inference. *Economics and Philosophy*, 6:207–234, 1990.

[Jensen, 1996] Finn V. Jensen. *An Introduction to Bayesian Networks.* Springer, New York, 1996.

[Kautz and Selman, 1996] H. Kautz and B. Selman. Pushing the envelope: Planning, propositional logic, and stochastic search. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, volume Volume II, pages 1194–1201. MIT Press, Menlo Park, CA, 1996.

[Pearl and Verma, 1991] J. Pearl and T. Verma. A theory of inferred causation. In J.A. Allen, R. Fikes, and E. Sandewall, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, pages 441–452. Morgan Kaufmann, San Mateo, CA, 1991.

[Pearl, 1988] J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, San Mateo, CA, 1988.

[Pearl, 1993] J. Pearl. From conditional oughts to qualitative decision theory. In D. Heckerman and A. Mamdani, editors, *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, pages 12–20, San Mateo, CA, July 1993. Morgan Kaufmann Publishers.

[Pearl, 1994] J. Pearl. A probabilistic calculus of actions. In R. Lopez de Mantaras and D. Poole, editors, *Uncertainty in Artificial Intelligence 10*, pages 454–462. Morgan Kaufmann, San Mateo, CA, 1994.

[Pearl, 1995] J. Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–710, December 1995.

[Pearl, 1997] J. Pearl. On the identification of nonparametric structural models. In M. Berkane, editor, *Latent Variable Modeling with Application to Causality*, pages 29–68. Springer-Verlag, 1997.

[Pearl, 1998] J. Pearl. On the definition of actual cause. Technical Report R-259, Department of Computer Science, University of California, Los Angeles, CA, 1998.

[Pearl, 1999] J. Pearl. Probabilities of causation: Three counterfactual interpretations and their identification. Technical Report R-261, Department of Computer Science, University of California, Los Angeles, CA, 1999. Forthcoming, *Synthese*, Vol. 121, No. 1, 2000.

[Pearl, 2000] J. Pearl. *Causality: Models, Reasoning, and Inference.* Cambridge University Press, New York, 2000.

[Spirtes *et al.*, 1993] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search.* Springer-Verlag, New York, 1993.

[Tan, 1994] S-W. Tan. Qualitative decision theory. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, volume Volume II, pages 928–933. MIT Press, Menlo Park, CA, 1994.

[Tian and Pearl, 1999] J. Tian and J. Pearl. Probabilities of causation: Bounds and identification. Technical Report R-271, Department of Computer Science, University of California, Los Angeles, CA, 1999.

[Tian *et al.*, 1998] J. Tian, A. Paz, and J. Pearl. Finding minimal separating sets. Technical Report R-254, University of California, Los Angeles, CA, 1998.

[Wilson, 1995] N. Wilson. An order of magnitute calculus. In P. Besnard and S. Hanks, editors, *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 548–555, San Francisco, CA, 1995. Morgan Kaufmann.

F RELEVANCE TO MICRO

The problem of reasoning and acting under uncertainty, lies at the center of computer automation and, in this general sense, the proposed research is central to the objectives of the MICRO program. More specifically, the introduction of Bayes networks as the primary scheme for representing uncertainty in computer systems owes much of its development to the research conducted at UCLA, partly supported by MICRO projects beginning 1988. First commercial applications of Bayesian networks were found in medical diagnosis and include systems such as PATHFINDER, INTELLIPATH and CPSC. Currently, these systems are used by hundreds of hospitals and medical schools nation wide. Another application system, in the area of power-generator monitoring (GEMS) has been developed by General Electric (Schenectady, NY) in collaboration with EPRI (Palo Alto CA), and is available commercially. Pilot systems in such diverse applications as software debugging, information retrieval and system troubleshooting are described in the March, 1995 issue of the CACM (Special issue on practical applications of Bayesian networks).

The cooperating company, BizRate.com is a California corporation that is one of the leading developers of E-commerce systems and services. BizRate.com has several projects in-house aimed at improved user interface, based on data mining and advanced models of user preference dynamics. The technical liason person from the cooperating company is:

Dr. George Rebane
VP, Advanced Projects
BizRate.com
4053 Redwood Avenue phone: (310) 305-3506
Los Angeles, CA 90066 fax: (310) 305-7737

G PERSONNEL

The major portion of the proposed research will be carried out by the Principal Investigator, Professor Judea Pearl and two Ph.D. students at the Computer Science Department of UCLA. The following is a biographical sketch of the principal investigator.

Judea Pearl is a Professor of Computer Science at UCLA where he also is the Director of the Cognitive Systems Laboratory.

