Discovering Hidden Dispositions and Situational Factors in Causal Relations by Means of Contextual Independencies

Manon J. Sanscartier (mjs696@mail.usask.ca) and Eric Neufeld (eric@cs.usask.ca)
Department of Computer Science, University of Saskatchewan; 57 Campus Drive
Saskatoon, Saskatchewan, Canada S7K 5A9

Abstract

Correspondent inferences in attribution theory deal with assigning causes to behaviour based on true dispositions rather than situational factors. In this paper, we investigate how knowledge representation tools in Artificial Intelligence (AI), such as Bayesian networks (BNs), can help represent such situations and distinguish between the types of clues used in assessing the behaviour (dispositional or situational). We also demonstrate how a discovery algorithm for contextual independencies can provide the information needed to separate a seemingly erroneous causal model (considering dispositions and situations together) into two more accurate models, one for dispositions and one for situations.

Introduction

In the determination of causal attribution, we are interested in not only the true cause of behaviour, but also in how we, as human adults, assign a cause to another person's behaviour, whether the inferred cause is true or not. When seeking to understand another individual's behaviour, people generally make use of information that can be classified into two categories of causes, namely *situational* factors, and *dispositional* causes. Situational factors explain actions in terms of a social setting or environment, while dispositions are causes based on characteristics of the person whose behaviour we seek to understand. When attributing a cause to a person, it is very important that the inference comes from dispositional factors, and not situational ones. Unfortunately, the distinction between the two is often blurred in data. For example, a job applicant who fails to attend a recruitment meeting may be perceived as anti-social or uninterested (disposition), when in reality, the individual lives out of province, and will only relocate if hired (situation). A more thorough examination of the context of the situation painted by the available information may reveal hidden clues about the nature of the factors being considered (situational or dispositional). Discovery of such clues (context-specific independencies) may yield more accurate causal models to describe the situation at hand.

In the late 1990s, psychologists and educators began to value the need for consideration of the natural context in which humans perform problem solving tasks. This relatively new emphasis is refered to as *situated cognition* (Seifert, 1999). In this paper, we address how consideration of context can help uncover hidden factors about individuals and how the discovered independencies will improve/change our believed causal model, by isolating situational factors and true dispositions, to distinguish between the causal repercussions in both cases. For the remainder of the paper, the terms factor and variable will be used interchangeably. There are two kinds of hidden variables, we will call them *unmeasured-out* and *unmeasured-in*. The first type is present when the relevant information is simply not in the model. Alternately, information could be hidden inside a variable, typically by means of an independency that holds only in a particular context. We will call this scenario *unmeasured-in*. This context-specific independence (CSI) has been mainly studied in the context of reasoning with uncertainty, and methods of inference for such independencies exist as well (Boutilier et al., 1996). The distinction between the two classes of hidden variables is necessary for making correct representational decisions in adult causal judgment when faced with seemingly erroneous data.

We then focus on the class of unmeasured-in variables and show how statistical methods for discovering contextual independencies in Bayesian networks can help us discover these hidden variables. Ignoring this kind of hidden variable results in incorrect inferences about particular subgroups of individuals. More interestingly, we show that if the contextually hidden data is considered, it will help us learn whether the attribution was based on a person's true disposition or on situational factors. In addition, we may discover that two different causal models should be used for the same scenario based on the type of attribution that was made (dispositional or situational). We present a method for correcting such erroneous models by finding the hidden contextual variables.

The remainder of this paper is organized as follows. First, we present some background information. We discuss Correspondent Inferences in attribution theory, and follow with the role of Bayesian networks and causal models as representation and inference tools. Then we give an example of a distribution containing hidden factors that cannot be inferred without consideration of context. We then discuss two types of hidden variables and how we distinguish between the two. Finally, we discuss a method for discovering hidden variables from data, by means of contextual independencies. In our conclusions, we outline some potential future direction for this work.

