Causal Satisficing And Markoff Models In The Context Of Data Mining

Lawrence J. Mazlack

Applied Computational Intelligence Laboratory
University of Cincinnati
Cincinnati, Ohio 45221-0030
mazlack@uc.edu

Abstract - Causal reasoning occupies a central position in human reasoning. In many ways, causality is granular. This is true for commonsense reasoning as well as for mathematical and scientific theory. Knowledge of at least some causal effects is imprecise. Perhaps, complete knowledge of all possible factors might lead to a crisp description of whether an effect will occur. However, in our commonsense world, it is unlikely that all possible factors can be known. In commonsense, every day reasoning, we use approaches that do not require complete knowledge. We need an algorithmic way of handling imprecision if we are to computationally handle causality. Perhap, fuzzy Markoff models might be useful. People recognize that a complex collection of elements causes a particular effect, even if the precise elements of the complex are unknown. Perhaps Markoff chains might be used to build these complexes. It may be more useful to work on a larger grain size. This may reduce the need to learn extensive hidden Markoff models, which in computationally expensive. Perhaps, a satisficing solution would be to develop large grained solutions and then only go to the finer grain when the impreciseness of the large grain is unsatisfactory.

INTRODUCTION

Commonsense causal reasoning occupies a central position in human reasoning. It plays an essential role in human decision-making. Considerable effort has been spent examining causation. Philosophers, mathematicians, computer scientists, cognitive scientists, psychologists, and others have formally explored questions of causation beginning at least three thousand years ago with the Greeks.

Whether causality can be recognized at all has long been a theoretical speculation of scientists and philosophers. At the same time, in our daily lives, we operate on the commonsense belief that causality exists.

Causal relationships exist in the commonsense world. When a glass is pushed off a table and breaks on the floor, we might say that being pushed from the table caused the glass to break. (Although, being pushed from a table is not a *certain* cause of breakage; sometimes the glass bounces and no break occurs; or, someone catches the glass before it hits the floor.) Counterfactually, usually (but not always), not falling to the floor prevents breakage. (Sometimes, a glass breaks when it is tipped over on the table.) So, knowledge of at least some causal effects is imprecise for both positive and counterfactual descriptions. Perhaps, complete knowledge of all possible factors might lead to a crisp description of whether an effect will occur. However, in our commonsense world, it is unlikely that all possible factors can be known. It is also unlikely that it may be possible to fully know, with certainty, all of the elements involved.

This lack of complete, precise knowledge should not be discouraging. We do things in the world by exploiting our commonsense *perceptions* of cause and effect. When trying to precisely reason about causality, we need complete knowledge of all of the relevant events and circumstances. In commonsense, every day reasoning, we use approaches that do not require complete knowledge. Often, approaches follow what is essentially a *satisficing* [Simon, 1955] paradigm. The use of non-optimal mechanisms does not necessarily result in ad hocism [Goodrich, 2000], for example, Zadeh "[1998] questions the feasibility (and wisdom) of seeking for optimality given limited resources. However, in resisting naive optimizing, Zadeh does not abandon the quest for justifiability, but instead resorts to modifications of conventional logic that are compatible with linguistic and fuzzy understanding of nature and consequences."

Our commonsense understanding of the world tells us that we have to deal with imprecision, uncertainty and imperfect knowledge. This is also the case of our scientific knowledge of the world. We need an algorithmic way of handling imprecision if we are to computationally handle causality. Models are needed to algorithmically consider causes. These models may be symbolic or graphic. A difficulty is striking a good balance between precise formalism and commonsense imprecise reality

In commonsense reasoning, people recognize that a complex collection of elements causes a particular effect, even if the precise elements of the complex are unknown. People may not know what events are in the complex; or, what constraints and laws the complex is subject to. Sometimes, the details underlying an event are known to a fine level of detail, sometimes not. The level of detail can reasonably be called the event's grain size.

In general, people are more successful in reasoning about a few large grain sized events than many fine grained events that might make up a complex that fully describes the large grained event. When using commonsense reasoning, people do not always even need to know the extent of the underling complexity. Complexes may be best handled on a black-box, large grained basis. That is, it can be recognized that a complex exists; but we do not necessarily need to deal with the deals internal to the complex.

COMPLEXES

When events happen, there are usually other related events. The entire collection of events can be called a complex. The events can be called the elements of the complex.

