Characterising Normal Forms for Informational Independence

L.C. van der Gaag & J.-J.Ch. Meyer
Utrecht University, Department of Computer Science
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands
e-mail: {linda,jj}@cs.ruu.nl

Abstract

The concept of *informational independence* plays a key role in most knowledge-based systems. Although defined differently in different contexts, its basic properties have been axiomatised by J. Pearl and his co-researchers. Pearl's axiomatic system offers a set of inference rules for deriving new independence statements from an initial set of statements and as such allows for a *normal form* for representing independences. Pearl's axioms, however, focus on mutually disjoint sets of variables only. We show that focusing on disjoint sets of variables can hide various interesting properties of independence. We extend Pearl's axiomatic system to capture these properties and thereby introduce a new normal form for informational independence.

1 Introduction

The concept of *informational independence* pervades most knowledge-based systems. The concept is used for example for demarcating a system's scope. But more importantly, it is making effective use of knowledge about independences that renders knowledge-based systems capable of dealing with the computational complexity of their problem-solving tasks.

The concept of informational independence generally is defined in different terms in different contexts. For example, in knowledge-based systems built on probability theory, informational independence is identified with statistical independence among sets of variables; in constraint satisfaction systems, informational independence is defined in terms of constraints — two variables are said to be independent if restricting the domain of one variable leaves the other one's domain unaltered. Despite these definitional differences, the various perspectives on informational independence share the same basic properties.

The basic properties of informational independence have been identified and taken to constitute an *axiomatic system for informational independence* by J. Pearl and his co-researchers [1, 2]. There are many advantages to an axiomatic system for informational independence. Such a system offers a set of inference rules for computing new independence statements from an initial set of statements. As such, it allows for a *normal form* for independence and, hence, provides for a concise representation of a set of independences. An axiomatic system may further be used for verifying whether a new statement logically follows from a set of independence statements and for studying inconsistency among independences. These advantages fall in with the advantages of axiomatisation in general.

Pearl's axiomatic system for informational independence focuses on independences among *mutually disjoint* sets of variables only. As a consequence, its associated normal form also focuses on mutually disjoint sets. The concept of informational independence itself, however, is not restricted to mutually disjoint sets of variables. In this paper, we examine the consequences of Pearl's restricted focus in the context of probability theory. We show that focusing on disjoint sets of variables can hide various interesting properties of informational independence. To capture these properties, we identify several new axioms to include in Pearl's axiomatic system. The thus extended axiomatic system allows for a new normal form for informational independence that is more general in scope than Pearl's normal form in that it also addresses independences among overlapping sets of variables.

The remainder of the paper is structured as follows. We review Pearl's axiomatic system for informational independence in Section 2 and discuss the extension Pearl has proposed for his system in Section 3. The main results of our analysis of the restricted focus on disjoint sets of variables, including a new normal form for informational independence, are presented in Section 4; in a forthcoming paper, we will provide further details and full proofs. The present paper is rounded off with some conclusions in Section 5.

2 Independence Revisited

The concept of informational independence has been studied in various different contexts. Especially in the context of *probability theory* has independence been a subject of extensive studies, see for example [3, 4]. The main objective of the early statistical studies was to identify and express in algebraic form independence relations of probability distributions to allow for comparison and classification. J. Pearl and his co-researchers were among the first to formalise properties of independence in an axiomatic system and to develop a logic for informational independence [5, 6, 7].

In the context of probability theory, the concept of independence is generally introduced in terms of numerical quantities: the independence relation of a probability distribution is taken to be implicitly embedded in the probabilities involved. A definition of independence in terms of numbers suggests that, in order to determine whether two sets of variables are (conditionally) independent, several probabilities have to be computed and several equalities have to be tested; moreover, such a definition suggests that for determining independence a probability distribution on the variables discerned has to be explicitly available. In contrast, humans tend to be able to state directly, with conviction and consistency, whether or not two sets of variables are independent. Such statements of independence typically are issued qualitatively, without any reference to exact probabilities. Based on these observations, Pearl argues that the concept of independence is far more basic to human reasoning than its numerical definition suggests and that in fact the definition of independence in terms of probabilities may be looked upon as a *quantitative* way of capturing the basic concept which is *qualitative* in nature [1]. Pearl's aim in designing an axiomatic system for informational independence now is to provide an explicit qualitative definition of independence.