Background Information

In the introduction, we distinguished between true dispositions and situational factors. In this section, we highlight particularities about each that motivate the need for an understanding of subsets of the data used in making causal judgments. We then discuss *Bayesian networks* and how they are a useful representational tool for causal relations. Finally, we discuss causal models and present an example.

Correspondent Inferences in Attribution

As stated in the introduction, situational factors explain actions in terms of a social setting or environment, while dispositions are causes based on characteristics of the person in question. Jones and Davis' Correspondent Inference theory (Jones and Davis, 1965) suggests that we use information about the behaviour of a person as well as effects of the particular behaviour to make a correspondent inference, in which the behaviour is either attributed to a disposition or a situation, and is based on a sole observation. This theory is interesting for hidden variable discoveries, as we have a single observation about each individual, and discover independencies between variables when we look at a group of individuals performing a similar task.

Bayesian Networks and Causality

A *Bayesian network* (BN) (Pearl, 1988) is a directed acyclic graph with a conditional probability distribution associated with each node. The topology of the graph encodes the information that the joint distribution of all variables in the graph is equal to the product of the local distributions. We can interpret the joint distribution as being everything we know about a group of individuals, and the local distributions as being a subset of information directly related to a particular inquiry about the group of users. BNs compactly represent joint probability distributions, and reason efficiently with those representations. There is significant literature on inference; (Pearl, 1988) is a good place to start.

BN practitioners noticed early on that typical independence assumptions (unconditional independence of diseases, conditional independence of symptoms) in the diagnosis domain, for example, tended to orient arcs in the direction of causality. Pearl and Verma (Pearl and Verma, 1991) provided probabilistic definitions of causality that explained this phenomenon, but also provided algorithms for learning cause-effect relationships from raw data.

The definitions of Pearl and Verma are subtle, but the algorithm itself is simple, and works as follows. Although covariance does not imply causation, covariance implies the *presence* of causality. If A and B covary, then either A causes B, B causes A, or A and B have a common cause C. Thus, covariance implies a disjunction of causal relations. Combinations of conditional independencies and dependencies in the data can eliminate certain disjuncts. For example, if A and C covary, B and C covary, and A and B are unconditionally independent, one expects causal links to be directed from A and B to C, as all other causal scenarios lead to logical errors.

Causal Models

Several authors express causal models in probabilistic terms because, as argued by Suppes (Suppes, 1970), most causal statements in everyday conversation are a reflection of probabilistic and not categorical relations. For that reason, probability theory provides an adequate framework for reasoning with causal knowledge (Good, 1983; Reichenbach, 1956). Pearl's *causal models* provide the mechanism and structure needed to allow for a representation of causal knowledge based on the presence and absence of *probabilistic conditional independencies* (CIs) (Pearl, 1988).

Definition 1 *A causal model (Pearl and Verma, 1991) of a set of random variables R can be represented by a directed acyclic graph (DAG), where each node corresponds to an element in R and edges denote direct causal relationships between pairs of elements of R.*

The direct causal relations in the causal model can be expressed in terms of CIs.

Definition 2 *Let $R = \{A_1, A_2, \ldots, A_n\}$ denote a finite set of discrete variables, where each variable $A \in R$ takes on values from a finite domain V_A. We use capital letters, such as A, B, C, for variable names and lowercase letters a, b, c to denote outcomes of those variables.*

Let X and Y be two disjoint subsets of variables in R and let $Z = R - \{X \cup Y\}$. We say that Y and Z are conditionally independent given X, denoted $I(Y, X, Z)$ if, given any $x \in V_x$, $y \in V_y$, then for all $z \in V_z$

$$p(y|x, z) \;=\; p(y|x), \text{ whenever } p(x, z) > 0.$$

With the causal model alone, we can express portions of the causal knowledge based on the CIs in the model. The conditional probabilities resulting from the CIs defined in the model can be formally expressed for all configurations in the Cartesian product of the domains of the variables for which we are storing conditional probabilities.