A mechanism [Simon, 1991] or a "causal complex" [Hobbs 2001, 2003] is a collection of events whose occurrence or non-occurrence results in a consequent event happening. A causal complex is the *complete* set of events and conditions necessary for the causal effect (consequent) to occur. Hobbs suggests that using a causal complex does not require precise, complete knowledge of the complex.

Hobbs [2001] uses the example of turning on an electric light. He says: "It is natural to say that when you flip a light switch, you cause the light to go on. But it would not happen if a whole large system of other conditions were not in place. The wiring has to connect the switch to the socket, and be intact. The light bulb has to be in good working order. The switch has to be connected to a system for supplying electricity. The power plant in that system has to be operational. And so on. Flipping the light switch is only the last small move in a large-scale systems of actions and conditions required for the light to go on."

Sometimes, it is enough to know what happens at a large grained level; at other times it is necessary to know the fined grained result. For example, if *Robin believes that flipping the light switch causes the electric light to start*, it is may be enough if Robin engages an electrician when his lights do not go on. On the other hand, the automobile mechanic needs to know a finer grained view of the causal complex than does Robin.

Each complex can be considered to be a granule. Larger complexes can be decomposed into smaller complexes. Thus, going from large grained to small grained. This paper will consider the issue of necessary complex granularity later.

Hobbs uses first order logic to describe his causal complexes. Similarly, Pearl [2000] develops probabilistic causal networks of directed graphs (DAGs).

The causal complexes explicitly considered by Hobbs and Pearl have a required structure that may be overly restrictive for commonsense causal understanding, namely:

- If *all* of the events in the causal complex happen, then the effect will occur
- There is nothing in the causal complex that is irrelevant to the effect

These requirements are probably too precise and extensive to be realized in a commonsense world. Sometimes, only some of the events need to happen. So, if all of the events happen, the effect may happen. If some of the events happen, the effect may happen. In the commonsense world, we rarely whether all of the events are in a complex are necessary. For example, a woman may want to attract the attention of a man. She may do a large number of things. If she does attract the man, she may never know which things were relevant and which were not. A better approach is to incorporate granulation using rough sets and/or fuzzy sets to soften the need for preciseness. And then accept impreciseness in the description.

Recognizing Causality

Recognizing causality has is of interest in many areas. Of particular interest to this paper are areas where the analysis is non-experimental.

Perhaps, the applied area that has the greatest history of attempting to deal with causality and non-observational data is economics. Econometrics is distinguished from statistics by econometrics interest in establishing causation [Hoover, 2003].

How and if causality can be recognized has been a significant area of discussion. Some of this discussion mirrors discussion that has gone on in the computational sciences. Hoover [2003] provides a good entry to the discussion of causality in economics.

Granger [1969] defined causality depends on one-way, time ordered conception of causality. In contrast, Simon [1952, 1953] provides an analysis of causality that does not rely on time order. This issue is of importance to those modeling causality in data discovery.

Friedman [1949] argues that any cause that we isolate is never the whole cause and that every direct cause itself has its own direct causes, sot that networks of causation spread synchronically across the economy and diachronically back into the mists of time.

Hume [1777/1902, p 165], as a philosopher, suggested that causal statements are really about constant conjunction and time-ordering. Hume [1742/1985, p 304] as an economist, "it is of consequence to know the principle whence any phenomenon arises, and to distinguish between a cause and a concomitant effect."

Data mining analyzes data previously collected; thus it is non-experimental.

There are several different data mining products. The most common are *conditional rules* or *association rules*. Conditional rules are most often drawn from induced trees while association rules are most often learned from tabular data.

- ***Conditional rule:***
  IF Age < 20
  THEN Income < $10,000
  with {belief = 0.8}
- ***Association rule:***
  Customers who
  buy beer and sausage
  also tend to buy hamburger
  with {confidence = 0.7}
  in {support = 0.2}

At first glance, these association rules seem to imply a causal or cause-effect relationship. That is: *A customer's purchase of both sausage and beer causes the customer to also buy hamburger.* All that is discovered is the *existence* of a statistical relationship between the items. They have a degree of joint occurrence. The *nature* of the relationship is not specified. We do not know whether the presence of an item or sets of items causes the presence of another item or set of items; or the converse, or some other phenomenon causes them to occur together.