We begin our review of Pearl's axiomatic system for informational independence by introducing some notational convention. Let V be a finite set of (discrete) variables and let Pr be a joint probability distribution on V. Then, the *independence relation* $I_{\mathrm{Pr}} \subseteq \mathcal{P}(V) \times \mathcal{P}(V) \times \mathcal{P}(V)$ of Pr is defined by $(X, Z, Y) \in I_{\mathrm{Pr}}$ if and only if $\Pr(X = x \mid Y = y \wedge Z = z) = \Pr(X = x \mid Z = z)$ for all value assignments x, y, z to the sets of variables $X, Y, Z \subseteq V$, respectively. In the sequel, we will omit the subscript Pr from the notation I_{Pr} as long as ambiguity cannot occur. Also, we will write $I(X, Z, Y)$ to denote $(X, Z, Y) \in I$ and $\neg I(X, Z, Y)$ to denote $(X, Z, Y) \notin I$. A statement $I(X, Z, Y)$ of a probability distribution's independence relation I is called an *independence statement*. In qualitative terms, an independence statement $I(X, Z, Y)$ expresses that in the context of information about Z information about Y is irrelevant with respect to X.

In designing his axiomatic system for informational independence, Pearl identifies various properties that are satisfied by any probability distribution's independence relation and takes these as axioms for the qualitative concept of informational independence [1].

Definition 2.1 *Let V be a finite set of variables. A* semi-graphoid independence relation *on V is a ternary relation $I \subseteq \mathcal{P}(V) \times \mathcal{P}(V) \times \mathcal{P}(V)$ such that I satisfies the properties*

- $I(X, Z, Y) \rightarrow I(Y, Z, X)$;
- $I(X, Z, Y \cup W) \rightarrow I(X, Z, Y) \wedge I(X, Z, W)$;
- $I(X, Z, Y \cup W) \rightarrow I(X, Z \cup W, Y)$;
- $I(X, Z, Y) \wedge I(X, Z \cup Y, W) \rightarrow I(X, Z, Y \cup W)$;

for all mutually disjoint sets $X, Y, Z, W \subseteq V$. A graphoid independence relation *I on V is a semi-graphoid independence relation on V such that I satisfies the additional property*

- $I(X, Z \cup W, Y) \wedge I(X, Z \cup Y, W) \rightarrow I(X, Z, Y \cup W)$;

for all mutually disjoint sets $X, Y, Z, W \subseteq V$.

The properties stated in Definition 2.1 with each other convey the idea that learning irrelevant information does not alter the independences among the variables discerned; for a discussion of the qualitative meanings of these properties, we refer the reader to [1]. Note that from Definition 2.1 and the basic axioms of probability theory, we have that the independence relation of any probability distribution Pr is a semi-graphoid independence relation; furthermore, if Pr is strictly positive, that is, if Pr does not comprise any non-trivial zero probabilities, then its independence relation is a graphoid independence relation. In the sequel, Definition 2.1 will be referred to as *Pearl's (restricted) axiomatic system for informational independence.*

3 Pearl's Extended Axiomatic System

Pearl's axiomatic system for informational independence involves axioms for mutually disjoint sets of variables only. The basic concept of independence, however, is not restricted to mutually disjoint sets of variables. In the context of probability theory, a probability distribution's independence relation typically includes independence statements involving overlapping sets of variables. To fully capture the basic concept of independence, an axiomatic system should therefore provide axioms, not just for mutually disjoint sets, but for overlapping sets of variables as well. In this section, we review the extension Pearl has proposed for this purpose for his axiomatic system.

In the context of probability theory, the properties in Definition 2.1, although stated to hold for mutually disjoint sets of variables only, hold for overlapping sets as well, as is easily verified. The axioms of Pearl's definition of informational independence therefore are generalised straightforwardly to apply to overlapping sets of variables. In addition, Pearl proposes including an extra axiom in his system. This axiom has its motivational foundation in the property

$$I(X, Z, Y) \leftrightarrow I(X \Leftrightarrow Z, Z, Y \Leftrightarrow Z)$$

for all sets $X, Y, Z \subseteq V$, first identified by A.P. Dawid to hold for any probability distribution's independence relation [3]. To allow for deriving Dawid's property for overlapping sets of variables, Pearl introduces the axiom

$$I(X, Z, Z)$$

for all sets $X, Z \subseteq V$ [1]. This axiom asserts that, once information about Z is known, learning information about Z is irrelevant with respect to any set of variables X. The validity of the axiom in the context of probability theory is easily verified. Pearl's *extended* axiomatic system for informational independence is summarised in the following definition.