Definition 3 *Let X and Y be two subsets of variables in R such that $p(y) > 0$. We define the conditional probability distribution (CPD) of X given $Y = y$ as:*

$$p(x|y) \;=\; \frac{p(x, y)}{p(y)}, \text{ implying } p(x, y) \;=\; p(y) \cdot p(x|y) \quad (1)$$

for all configurations in $V_x \times V_y$.

Definition 4 *A causal theory is a pair $T =< D, \theta_D >$ consisting of a DAG D along with a set of CPDs θ_D consistent with D. To each variable $A \in R$, there is an attached CPD $p(A_i|Y_i \ldots Y_n)$ describing the state of a variable A_i given the state of its parents $Y_i \ldots Y_n$.*

Example of a Causal Model

Company ABC is interested in better understanding what type of applicant is likely to be a successful employee within the company. ABC is a large corporation and receives applications from across the country. The CEO likes to interview as many qualified applicants as possible. However, although a large percentage of applicants meet all the requirements, to reduce recruitment cost the CEO would like to interview only a subset of the qualified applicants. The CEO would like to learn more about the employees of his company to understand what type of applicant would likely be successful in interview.

The causal model in Figure 1 describes the causal relationship between 5 variables directly related to the potential success in interview of a typical applicant, including the success variable itself. For simplicity, we assume each variable is binary. The 5 variables are the following: *(A)pplicant's experience with dealing with the public*, *(W)eekend outings organized by company regularly to promote dynamics within personnel*, *(P)reparation for interview*, *(R)esearch about company done by applicant prior to interview*, and finally *(S)uccess in job interview.*

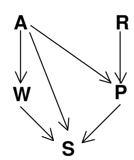

Figure 1: Causal model for job interview.

According to the DAG, there is a direct causal relationship between the applicant's experience with the public (A) and their interest in making their involvement in the company a part of their social life (W). There is also a direct causal influence from A to P, the time and effort spent on job interview preparation. Finally, the last causal relationship emerging from A is clear, namely that there is a relationship between A and a successful job interview (S). Researching the company prior to the job interview (R) is causally related to preparation for the interview (P), which in turn is directly causally related to S, a successful interview. Finally, an interest in socializing outside work hours (W) is directly related to a successful job interview. The corresponding causal theory attaches to variables A, R, P, W, and S respectively the following CPDs:

$$p(A),\ p(R),\ p(P|A,R),\ p(W|A) \text{ and } p(S|A,W,P). \quad (2)$$

Although the causal model in Figure 1 seems reasonable and intuitive, we will see later that discovery of hidden variables paints a different picture that can lead to bad hiring decisions if left unattended. Although the notion of causation is frequently associated with concepts of necessity and functional dependence, "causal expressions often tolerate exceptions, primarily due to missing variables and coarse descriptions" (Pearl and Verma, 1991).

Classes of Hidden Variables

An important distinction needs to be made between types of hidden variables to allow for accurate consideration of context and correction of causal models. We call a variable *unmeasured-out* when the relevant information is simply not in the model. This type of omission can yield a model erroneous in that the data may indicate a direct causal relationship between two variables, when in reality the two variables are simply the effects of a common cause. This type of false causal conclusion is referred to as "spurious association", and Pearl and Verma's (Pearl and Verma, 1991) Inductive Causality (IC) algorithm can detect the presence of such associations, although the algorithm cannot rectify the problem. The engine can't provide the user with the factor or set of factors that is a common cause to the spurious association: "No causes in, no causes out" (Cartwright, 1989).

