Available online at www.sciencedirect.com

ScienceDirect

Journal of Applied Logic 7 (2009) 188–205

JOURNAL OF APPLIED LOGIC

www.elsevier.com/locate/jal

Causal discovery and the problem of ignorance. An adaptive logic approach

Bert Leuridan [1]

Centre for Logic and Philosophy of Science, Ghent University, Belgium

Received 3 July 2007; received in revised form 18 October 2007; accepted 7 November 2007

Available online 20 February 2008

Abstract

In this paper, I want to substantiate three related claims regarding causal discovery from non-experimental data. Firstly, in scientific practice, the *problem of ignorance* is ubiquitous, persistent, and far-reaching. Intuitively, the problem of ignorance bears upon the following situation. A set of random variables $\mathbb{V}$ is studied but only *partly* tested for (conditional) independencies; i.e. for some variables A and B *it is not known whether they are* (*conditionally*) *independent*. Secondly, Judea Pearl's most meritorious and influential algorithm for causal discovery (the **IC** algorithm) cannot be applied in cases of ignorance. It presupposes that a *full list* of (conditional) independence relations is on hand and it would lead to unsatisfactory results when applied to *partial* lists. Finally, the problem of ignorance is successfully treated by means of **ALIC**, the *adaptive logic* for causal discovery presented in this paper.

Keywords: Causal discovery; Adaptive logic

1. Introduction: causal discovery and the problem of ignorance

Since the end of the 1980s, the interrelations between probability theory, graph theory and causal discovery have been studied by increasing numbers of research groups. Different algorithms have been developed to infer causal relations from non-experimental statistical data.[2]

In this paper, I will discuss Judea Pearl's **IC** algorithm [18, pp. 50–51]. It is one of the best known algorithms for causal discovery, or '*i*nductive *c*ausation', and its merits can hardly be overrated. Nevertheless, it faces a hard and very important problem. In this section, I want to substantiate three related claims. Firstly, in scientific practice, the *problem of ignorance* is ubiquitous, persistent, and far-reaching. Intuitively, the problem of ignorance bears upon the following situation. A finite set of random variables $\mathbb{V}$ is studied but only *partly* tested for (conditional) independencies; i.e. for some variables A, B and for some sets of variables $\mathbf{Q}$ *it is not known whether A and B are independent* (*conditional on* $\mathbf{Q}$). So 'ignorance' should not be understood as 'probabilistic knowledge', or 'degree of belief <1', but as the

E-mail address: Bert.Leuridan@Ugent.be.
[1] Ph.D. fellow of the Research Foundation—Flanders (FWO).
[2] See Spirtes et al. [19, Chapters 5–6] and Neapolitan [15] for an overview of several different such algorithms.

1570-8683/$ – see front matter
doi:10.1016/j.jal.2007.11.004

existence of 'undecided independencies'.[3] Secondly, the **IC** algorithm cannot be applied in cases of ignorance. It presupposes that a *full list* of (conditional) independence relations is on hand.[4] If it would be applied to *partial* lists, this would moreover lead to unsatisfactory results. Thirdly, the problem of ignorance can be solved without losing the strong points of the **IC** algorithm, viz. by means of an *adaptive logic for causal discovery*.[5]

Let me shortly dwell on the first claim. In scientific practice, for example in the social sciences or in epidemiology, the problem of ignorance is *ubiquitous*. This may be illustrated by an example. The influence of many different factors on cognitive skills and educational achievement are studied by many different research groups. Some research groups focus on cultural factors, some on sociological factors, still others on psychological ones. Other research groups focus on biological, chemical or other factors. In total, hundreds of variables are studied.[6] The combined research of all these groups gives rise to the confirmation and disconfirmation of many *CIRs* and *UIRs*, as in each study several possible confounders are tested for.[7] However, not all possible (conditional) independencies are tested.

The problem of ignorance is moreover *persistent*. Even if many conditional independence relations can be ruled out *a priori*, on the basis of reliable background knowledge, micro-level knowledge, common sense arguments, ..., many others will still be undecided.