Definition 3.1 *Let V be a finite set of variables. An* extended *semi-graphoid independence relation on V is a ternary relation $I \subseteq \mathcal{P}(V) \times \mathcal{P}(V) \times \mathcal{P}(V)$ such that I satisfies the properties*

- $I(X, Z, Y) \rightarrow I(Y, Z, X)$;
- $I(X, Z, Y \cup W) \rightarrow I(X, Z, Y) \wedge I(X, Z, W)$;
- $I(X, Z, Y \cup W) \rightarrow I(X, Z \cup W, Y)$;
- $I(X, Z, Y) \wedge I(X, Z \cup Y, W) \rightarrow I(X, Z, Y \cup W)$;
- $I(X, Z, Z)$;

for all *sets $X, Y, Z, W \subseteq V$. An* extended *graphoid independence relation I on V is an extended semi-graphoid independence relation on V such that I satisfies the additional property*

- $I(X, Z \cup W, Y) \wedge I(X, Z \cup Y, W) \rightarrow I(X, Z, Y \cup W)$;

for all *sets $X, Y, Z, W \subseteq V$.*

Note that from Definition 3.1 and the basic axioms of probability theory, we have that the independence relation of any probability distribution Pr is an extended semi-graphoid independence relation; furthermore, if Pr is strictly positive, then its independence relation is an extended graphoid independence relation.

The following lemma confirms that Dawid's property for overlapping sets of variables can be derived from Pearl's extended axiomatic system.

Lemma 3.2 *Let V be a finite set of variables and let I be an extended semi-graphoid independence relation on V. Then, $I(X, Z, Y) \leftrightarrow I(X \Leftrightarrow Z, Z, Y \Leftrightarrow Z)$, for all sets $X, Y, Z \subseteq V$.*

4 On Normal Forms

In his work on graphical representations of independence relations, Pearl builds on the *restricted* axiomatic system for informational independence, that is, on the system that involves axioms for mutually disjoint sets of variables only [1]. In this section, we address the question whether this focus on disjoint sets of variables is an essential one. To this end, we examine an independence relation's set of statements that involve overlapping sets of variables. Note that, if any such statement can be derived by the extended axiomatic system from the relation's set of statements that involve mutually disjoint sets of variables only, then Pearl's focus on disjoint sets is not essential and any theory built on the restricted axiomatic system holds for independence relations in general. In fact, Pearl's extended axiomatic system can then be looked upon as allowing for a normal form for informational independence that warrants building on the restricted system.

Definition 4.1 *Let V be a finite set of variables and let I be an extended semi-graphoid independence relation on V. An independence statement $I(X, Z, Y)$ is said to be in* Pearl normal form *if the sets $X, Y, Z \subseteq V$ are mutually disjoint.*

In addressing Pearl's focus on disjoint sets of variables, we distinguish between strictly positive probability distributions and distributions more in general.

4.1 Strictly Positive Distributions

In designing his axiomatic system for informational independence, Pearl has singled out strictly positive probability distributions because these distributions satisfy properties that are not satisfied by probability distributions in general. This special attention for strictly positive distributions has resulted in the concept of a graphoid independence relation as a special type of semi-graphoid independence relation. We now reconsider the independence relation of a strictly positive distribution and examine its set of independence statements that involve overlapping sets of variables.

Strictly positive probability distributions are well-known for the property that they do not embed any *functional dependences* among their variables. In a strictly positive distribution, therefore, there are no variables whose value is completely determined by some other variables' values. This lack of functional dependences in strictly positive probability distributions is reflected in their independence relations.

Proposition 4.2 *Let V be a finite set of variables. Let* Pr *be a strictly positive joint probability distribution on V and let I be its independence relation. Then,*

$$\neg I(\{V_i\}, Z, \{V_i\})$$

for each variable $V_i \in V$ and all sets $Z \subseteq V \Leftrightarrow \{V_i\}$.