Sometimes, the relationship might be causal; for example, if *someone drinks beer* and then *becomes inebriated,* there may be a causal relationship. On the other hand, if *someone wears a 'lucky' shirt* and then *wins a lottery*, there may not be a causal relationship. Recognizing true causal results greatly enhances the decision value of data mining results.

Purely accidental relationships are not nearly of the same level of interest as causal relationships. For example, if it is true that buying both *beer and sausage* somehow causes someone to *buy beer*, then a merchant might profitably put *beer* (or *sausage*) on sale and then increase the price of *hamburger* to compensate for the sale price. On the other hand, knowing that *bread* and *milk* are often purchased together may not be useful information as both products are commonly purchased on every store visit. What might be of interest is discovering if there is a causal relationship between the purchase of *bananas* and something else. (It turns out that *bananas* are the most frequently purchased food item at Wal-Mart [Nelson, 1998]).

One of the reasons why association rules are used is to aid in making retail decisions. However, simple association rules may lead to errors. Errors might occur; either if causality is recognized where there is no causality; or if the direction of the causal relationship is wrong [Silverstein, 1998*a*] [Mazlack, 2003*b*].

Causality Recognition

Various causality tools have been suggested. Different subject domains appear to have different preferences.

Different aspects of causality have been examined. The idea of "positive" causation ($\alpha \rightarrow \beta$) is at the core of common sense causal reasoning. Often a positive causal relationship is represented as a network of nodes and branches [Mazlack, 2003*a*].

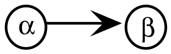

Figure 1. Diagram indicating that α is causally dependent on β.

Negation or counter factuals ($\neg\alpha \rightarrow \neg\beta$) also have a place; although it may result in errors in reasoning. For example, the rule:

If a person drinks wine, they may become inebriated

cannot be simply negated to

If a person <u>does not</u> drink beer, they will <u>not</u> become inebriated.

Effects can be *over determined;* that is: more than one item can cause an effect. In this case, people may also drink *beer* to excess and also become inebriated.

Other ideas that are sometimes involved in causal reasoning are *causal uncorrelatedness* [Shafer, 1999] where if two variables have no common cause they are causally uncorrelated. This occurs if there are no single events that cause them to both change. Similarly, causal independence occurs when speaking about probabilities.

Similarly, Dawid [1999] focuses on the negative; i.e., when α does not affect β. Dawid speaks in terms of *unresponsiveness* and *insensitivity*. β is unresponsive to α if whatever the value of α might be set to, β will be unchanged. β is insensitive to α if whatever the value α may be set, the uncertainty about β will be unaffected. Along the same vein, Shoham [1990, 1991] distinguishes between *causing, enabling,* and *preventing.* The enabling factor is often considered to be a causal factor. Shoham distinguished between background (enabling) conditions and foreground conditions. The background (enabling) conditions are inferred by default. For example [Shoham, 1991]: "If information is present that the key was turned and nothing is mentioned about the stated about the state of the battery, then it is inferred that the motor will start, because the battery is assumed, by default to be alive. Given this distinction, causing is taken to refer to the foreground conditions where enabling and preventing refer to the background conditions (in this example, turning the key causes the motor to start, the live battery enables it, the dead battery prevents it)."

Standard Statistics

The standard method in the experimental sciences of recognizing causality is to perform randomized, controlled experiments (observational data). Depending on their design, randomized experiments may remove reasons for uncertainty whether or not a relationship is casual. However, large data sets are typically the subject of data mining. Even if some experimentation is possible, the amount of experimentation in contrast to the amount of data to be mined will be small. This said, some work has been done using chi-squared testing to reduce the search space [Silverstein, 1998].

Directed Graphs

Various graph based Bayesian based methods have been suggested to recognize causality. Probably the best known is the class of methods based on Directed Acyclic Graphs (DAGs). The most fully developed approach is Pearl [2000]. Silverstein [1998] followed a similar approach.

Pearl [1991] and Spirtes [1993] make the claim that it is possible to infer causal relationships between two variables from associations found in observational (nonexperimental) data without substantial domain knowledge. Spirtes claim that directed acyclic graphs can be used if (a) the sample size is large and (b) the distribution of random values is faithful to the causal graph. Robins [1999] argues that their argument is incorrect. Lastly, Scheines [1994] only claims that in some situations will it be possible to determine causality. Their discussion is tangential to the focus of this paper; going deeply into their discussion is outside of the scope of this paper. It is enough to note that these methods are possibly the most thoroughly developed methods of computational causal analysis.