Finally, the problem of ignorance is *far-reaching*. If the causal interpretation of non-experimental data requires that all (conditional) independencies are decided, as in the **IC** algorithm, then observational science would lose large part of its materiality. No causal knowledge could ever be obtained in the short run.[8]

In Section 5, I will present **ALIC**, the adaptive logic for causal discovery which properly solves the problem of ignorance while doing justice to the merits of Pearl's **IC**. First, however, I will present the formal background to **IC** (Section 2), and also the algorithm itself (Section 3). Then I will present **LIC**, a non-adaptive logic for causal discovery (Section 4). By itself, **LIC** has little to add to the **IC** algorithm, but its significance derives from the role it plays in the formulation of **ALIC**.

2. Formal background to causal discovery

In this section I will shortly present the formal background to **IC**. Section 2.1 deals with directed acyclic graphs and their relations with probability distributions. Section 2.2 is on faithful indistinguishability classes and patterns. Finally, Section 2.3 treats of the graphoid axioms and their meta-theoretic properties (the relevance of which will prove in Sections 4.4 and 5.5).

2.1. Directed acyclic graphs and probability distributions

In this paper, causal structures will be described by means of directed acyclic graphs, or *DAGs*. A *graph* $G = \langle \mathbb{V}, \mathbb{E} \rangle$ consists of a finite set of *vertices* $\mathbb{V}$ and a finite set of *edges* $\mathbb{E}$. In a *directed graph*, all edges are *directed* ($\rightarrow$). Two vertices A and B are *adjacent* ($A - B$) iff either $A \rightarrow B \in \mathbb{E}$, or $B \rightarrow A \in \mathbb{E}$. There is a *path* between A and B iff there is a sequence of adjacent vertices, beginning with A and ending with B. A *directed path* ($A \Rightarrow B$) is a path that has no *colliders* ($X \rightarrow Y \leftarrow Z$) or *forks* ($X \leftarrow Y \rightarrow Z$). A path that contains no vertex more than once is *acyclic*. A *directed, acyclic graph* (*DAG*) is a graph that contains no directed, cyclic paths.

[3] In 2006, Rolf Haenni and Stephan Hartmann devoted a special issue of *Minds and Machines* to the topic of Causality, Uncertainty and Ignorance. Unfortunately, none of the papers in question treated 'ignorance' in the sense just stated, viz. the presence of undecided independencies [8].

[4] In the rest of this paper, I will use the following abbreviations. '*UIR*' will stand for 'unconditional independence relation'. Likewise, '*CIR*' will stand for 'conditional independence relation'. '*IR*' will be used as an umbrella term. Note that the relations in question are particular relations; they are relations between particular (sets of) variables.

[5] Since the adaptive logic to be presented is based on **IC**, it involves only finite sets of finite variables. Like **IC**, it also assumes causal sufficiency (i.e. if two variables under study share a common cause, this cause is observed, too).

[6] The ISI Web of Knowledge cites hundreds of articles on "educational achievement", published between 2000 and 2007. Many of them report non-experimental data (e.g. the National Education Longitudinal Study). The scope of the factors studied is huge: it ranges from television viewing, social capital and self-esteem over parasites, maternal smoking, birth order, ethnicity, ...

[7] I use 'confirmation' and 'disconfirmation' in a loose sense, here.

[8] The problem of ignorance not only lurks in scientific practice, but also in everyday human reasoning. Humans often endorse or disaffirm *CIRs* and *UIRs* between many different variables, thereby leaving undecided a large number of *IRs*.