The other type of hidden variable is *unmeasured-in* and it is a genuine causal relationship that is hidden inside a variable, typically by means of an independency that holds in a particular context. The causal relationships known about a particular domain are probabilistic conditional independencies (CIs) found in the data. For a CI to hold, it must be true for every configuration in the dataset. Whenever we find such truth in the data, we ensure there is no direct causal relationship between the variable on which we condition, and the variable that is deemed independent. In such cases, we can remove the causal link from the causal model. Since this CI must hold for every value in a CPD to be considered independent, any subset of values for which an independence holds simply gets ignored. The distinction between dispositional and situational factors can easily become blurred in a model that only admits CI and result in the entire causal model seeming erroneous. Note that discovering unmeasured-in variables can provide clues on where unmeasured-out variables may need to be considered.

We show how statistical methods for discovering contextual independencies in Bayesian networks can help us discover these hidden variables. More interestingly, we show that if the contextually hidden data were considered, it would help us learn much about a particular type of individual, based on the reason for their behaviour, namely a situation or a true disposition. We present a method for correcting such erroneous models by finding the hidden contextual variables.

Discovery of Hidden Variables

Since BNs operate on the general notion of CIs, it is difficult to consider hidden variables in the data or even to be aware of their presence. In this section, we first instantiate a CPD from our running example, which is based solely on CI. We then introduce context-specific independence (CSI) and discuss how it allows us to consider contexts and therefore have a starting point for finding hidden variables. We illustrate this with our running example. Finally, we show how the discovery of CSIs helps refine and correct our existing causal model.

Instantiation of a CPD

In his attempts to understand applicants and their potential fit within the company, while not interviewing all qualified applicants, the CEO of ABC gathers factors about the applicants that he feels are relevant indicators of success. For every hiring session, he organizes an informal social recruiting session specifically for the applicants, and although not mandatory, he expects most candidates to attend. Since this session is an indicator of motivation and interest, the CEO compiles the applications of those who didn't attend the session to look for indicators of a lower applicant success rate, which is exactly what one would expect. Based on the arrows in the causal model in Figure 1, the variables having a direct relationship with successful interview S are A, P, and W. The associated CPD for $p(S|A, W, P)$ is presented in Figure 2.

A	W	P	S	$p(S\|A, W, P)$
0	0	0	0	0.80
0	0	0	1	0.20
0	0	1	0	0.10
0	0	1	1	0.90
0	1	0	0	0.80
0	1	0	1	0.20
0	1	1	0	0.10
0	1	1	1	0.90
1	0	0	0	0.15
1	0	0	1	0.85
1	0	1	0	0.15
1	0	1	1	0.85
1	1	0	0	0.05
1	1	0	1	0.95
1	1	1	0	0.05
1	1	1	1	0.95

Figure 2: The CPD $p(S|A, W, P)$.

Based on the information in the distribution, we see that some applicants who did not attend the session were very successful in interview while others were not. There is no clear indication that not attending the recruitment session had a direct impact on overall success. If that were the case, all probability values in the distribution would be quite low since none, or few of the applicants from this group would have had successful interviews. Below, we see how a discovery method for hidden variables reveals strong influences hidden in this seemingly inconclusive CPD, and revealing situational factors about the individuals that are not to be attributed to true dispositions about the person, but rather to the situation.

Context-Specific Independence (CSI)

Boutilier et al. (Boutilier et al., 1996) formalized the notion of context-specific independence. Without CSI, it is only possible to establish a causal relationship between two variables if a certain set of CIs is absent for all values of a variable in the distribution. With CSI, we can recognize CIs that hold for a subset of values of a variable in a distribution. Therefore, we can account for a situation of the individual that *doesn't* reflect a true disposition of that person, without disregarding occasions when a similar inference would be rightfully attributed to a true disposition. CSI is a CI that holds only in a particular context. Discovery of CSI can help us build more specific causal models instead of a single causal model ignoring particular subsets of values. CSI is defined as follows.