The Bayesian directed graph methods all have similar liabilities, specifically:

- *Discrete or continuous data must be reduced to Boolean values.*

  Objection: This is an early technique that was and is used in data mining when analyzing market basket data. However, it is essentially flawed. Quantities do matter; some data cooccurrences are conditioned on there being a sufficiency of a co-occurring attribute. Also, some relationships may be non-linear based on quantity [Mazlack, 2003*b*].

- *There is no missing data.*

  Objection: This is at variance with day-to-day experience. There is almost always missing data of some sort. Data collection is rarely fully representative and complete. Incremental data is often acquired that is at variance with previously acquired data. What is needed is a methodology that is not brittle in the face of incompleteness.

- *Causal relationships are not cyclic, either directly or indirectly (through another attribute).*

  Objection: This is at variance with our commonsense understanding of the world.

  It is possible that there might be mutual dependencies [Mazlack, 2003]; i.e., $\alpha \rightarrow \beta$ as well as $\beta \rightarrow a$. It seems to be possible that they do so with different strengths. They can be described as shown in the following figure where $S_{i,j}$ represents the strength of the causal relationship from i to j. It would seem that the strengths would be best represented by an approximate belief function, either quantitatively or verbally. There would appear to be two variations:

  ◊ *Different causal strengths for the same activity, occurring at the same time:*

  For example, α could be *short men* and β could be *tall women*. If $S_{\alpha,\beta}$ meant the strength of desire for a social meeting that was caused in *short men* by the sight of *tall women*, it might be that $S_{\alpha,\beta} > S_{\beta,\alpha}$.

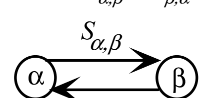

Figure 2. Cyclic relationship: Mutual dependency. [Mazlack, 2003*a*]

Additionally, Simon [1991] and Shoham [1991] identify cases where causality is simultaneous.

Some argue that causality should be completely asymmetric and if it appears that items have mutual influences it is because there is another cause that causes both. A problem with this is that it can lead to eventual regression to a first cause; whether this is true or not, it is not useful for commonsense representation.

◊ *Different causal strengths for symmetric activities, occurring at different times.*

It would seem that if there were causal relationships in market basket data, there would often be imbalanced dependencies. For example, if a customer first buys strawberries, there may be a reasonably good chance that she will then buy whipped cream. Conversely, if she first buys whipped cream, the subsequent purchase of strawberries may be less likely. This situation could also be represented by the previous figure. However, the issue of time sequence would be poorly represented. A graph representation could be used that implies a time relationship. Nodes in a sequence closer to a root could be considered to be earlier in time than those more distant from the root. Redundant nodes would have to be inserted to capture every alternate sequence. For example, one set of nodes for when strawberries are bought before whipped cream and another set when whipped cream is bought before strawberries. However, this representation is less elegant and not satisfactory when a time differential is not a necessary part of causality. It also introduces multiple nodes for the same object (e.g., strawberries, whipped cream); which at a minimum introduces housekeeping difficulties.

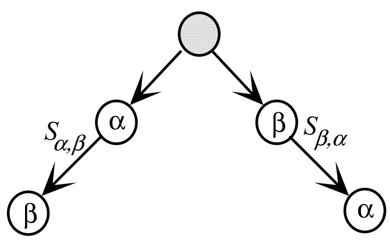

Figure 3. Alternative time sequences for two symmetric causal event sequences where representing differing event times necessary for representing causality. Nodes closer to the root occur before nodes more distant from the root. Causal strengths may be different depending on sequence. [Mazlack, 2003*a*]

◊ *Markov Stationary Condition* holds: Probabilities are time independent.

Objection: This does not correspond to our commonsense understanding of the world. If one event is dependent on two other causal events, if one causal event happens much earlier (or later) than the other causal event, we may well have a different result.

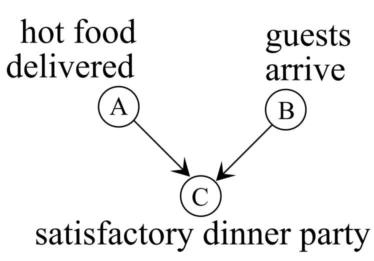

Figure 4. Case where differing times in causal events affects probability of causal result.