Directed acyclic graphs are closely connected to probability distributions. The vertices in $\mathbb{V}$ may denote random variables with a finite number of discrete values. In this paper, random variables are represented by italicized capital letters: e.g. $A, B, C, \ldots$ or $X, Y, Z, \ldots$ Values of variables are represented by italicized small letters: e.g. $a, b, c, \ldots$ or $x, y, z, \ldots$ Sets of variables are denoted by bold capital letters: e.g. $\mathbf{A}, \mathbf{B}, \mathbf{C}, \ldots$ or $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \ldots$ Configurations of values for all members of a set of variables are denoted by bold small letters: e.g. $\mathbf{a}, \mathbf{b}, \mathbf{c}, \ldots$ or $\mathbf{x}, \mathbf{y}, \mathbf{z}, \ldots$

Let $P(\mathbb{V})$ be a joint distribution over $\mathbb{V}$. P may verify some independence relations (*IRs*).

Definition 1 *((Un)conditional independence).* According to P, $\mathbf{A}, \mathbf{B} \subseteq \mathbb{V}$ are independent conditional on $\mathbf{Q} \subseteq \mathbb{V}$, in short $(\mathbf{A} \amalg_P \mathbf{B} \mid \mathbf{Q})$, iff $P(\mathbf{a} \mid \mathbf{b}, \mathbf{q}) = P(\mathbf{a} \mid \mathbf{q})$ for all $\mathbf{a}, \mathbf{b}$ and $\mathbf{q}$ (whenever $P(\mathbf{b}, \mathbf{q}) > 0$). Likewise, $\mathbf{A}$ and $\mathbf{B}$ are unconditionally independent, in short $(\mathbf{A} \amalg_P \mathbf{B})$, iff $P(\mathbf{a} \mid \mathbf{b}) = P(\mathbf{a})$ (whenever $P(\mathbf{b}) > 0$)[9] [18, p. 11].

A *DAG* $G = \langle \mathbb{V}, \mathbb{E} \rangle$ can be used to represent the *IRs* verified by $P(\mathbb{V})$. To that extent, P and G need to satisfy two conditions: the Markov condition and the Faithfulness condition. These conditions are formulated in terms of kinship relations between variables. B is a *parent* of A iff $B \to A \in \mathbb{E}$. *Parents*(A) is the set of parents of A. Other kinship relations are defined likewise. By convention, A is its own child and descendant, even though $A \to A$ and $A \Rightarrow A$ are ruled out in *DAGs*.

Convention 1. $A \in \mathit{Children}(A)$, and hence also $A \in \mathit{Descendants}(A)$.

Definition 2 *(Markov).* (See [18, p. 30].) $G = \langle \mathbb{V}, \mathbb{E} \rangle$ and $P(\mathbb{V})$ satisfy the Markov condition iff for every $A \in \mathbb{V}$,

$$(A \amalg \mathit{Nondescendants}(A) \setminus \mathit{Parents}(A) \mid \mathit{Parents}(A))$$

Given any graph G, the Markov condition generates a set of *IRs*. But probability distributions that are Markov to G may verify extra *IRs*, too. The faithfulness condition rules out such distributions.

Definition 3 *(Faithfulness).* Let $P(\mathbb{V})$ be a probability distribution generated by $G = \langle \mathbb{V}, \mathbb{E} \rangle$ according to the Markov condition. G and P satisfy the faithfulness condition iff every *IR* true in P is entailed by the Markov condition applied to G [19, p. 13].

As I stated earlier, a *DAG* $G = \langle \mathbb{V}, \mathbb{E} \rangle$ may be used to represent the *IRs* verified by some $P(\mathbb{V})$. If P is Markov and faithful to G, then all and only those *IRs* that are entailed by the Markov condition applied to G are true in P. It is difficult, however, to delineate the set of these relations. The graph-theoretical concept of d-separation provides an easy means to do this [19, p. 44].

Definition 4 *(d-separation).* Let $G = \langle \mathbb{V}, \mathbb{E} \rangle$ be a *DAG*. If $\mathbf{Q} \subset \mathbb{V}$ and $A, B \in \mathbb{V} \setminus \mathbf{Q}$, then A and B are d-separated given $\mathbf{Q}$ in G, in short $(A \amalg_G B \mid \mathbf{Q})$ iff there is no path U between A and B, such that

1. for every collider $\ldots \to C \leftarrow \ldots$ on U, $\mathit{Descendants}(C) \cap \mathbf{Q} \neq \emptyset$,[10]
2. and no other vertex on U is in $\mathbf{Q}$.