Definition 5 *Let X, Y, Z, C be pairwise disjoint subsets of variables in R, and let $c \in V_c$. We say that Y and Z are conditionally independent given X in context $C = c$ (Boutilier et al., 1996), denoted $I_{C=c}(Y, X, Z)$ if,*

$$p(y|x, z, c) = p(y|x, c), \text{ whenever } p(x, z, c) > 0.$$

Note that since we are dealing with partial CPDs, a more general operator than the multiplication operator is necessary for manipulating CPDs containing CSIs. This operator, formalized by Zhang and Poole (Zhang and Poole, 1999) is called the *union-product* operator and we represent it with the symbol $\odot$. Due to space limitations, we do not discuss the details of union-product here.

CSI Discovery

The CEO of ABC did not consider context. In this subsection, we see that a consideration of context changes the original model in Figure 1. We use a CSI detection method called Refine-CPD-tree (Butz and Sanscartier, 2002a). The method is based on a tree representation of a CPD. Using this algorithm, we can see if a tree reduction is possible. If such a reduced tree exists, the data contains a CSI, which is an indication of a hidden variable that could perhaps correct a faulty model that may otherwise appear correct. The detection method works as follows: Given a tree representation of a CPD, if all children of a node A are identical, then replace A by one of its offspring, and delete all other children of A.

In our running example, we have a CPD that contains all available information relevant to making a decision about the potential success of an interview by a job applicant, as depicted in Figure 2. Recall that no variables can be removed from that distribution based on CI, since the independence would have to hold for all values in the distribution. The Refine-CPD algorithm can determine if context-specific independencies reside in the data. The CPD in Figure 2 can be represented as the CPD-tree in Figure 3.

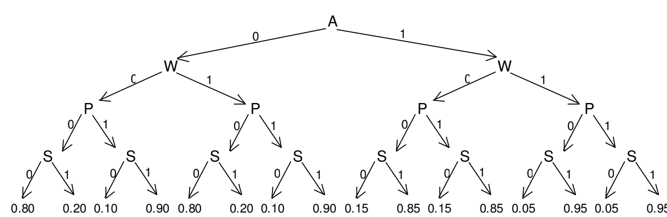

Figure 3: Initial CPD-tree for $p(S|A, W, P)$.

Running the Refine-CPD algorithm yields the refined CPD-tree in Figure 4. The variable W no longer appears on the left side of the tree, in the context $A = 0$. In addition, on the right side of the tree, in context $A = 1$, the variable P no longer appears. This suggests a hidden relationship in variable A in context $A = 0$ and in context $A = 1$.

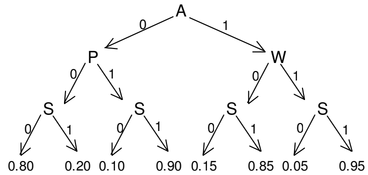

Figure 4: Refined CPD-tree for $p(S|A, W, P)$.

Uncovering Hidden Variables

The previous subsection showed that a CSI discovery algorithm can uncover hidden relationships in a CPD when no causal independencies can be inferred by considering the entire dataset. The example showed that some contexts of A may help explain the relevance of the applicants' absence to the recruitment session. If we look again at Figure 2, and consider $A = 0$ and $A = 1$ separately, we can observe that removing W from the distribution in configurations where $A = 0$ doesn't change the likelihood of occurrence of S, whereas such a removal would be impossible in the context $A = 1$. In $A = 0$, $p(S|A = 0, P, W) = 0.80$ when $P = 0$ and $S = 0$, 0.20 when $P = 0$ and $S = 1$, 0.90 when $P = 1$ and $S = 0$, and finally, 0.10 when $P = 1$ and $S = 1$. In context $A = 1$, saying $p(S|A = 1, P, W) = 0.15$ when $P = 0$ and $S = 0$, is not completely correct since it is also true that in context $A = 1$, $p(S|A = 1, P, W) = 0.05$ when $P = 0$ and $S = 0$. In the first case of context $A = 1$, $W = 0$, while in the second case, $W = 1$. Therefore, the value of W does change the probability of successful interview in context $A = 1$, so no removal is possible. We conclude that in context $A = 0$, variables S and W are independent given variable P. Such a separation is legal since no information is lost due to the union-product operator. From the resulting CPDs, we may now make more adequate judgments about the individuals. The CPD after refinement is presented in Figure 5.