◊ The *Markoff Condition* holds: Let *A* be a node in a causal Bayesian network, and let *B* be any node that is not a descendant of *A* in the network. Then the Markoff (Markov) condition holds if *A* and *B* are independent, conditioned on the parents of *A*. The intuition of this condition is: If *A* and *B* are dependent, then *B* must either be (a possibly indirect) cause of *A* or (possibly indirectly) caused by *A*. In the second case, *B* is a descendant of *A*, while in the first *B* is an ancestor of *A* and has no effect on *A* once *A*'s immediate parents are fixed.

This makes sense in the following example.

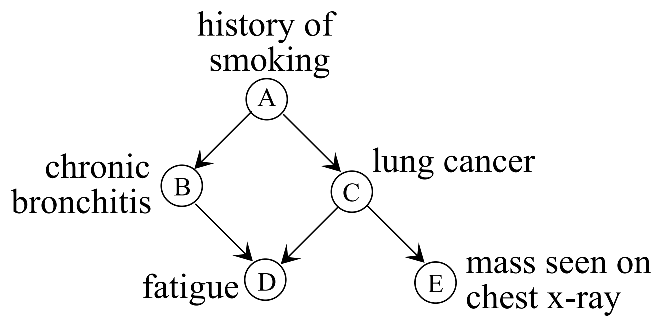

Figure 5. "Memoryless" Markoff condition holds

However, not all of our commonsense perceptions of causality work this way. Often, we believe that history matters as in the following example.

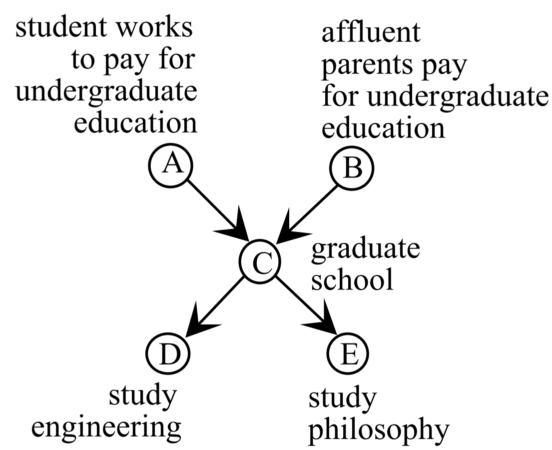

Figure 6. Causality where memory play a part.

Markoff Models

Markoff (Markov) Models treat a system as a series of states with specific, constant rate transitions between them. At all times, the system is in exactly one state. (Transitions are considered to be instantaneous.) The only information available is the current state, the allowed transitions, and the probability of these transitions. Such a system is referred to as memoryless, as is said to posses the *Markoff property*. This means that the system is totally characterized by its current state. None of the past states or transitions have any effect on the transitions out of the current state.

Quantitatively describing the relationships between the nodes can be complex. One possibility is an extension of the random Markoff model, shown in Figure 7. The state value is 1/0 as an event either happens or does not.

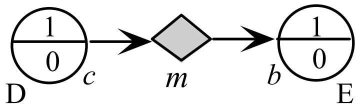

Figure 7. Random Markoff model: c = P(D), m = the possibility/probability that when D is present, the causal mechanism brings about E, b = the probability that some other (unspecified) causal mechanism brings about E.

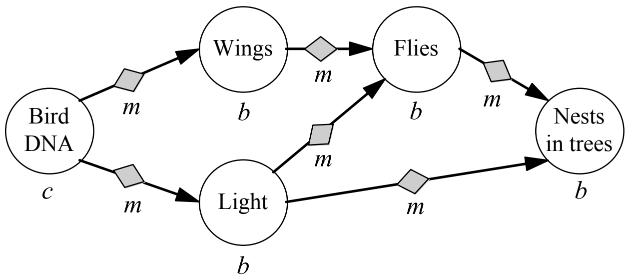

Figure 8. Markoff model applied to a "bird" example [Rehder, 2002]

Perhaps, fuzzy sets and fuzzy Markoff algorithms in particular might be a good way of approaching causality represented by Markoff models. What we must work with is inherently imprecise and incomplete.

Fuzzy sets are useful for dealing with any situation where the exact value of a variable is unknown. Instead of a guess of the value of the variable (which can easily be wrong), or a distribution of its possible values (which is usually unknown, so this problem reduces to a guess), fuzzy logic deals with the *possibility* of the variable taking on a set of values. In this way, it assumes less, and shows explicitly both what is and is not known.