The property stated in Proposition 4.2 expresses that in a strictly positive probability distribution information about a variable V_i *cannot* be irrelevant with respect to V_i itself *as long as no information about V_i is available as yet.* So, this property conveys that indeed there do not exist any functional dependences among the variables discerned. Note that, if information about V_i *is* available, that is, if $V_i \in Z$, we have $I(\{V_i\}, Z, \{V_i\})$ by Pearl's extended axiomatic system.

The concept of an extended graphoid independence relation has been designed to capture independence relations of strictly positive probability distributions. Unfortunately, this concept does not reflect the property stated in Proposition 4.2: an extended graphoid independence relation may very well include statements $I(\{V_i\}, Z, \{V_i\})$ for some $V_i \in V$ and $Z \subseteq V \Leftrightarrow \{V_i\}$. To exclude such statements, we add to Pearl's extended axiomatic system the axiom

$$\neg I(\{V_i\}, Z, \{V_i\})$$

for each variable $V_i \in V$ and all sets $Z \subseteq V \Leftrightarrow \{V_i\}$, for extended graphoid independence relations. Note that, in contrast with all other axioms in Pearl's extended axiomatic system, this axiom is *negative* in the sense that it explicitly states which independence statements do *not* hold. In fact, the axiom expresses a dependence and may be looked upon as providing for a *dependence closure*.

To conclude, we address the normal form for informational independence in strictly positive probability distributions allowed for by our new extended axiomatic system. From the new axiom we have that the independence relation of a strictly positive probability distribution does not comprise any statement of the form $I(\{V_i\}, Z, \{V_i\})$ with $V_i \notin Z$. In fact, the independence relation does not even include statements of the form $I(X, Z, Y)$ with $X \cap Y \not\subseteq Z$, as is stated more formally in the following lemma.

Lemma 4.3 *Let V be a finite set of variables. Let* Pr *be a strictly positive joint probability distribution on V and let I be its independence relation. Then, $\neg I(X, Z, Y)$ for all sets $X, Y, Z \subseteq V$ with $X \cap Y \not\subseteq Z$.*

From Lemma 4.3 we have that *all* statements of the independence relation of a strictly positive probability distribution that involve overlapping sets of variables are of the form $I(X, Z, Y)$ with $X \cap Y \subseteq Z$. From Lemma 3.2 we now have that, if $I(X, Z, Y)$ is a valid statement in I, then $I(X \Leftrightarrow Z, Z, Y \Leftrightarrow Z)$ is also in I. In addition, we have that $I(X, Z, Y)$ can be derived from this statement by the (new) extended axiomatic system. Note that the statement $I(X \Leftrightarrow Z, Z, Y \Leftrightarrow Z)$ is in Pearl normal form. So, for any independence statement s in I that is not in Pearl normal form, there exists in I a statement s' that is in Pearl normal form such that $s' \Leftrightarrow s$. This property is depicted schematically in Figure 1. In the figure, I represents the set of all statements in the independence relation at hand; I_{PNF} represents the subset of statements that are in Pearl normal form.

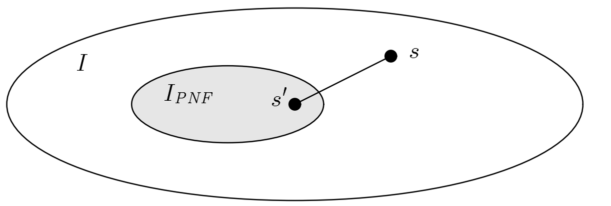

Figure 1: The Derivability of Statements in $I \Leftrightarrow I_{PNF}$ for Strictly Positive Probability Distributions.

From the above observations, we conclude that in view of independence relations of strictly positive probability distributions, Pearl's focus on mutually disjoint sets of variables is not essential: any statement from such an independence relation that is not in Pearl normal form can be derived by our new extended axiomatic system from the relation's set of statements that are in normal form. Pearl normal form therefore suffices for representing the relation at hand.

4.2 General Probability Distributions

While strictly positive probability distributions do not embed any functional dependences among their variables, probability distributions in general may very well do so: a probability distribution may involve one or more *deterministic* variables whose value is completely determined by some other variables' values. We now reconsider independence relations of probability distributions that may embed functional dependences; in doing so, we once more examine their sets of independence statements that involve overlapping sets of variables.