Isolation of contexts suggests different causal models depending upon the value of A. An examination of the semantics of the reduction reveals that in context $A = 0$ (no experience with the public), variable W plays no role in estimating the success of the candidate's interview. Recall that variable W dealt with the candidate's interest in participating in company weekend outings. Since this subset of candidates have no experience with the public and do not seem eager to participate in weekend outings, we are lead to believe that their absence from the recruitment session was due to a true disposition of the person. Therefore, in context $A = 0$, perhaps a different set of variables may better explain what would cause these candidates' interviews to be successful. However, without the discovery of this CSI between S and W in context $A = 0$, we cannot make that conclusion. On the other hand, in context $A = 1$ (experience with the public), we notice that those who didn't attend the recruiting session were influenced by the weekend outings W. Their probability of success was higher when the value of W was equal to 1. Therefore, it is important to keep W in the model for that second subset of candidates since knowing W *does* change our belief in S. However, still in context $A = 1$, after running the discovery algorithm, variable P disappears. Recall that variable P dealt with preparation for the interview. Since P doesn't affect our belief in S in context $A = 1$, we can conclude that these individuals' performance is not affected by whether they prepare for the interview or not. Given that and the fact that they are eager to participate in weekend outings, it is difficult to attribute their non-attendance to the recruiting session to a true disposition. With this new knowledge acquired from the discovery of an unmeasured-in variable, we have enough information to believe that there is something particular about candidates who didn't attend the session, but yet have experience with the public and are eager to socialize with co-workers. With this information, we can look at the applications of those particular applicants to see if our unmeasured-in discovery leads us to discover that perhaps some important information has been left out of the model (unmeasured-out), but for which we could not see the importance unless we discovered the unmeasured-in variable. In this case, we may discover that such candidates all live outside the city, and therefore could not attend the session despite their desire to socialize. This new information would also coincide with their desire for socializing with co-workers on weekends (moving to a new city for a job), and their experience with the public would be a much better indicator of their success in interview than their amount of preparation (unlike their $A = 0$ counterpart). In context $A = 1$, behaviour should clearly be attributed to the situation rather than a true disposition. From this analysis, it is clear that different causal models should be used for the two groups, as the factors that would lead to a successful interview differ greatly between the two. We now see how we can correct the causal models based on the discovered independencies.

AWPS	$p(E\|A,W,P)$
0000	0.80
0001	0.20
0010	0.10
0011	0.90
0100	0.80
0101	0.20
0110	0.10
0111	0.90
1000	0.15
1001	0.85
1010	0.15
1011	0.85
1100	0.05
1101	0.95
1110	0.05
1111	0.95

(i)

AWPS	$p(S\|A{=}0,W,P)$
0000	0.80
0001	0.20
0010	0.10
0011	0.90
0100	0.80
0101	0.20
0110	0.10
0111	0.90

AWPS	$p(S\|A{=}1,W,P)$
1000	0.15
1001	0.85
1010	0.15
1011	0.85
1100	0.05
1101	0.95
1110	0.05
1111	0.95

(ii)

APS	$p(S\|A{=}0,P)$
000	0.80
001	0.20
010	0.90
011	0.10

AWS	$p(S\|A{=}1,W)$
100	0.15
101	0.85
110	0.05
111	0.95

(iii)

Figure 5: Variables S and W are conditionally independent given P in context $A = 0$, while S and P are conditionally independent given W in context $A = 1$.