Cheok [2002] shows how the probabilities in the classic Markoff model may be extended to fuzzy possibilities. The generalized hidden Markoff Model can be defined with the same type of model parameters, but with a different mathematical basis than the classic one. Fuzzy logic replaces probability theory and this leads to a new definition of the model variables. The structure in terms of states and observations, remains the same.

There has been a fair amount of work in fuzzy Markoff models [Jeanpierre, 2002] [Leuschen, 1998] [Zadeh, 1998]. A fair amount of the work has a control theory or medical diagnostic orientation. Whether it can be extended to representing causality is to be seen.

Increasing Grain Size

Depending on the goal of the work, it may be more useful to work on a larger grain size. This may reduce the need to learn extensive hidden Markoff models, which in computationally expensive. Perhaps, a satisficing solution would be to develop large grained solutions and then only go to the finer grain when the impreciseness of the large grain is unsatisfactory. For example, if instead of a complex in the form of Figure 8, perhaps a diagram similar to Figure 9, would be sufficient. This form of representation is widely used in automata design, its extension to causal reasoning and imprecise reasoning would seem to be reasonable.

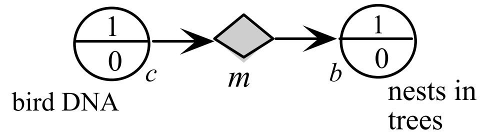

Figure 9. Greater grain size (higher level) abstraction of Figure 8.

The greater grain size would be more useful in commonsense reasoning. It also might be more computationally tractable in the discouvery process.

Epilogue

Commonsense causal reasoning occupies a central position in human reasoning. It plays an essential role in human decision-making. Considerable effort has been spent examining causation. Philosophers, mathematicians, computer scientists, cognitive scientists, psychologists, and others have formally explored questions of causation beginning at least three thousand years ago with the Greeks.

Whether causality can be recognized at all has long been a theoretical speculation of scientists and philosophers. At the same time, in our daily lives, we operate on the commonsense belief that causality exists.

Causal relationships exist in the commonsense world. Knowledge of at least some causal effects is imprecise for both positive and counterfactual descriptions. Perhaps, complete knowledge of all possible factors might lead to a crisp description of whether an effect will occur. However, in our commonsense world, it is unlikely that all possible factors can be known. It is also unlikely that it may be possible to fully know, with certainty, all of the elements involved.

This lack of complete, precise knowledge should not be discouraging. We do things in the world by exploiting our commonsense *perceptions* of cause and effect. When trying to precisely reason about causality, we need complete knowledge of all of the relevant events and circumstances. In commonsense, every day reasoning, we use approaches that do not require complete knowledge. Often, approaches follow what is essentially a *satisficing* paradigm.

Our commonsense understanding of the world tells us that we have to deal with imprecision, uncertainty and imperfect knowledge. This is also the case of our scientific knowledge of the world. We need an algorithmic way of handling imprecision if we are to computationally handle causality. Models are needed to algorithmically consider causes.

In commonsense reasoning, people recognize that a complex collection of elements causes a particular effect, even if the precise elements of the complex are unknown. People may not know what events are in the complex; or, what constraints and laws the complex is subject to. Sometimes, the details underlying an event are known to a fine level of detail; sometimes not. The level of detail can reasonably be called the event's grain size. In general, people are more successful in reasoning about a few large grain sized events than many fine grained events. Perhaps, this can transfer over to computational models of causality.

Perhaps Markoff chains might be used to build these complexes. Depending on the goal of the work, it may be more useful to work on a larger grain size. This may reduce the need to learn extensive hidden Markoff models, which in computationally expensive. Perhaps, a satisficing solution would be to develop large grained solutions and then only go to the finer grain when the impreciseness of the large grain is unsatisfactory.