In the previous section, we have seen that lack of functional dependences in a probability distribution gives rise to the property $\neg I(\{V_i\}, Z, \{V_i\})$ for each variable $V_i \in V$ and all sets $Z \subseteq V \Leftrightarrow \{V_i\}$. In the presence of functional dependences this property no longer holds, as the dependences are reflected in the independence relation at hand: the independence relation of a probability distribution that embeds one or more functional dependences among its variables typically includes statements of the form $I(\{V_i\}, Z, \{V_i\})$ with $V_i \notin Z$. Such an independence statement expresses that in the context of information about Z, information about V_i is irrelevant with respect to V_i itself, *even if no information about V_i is available as yet.* From a qualitative point of view, independence statements of this form may seem couterintuitive: intuitively, no variable can be independent from itself. Recall, however, that we build on the *probabilistic* concept of independence which is defined in terms of numerical quantities: the statement $I(\{V_i\}, Z, \{V_i\})$ reflects that V_i's value is completely determined by the information about Z and, hence, that V_i is functionally dependent upon Z. The presence of statements of this form in a probability distribution's independence relation gives rise to the property stated in the following proposition.

Proposition 4.4 *Let V be a finite set of variables. Let* Pr *be a joint probability distribution on V and let I be its independence relation. Then,*

$$I(X, Z, X) \rightarrow I(X, Z, Y)$$

for all sets $X, Y, Z \subseteq V$.

The property stated in Proposition 4.4 expresses that, if information about X is irrelevant with respect to X itself in the context of some information about Z, then information about *any* set of variables is irrelevant with respect to X in this context. Note that this property cannot be derived from the axioms stated in Definition 3.1 and therefore is logically independent from Pearl's extended axiomatic system.

So far, we have considered in a distribution's independence relation the statements for the set of functionally dependent variables only. We now turn to the effect of the presence of functional dependences on the independences among the other variables discerned.

Proposition 4.5 *Let V be a finite set of variables. Let* Pr *be a joint probability distribution on V and let I be its independence relation. Then,*

$$I(X, Z, X) \wedge I(Y, Z, W) \rightarrow I(X \cup Y, Z, X \cup W)$$

for all sets $X, Y, Z, W \subseteq V$.

The property stated in Proposition 4.5 reflects that once the values of the functionally dependent variables in X are completely determined by some information about Z, then learning information about X cannot alter the independences among all other variables discerned. Note once more that this property cannot be derived from the axioms stated in Definition 3.1 nor from the property stated in Proposition 4.4. The following lemma shows that the property stated in Proposition 4.5 is easily generalised to a bi-implication.

Lemma 4.6 *Let V be a finite set of variables. Let* Pr *be a joint probability distribution on V and let I be its independence relation. Then, $I(X, Z, X) \wedge I(Y, Z, W) \leftrightarrow I(X \cup Y, Z, X \cup W)$ for all sets $X, Y, Z, W \subseteq V$.*

The concept of an extended semi-graphoid independence relation has been designed to capture independence relations of probability distributions in general. An extended semi-graphoid independence relation, however, need not satisfy the properties stated in the Propositions 4.4 and 4.5. We therefore add to Pearl's extended axiomatic system the axioms

$$I(X, Z, X) \rightarrow I(X, Z, Y)$$

and

$$I(X, Z, X) \wedge I(Y, Z, W) \rightarrow I(X \cup Y, Z, X \cup W)$$

for all sets of variables $X, Y, Z, W \subseteq V$, for semi-graphoid independence relations.

To conclude, we address the normal form for informational independence in probability distributions in general allowed for by our new extended axiomatic system. Any statement in the independence relation of a probability distribution that involves overlapping sets of variables is of one of the following forms

- $I(X, Z, Y)$ with $X \cap Y \subseteq Z$; or,
- $I(X, Z, Y)$ with $X \cap Y \not\subseteq Z$.

For statements of the form $I(X, Z, Y)$ with $X \cap Y \subseteq Z$, we observe that similar observations apply as for the independences in a strictly positive probability distribution. Note that if in a distribution's independence relation all statements involving overlapping sets of variables are of this form, then the distribution does not embed any functional dependences.