Correcting the Model

Since there is no longer mention of variable W in context $A = 0$, we can refine our causal model by removing the direct causal link between W and S, and similarly in context $A = 1$ for variable P. With the uncovered hidden contexts of variable A, when considering the probability of a successful job interview S, given all factors that have a direct causal link with S, the initial causal model in Figure 1 can be represented by two more specific causal models that account for differences between the two groups. Those refined causal models are illustrated in Figure 6, where the left side represents the refined model for context $A = 0$ (disposition), and the right side represents the refined model for context $A = 1$ (situation).

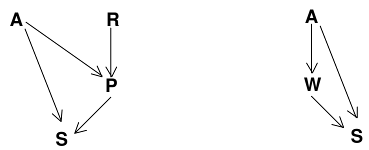

Figure 6: Causal Models After Discovery

Based on the more specific representations of the original causal model, it is now possible to categorize groups of individuals. Candidates in the context $A = 0$, where S and W are independent given P, are likely to be indifferent about the company's weekend activities, as they are disinclined to attend. Candidates in $A = 1$ are likely to be motivated by the idea of a social work culture, since they would be moving to a new city if they were hired. As for interview preparation, candidates in the context $A = 0$ are likely to spend more time and effort on preparation so that they feel more comfortable during the interview by contemplating as many interview scenarios as possible, due to their lack of interpersonal experience. Meanwhile, candidates in the context $A = 1$ are likely to spend less time preparing than those in context $A = 0$.

This example clearly indicates that the reason for attributing a cause to a particular individual differs greatly when we use clues about the situation surrounding the individual at the time of decision, rather than clues about a true disposition of the individual. In addition, the discovery of unmeasured-in hidden variables can help in identifying elements surrounding a situation (based on what variables remain in a context, and which ones disappear) to establish different causal models for dispositions and situations.

Conclusions and Future Work

In attribution theory, the *discounting principle* and the *covariation principle* help determine the attributions people make. In this paper, we have showed how considering context can help uncover true dispositions versus situational factors as explanations for behaviour, with the covariation principle. Contextual independencies can also provide clues regarding consideration of the weaker cause in the discounting principle. We are currently investigating this problem. In addition, the type of contextual discovery of independencies we use in this paper (CSI) can be generalized to deeper contexts, such as contextual weak independencies (CWI) (Butz and Sanscartier, 2002b).

References

Boutilier, C., Friedman, N., Goldszmidt, M., and Koller, D. (1996). Context-specific independence in bayesian networks. In *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, pages 115–123.

Butz, C. and Sanscartier, M. (2002a). A method for detecting context-specific independence in conditional probability tables. In *Third International Conference on Rough Sets and Current Trends in Computing*, pages 344–348.

Butz, C. and Sanscartier, M. (2002b). On the role of contextual weak independence in probabilistic inference. In *Fifteenth Canadian Conference on Artificial Intelligence (AI)*, pages 185–194.

Cartwright, N. (1989). *Nature, Capacities and their Measurements*. Clarendon Press, Oxford.

Good, I. (1983). A causal calculus. *British Journal for Philosophy of Science*, 11.

Jones, E. and Davis, K. (1965). From acts to dispositions: The attribution process in person perception. In Berkowitz, L., editor, *Advances in Experimental Social Psychology*, volume 2, Orlando, FL. Academic Press.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Fransisco USA.

Pearl, J. and Verma, T. (1991). A theory of infered causation. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, pages 441–452. Morgan Kaufmann.

Reichenbach, H. (1956). *The Direction of Time*. University of California Press, Berkeley.

Seifert, C. (1999). Situated cognition and learning. *The MIT Encyclopedia of the cognitive sciences*, pages 767–769.

Suppes, P. (1970). *A Probabilistic Theory of Causation*. North Holland, Amsterdam.

Zhang, N. and Poole, D. (1999). On the role of context-specific independence in probabilistic reasoning. In *Sixteenth International Joint Conference on Artificial Intelligence*, pages 1288–1293.