He received the B.S. degree in Electrical Engineering from Technion-Israel Institute of Technology, Haifa, Israel, in 1960; the MS.C. degree in physics from Rutgers University, New Brunswick, New Jersey, in 1965; and the Ph.D. degree in Electrical Engineering from the Polytechnic Institute of Brooklyn, Brooklyn, NY in 1965.

Before coming to UCLA, he worked at RCA Laboratories, Princeton, New Jersey, on super-conductive parametric and storage devices, and at Electronic Memories, Inc., Hawthorne, California, on advanced memory advises. His present interests include Knowledge-representation, probabilistic reasoning, constraint processing, non-standard logics, distributed computation, and learning.

Professor Pearl serves on the editorial boards of *Artificial Intelligence*, AAAI Press, Annals of Mathematics and AI, and the Encyclopedia of AI. He has published over 150 research papers, has authored two books: *Heuristics* (Addison-Wesley, 1984) and *Probabilistic Reasoning in Intelligent Systems,* (Morgan Kaufmann, 1988), and has edited *Search and Heuristics* (North-Holland, 1983) and *Reading in Uncertainty Reasoning* (with G. Shafer, Morgan Kaufmann, 1990).

Professor Pearl is a Fellow of IEEE and AAAI, a member of the National Academy of Engineering, and the winner of IJCAI Research Excellence Award for 1999.

H BUDGET

I JUSTIFICATIONS FOR MAJOR ITEMS

The item "Computer Networking Services" provides shared computing resources only available through the department computing facility. Starting January 1, 1989, the Computer Science Department established a departmental recharge unit approved by the POSSSE (Policy Committee on Sales and Service Activities and Service Enterprises) committee of the Chancellor's Office. This recharge is required in **all** contract and grant budgets. The recharge is computed at the rate of 7% of the salaries and benefits.

J CURRENT AND PENDING SUPPORT

CURRENT:

Title:	Probabilistic Networks for Automated Reasoning
Funding Agency:	National Science Foundation
Amount:	$80,719
Contracting Period:	12/1/99 thru 12/1/00

Title:	Advanced Reasoning Methods for Management of Uncertainty
Funding Agency:	MICRO Research Project supported by Rockwell Science Center and the State of California
Amount:	$17,500
Contracting Period:	8/1/99 thru 6/30/00

Title:	Dynamic Networks Techniques for Autonomous Planning and Control
Funding Agency:	U.S. Air Force/Office of Scientific Research
Amount:	$60,336
Contracting Period:	4/1/99 thru 11/30/00

PENDING:

	Proposed by a team of seven Principal Investigators: J. Pearl, A. Darwiche, W. Karplus, E. Coleman, R. Dechter, S. Irani, D. Roth
Title:	An Integrated Approach to Decision Making Under Uncertainty
Funding:	Office of Naval Research
Proposed Amount:	$2,934,000
Proposed Contracting Period:	4/1/00 thru 4/1/03

K LETTER OF INTENT

L LIST OF SUGGESTED REVIEWERS

Title of Proposal: Advanced Methods of Reasoning Under Uncertainty
Name of the P.I.: Judea Pearl
Area of Research: ____Microelectronics
_X_Computer Science
____Applications

Suggested Reviewers:

1. Name: Nils A. Nilsson
   Address: Stanford University
   Dept. of Computer Science & Elec. Eng.
   Margaret Jacks Hall
   Stanford, CA 94305
   Reviewer's fields of specialization: Artificial Intelligence
   Phone: (415) 723-3886 E:mail: nilsson@cs.stanford.edu

2. Name: Bart Selman
   Address: Cornell University, Computer Science Dept.
   4148 Upson Hall
   Reviewer's fields of specialization: Artificial Intelligence
   Phone: (607) 255-4428 E:mail: selman@cs.cornell.edu

3. Name: David Heckerman
   Address: Microsoft Corporation
   One Microsoft Way
   Reviewer's fields of specialization: Artificial Intelligence
   Phone: (425) 936-2662 E:mail: heckerma@microsoft.com

4. Name: Leora Morgenstern
   Address: IBM T.J. Watson Research
   P.O. Box 704, Mail Stop H1L05
   Yorktown Heights, NY 10598
   Reviewer's fields of specialization: Artificial Intelligence
   Phone: (914) 784-7151 E-mail: leora@watson.ibm.com

5. Name: Daphne Koller
   Address: Department of Computer Science
   Stanford University
   Stanford, CA 94305-9010
   Reviewer's fields of specialization: Artificial Intelligence
   Phone: (650) 723-6598 E-mail: koller@cs.stanford.edu