REFERENCES

A. Cheok, S. Chevalier, M. Kaynak, K. Sengupta, K. Chung [2002] "Use of a Novel Generalized Fuzzy Hidden Markov Model for Speech Recognition," *Proceedings. IEEE International Conference on Fuzzy Systems,* December 2-5, 2002 Melbourne, Australia

A. Dawid [1999] "Who Needs Counterfactuals" in **Causal Models and Intelligent Data Management** (ed) A. Gammerman) Springer-Verlag, Berlin

M. Friedman [1949] "The Marshallian Demand Curve," *Journal of Political Economy,* v 57, 463-495

M. Goodrich, W. Stirling, E. Boer [2000] "Satisficing Revisited," *Minds and Machines,* v 10, 79-109

C. Granger [1969] "Investigating Causal Relations By Econometric Models And Cross-Spectral Methods," *Econometrica,* v 37, 424-438

J. Hobbs [2001] "Causality," *Proceedings, Common Sense 2001, Fifth Symposium on Logical Formalizations of Commonsense Reasoning*, New York University, New York, May, 145-155

J. Hobbs [2003] "Causality And Modality: The Case Of 'Would'," to appear in *Journal of Semantics*

K. Hoover [2003] "Lost Causes," *HES Conference,* Presidential Address, Durham, North Carolina

D. Hume [1742/1985] **Essays: Moral, Political, And Literary**, Eugene Miller (ed.), Liberty Classics, Indianapolis, 1985

D. Hume [1777/1902] "An Enquiry Concerning Human Understanding," in L. Selby-Bigge (ed.) **Enquiries Concerning Human Understanding And Concerning The Principles Of Morals,** 2nd edition, Clarendon Press, Oxford, 1902

L. Jeanpierre, F Charpillet [2002] "Hidden Markov Models for Medical Diagnosis," *International Workshop on Enterprise Networking and Computing in Health Care Industry, HEALTHCOM'2002,* Nancy, France, June 2002

M. Leuschen, I. Walker, J. Cavallaro [1998] "Robot Reliability Through Fuzzy Markov Models," *Proceedings IEEE Annual Reliability and Maintainability Symposium,* Anaheim, CA, January 1998, 209-214,

L. Mazlack [2003*a*] "Commonsense Causal Modeling In The Data Mining Context," IEEE ICDM Proceedings, Melbourne, Florida, November 19 - 22, 2003

L. Mazlack [2003*b*] "Causality Recognition For Data Mining In An Inherently Ill Defined World," 2003 BISC FLINT-CIBI International Joint Workshop On Soft Computing For Internet And Bioinformatics, December, 2003

E. Nelson [1998, October 6] "Why WalMart sings, 'Yes, we have bananas'," The Wall Street Journal, B1

J. Pearl, T. Verma [1991] ""A Theory Of Inferred Causation," *Principles Of Knowledge Representation And Reasoning: Proceedings Of The Second International Conference,* Morgan Kaufmann, 441-452

J. Pearl [2000] **Causality**, Cambridge University Press, New York, NY

B. Rehder [2002] "A Causal-Model Theory of the Conceptual Representation of Categories," FLAIRS 2002 Workshop on Causality, Pensacola, May

R. Robins, L. Wasserman [1999], "On The Impossibility Of Inferring Causation From Association Without Background Knowledge," in (eds) C. Glymour, G. Cooper, **Computation, Causation, and Discovery** AAAI Press/MIT Press, Menlo Park, 305-321

R. Scheines, P. Spirtes, C. Glymour, C. Meek [1994] **Tetrad II: Tools For Causal Modeling**, Lawrence Erlbaum, Hillsdale, NJ

G. Shafer [1999] "Causal Conjecture," in **Causal Models and Intelligent Data Management** (ed) A. Gammerman) Springer-Verlag, Berlin

Y. Shoham [1990] "Nonmonotonic Reasoning And Causation," *Cognitive Science,* v14, 213-252

Y. Shoham [1991] "Remarks On Simon's Comments," *Cognitive Science*, v15, 301-303

C. Silverstein, S. Brin, R. Motwani, J. Ullman [1998] "Scalable techniques for mining causal structures," *Proceedings 1998 VLDB Conference,* New York, NY, August 1998, 594-605

H. Simon [1952] "On The Definition Of The Causal Relation," *The Journal Of Philosophy,* v 49, 517-528. Reprinted in Herbert A. Simon, **Models Of Man**, John Wiley, New York, 1957

H. Simon [1953] "Causal ordering And Identifiability," Reprinted in Herbert A. Simon, **Models Of Man**, John Wiley, New York, 1957

H. Simon [1991] "Nonmonotonic Reasoning And Causation: Comment," *Cognitive Science,* v 15, 293-300

L. Smolin [2004] "Atoms of Space And Time," *Scientific American,* January, 2004, 66-75

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

L. Zadeh [1998] "Maximizing Sets And Fuzzy Markoff Algorithms," *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews,* v 28, 9-15