We now consider the independence statements of the form $I(X, Z, Y)$ with $X \cap Y \not\subseteq Z$; without loss of generality we assume that $(X \cap Y) \cap Z = \varnothing$. From Lemma 4.6 we have that if $I(X, Z, Y)$ is a valid statement in I, then both $I(X \cap Y, Z, X \cap Y)$ and $I(X \Leftrightarrow Y, Z, Y \Leftrightarrow X)$ are in I. In addition we have that $I(X, Z, Y)$ can be derived from these statements by the new extended axiomatic system. Note that while the statement $I(X \Leftrightarrow Y, Z, Y \Leftrightarrow X)$ is in Pearl normal form, the statement $I(X \cap Y, Z, X \cap Y)$ is not. The latter statement, however, is also of a special, restricted form as it models functional dependence of $X \cap Y$ on Z; we say that the statement is in *functional form*. We now have that, for any independence statement s in I that is not in Pearl normal form nor in functional form, there exist in I a statement s' that is in Pearl normal form and a statement s'' that is in functional form such that $s' \wedge s'' \Leftrightarrow s$. This property is depicted in Figure 2. In the figure, I represents the set of all independence statements in the independence relation at hand; I_{FF} represents the set of all statements that are in functional form; I_{PNF} represents all statements from I that are in Pearl normal form.

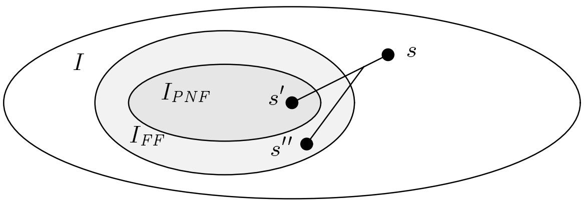

Figure 2: The Derivability of Statements in $I \Leftrightarrow (I_{PNF} \cup I_{FF})$ for General Probability Distributions.

From the above observations we conclude that in view of an independence relation of a probability distribution that embeds functional dependences, Pearl's focus on mutually disjoint sets of variables hides an interesting and important set of independence statements, namely the independence statements arising from functional dependences: these statements cannot be derived by Pearl's extended axiomatic system from the relation's set of statements that are in normal form. Pearl normal form therefore does not suffice for representing such an independence relation. We introduce a new normal form for this purpose in the following definition.

Definition 4.7 *Let V be a finite set of variables and let I be an extended semi-graphoid independence relation on V. An independence statement $I(X, Z, Y)$ is said to be in* normal form *if the sets $X, Y, Z \subseteq V$ are mutually disjoint, or if X and Z are mutually disjoint and $X = Y$.*

We would like to note that in the context of graphical representations of independence Pearl and his co-researchers have pointed out the necessity of explicitly modelling all functional dependences [2]; our new normal form falls in with and extends on their observations.

5 Conclusions

The concept of informational independence has been axiomatised by J. Pearl and his co-researchers. Pearl's axiomatic system focuses on independences among mutually disjoint sets of variables only. We have shown that this focus on disjoint sets can hide various interesting properties of informational independence, properties arising from functional dependences among the variables discerned. To capture these properties we have extended Pearl's axiomatic system with several new axioms. The thus extended axiomatic system allows for a normal form for informational independence that elucidates the necessity of explicitly representing the functional dependences among the variables of an independence relation.

References

[1] J. Pearl (1988). *Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference*, Morgan Kaufmann, Palo Alto.

[2] J. Pearl, D. Geiger, and T. Verma (1990). The logic of influence diagrams, in: R.M. Oliver and J.Q. Smith (eds). *Influence Diagrams, Belief Nets and Decision Analysis*, John Wiley & Sons, pp. 67 – 86.

[3] A.P. Dawid (1979). Conditional independence in statistical theory, *Journal of the Royal Statistical Society (Series B)*, vol. 41, pp. 1 – 31.

[4] S.L. Lauritzen (1982). *Lectures on Contingency Tables*, 2nd. Ed., University of Aalborg Press, Aalborg, Denmark.

[5] J. Pearl and A. Paz (1985). GRAPHOIDS: a graph-based logic for reasoning about relevance relations, in: B. Du Boulay, D. Hogg, and L. Steels (eds). *Advances in Artificial Intelligence 2*, North-Holland.

[6] J. Pearl and T.S. Verma (1987). The logic of representing dependencies by directed acyclic graphs, *Proceedings of the Sixth National Conference on Artificial Intelligence*, pp. 374 – 379.

[7] D. Geiger and J. Pearl (1988). On the logic of causal models, *Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence*, pp. 136 – 147.