If $\mathbf{X} \neq \emptyset$, $\mathbf{Y} \neq \emptyset$ and $\mathbf{Z}$ are three disjoint sets, then $\mathbf{X}$ is d-separated from $\mathbf{Y}$ given $\mathbf{W}$ iff every member of $\mathbf{X}$ is d-separated from every member of $\mathbf{Y}$ given $\mathbf{Z}$.

2.2. *Faithful indistinguishability classes and patterns*

The relation between conditional independence and d-separation is characterized by the following two theorems, the last of which lays at the basis of Judea Pearl's **IC** algorithm for causal discovery.

9 Conditional probability is defined as follows: $P(\mathbf{a} \mid \mathbf{b}) = P(\mathbf{a}, \mathbf{b})/P(\mathbf{b})$.

10 Note again that $C \in \mathit{Descendants}(C)$, by Convention 1.

Theorem 1. *$P(\mathbb{V})$ is Markov and faithful to a DAG $G = \langle \mathbb{V}, \mathbb{E} \rangle$ iff for all disjoint sets* $\mathbf{A}, \mathbf{B}$ *and* $\mathbf{Q}$, $(\mathbf{A} \amalg_P \mathbf{B} \mid \mathbf{Q})$ *iff* $(\mathbf{A} \amalg_G \mathbf{B} \mid \mathbf{Q})$

Proof. See Spirtes et al. [19, pp. 385–393]. □

Theorem 2. *If $P(\mathbb{V})$ is faithful to* some *DAG, then it is faithful to the DAG $G = \langle \mathbb{V}, \mathbb{E} \rangle$ iff*

1. *for all $A, B \in \mathbb{V}$, $A - B$ iff $\sim (A \amalg_P B \mid \mathbf{Q})$ for all $\emptyset \subseteq \mathbf{Q} \subseteq \mathbb{V} \setminus \{A, B\}$;*
2. *and for all $A, B, C \in \mathbb{V}$ such that $A - B - C$, but not $A - C$, $A \rightarrow B \leftarrow C$ is a subgraph of G iff $\sim (A \amalg_P C \mid \mathbf{Q} \cup \{B\})$ for all $\emptyset \subseteq \mathbf{Q} \subseteq \mathbb{V} \setminus \{A, B, C\}$.*

Proof. See Spirtes et al. [19, pp. 393–394]. □

The **IC** algorithm takes as its input a probability distribution P that is Markov and faithful to some underlying *DAG* G_0.[11] The intended output is the *DAG* G_0. In most cases, however, several *DAGs* are statistically indistinguishable from G_0—i.e. no non-experimental data can distinguish between G_0 and these other *DAGs*. *DAGs* that are statistically indistinguishable belong to the same *indistinguishability class* [19, pp. 59, 61]. There are several different concepts of indistinguishability and corresponding indistinguishability classes. In the rest of this section, I will discuss *faithful indistinguishability*. Two graphs $G = \langle \mathbb{V}, \mathbb{E} \rangle$ and $G' = \langle \mathbb{V}, \mathbb{E}' \rangle$ are faithfully indistinguishable (f.i.) iff for every $P(\mathbb{V})$, $P(\mathbb{V})$ is Markov and faithful to G iff it is Markov and faithful to G'. Whether or not two *DAGs* are f.i. can be easily verified by the following graphic criterion:

Definition 5 *(Faithful indistinguishability).* Two *DAGs* G and G' are faithfully indistinguishable iff (i) they have the same vertex set $\mathbb{V}$, (ii) they have the same underlying undirected graph: $A - B$ in G iff $A - B$ in G' and (iii) they have the same unshielded colliders: if $A - B - C$ and not $A - C$ in G or in G', then $A \rightarrow B \leftarrow C$ in G iff $A \rightarrow B \leftarrow C$ in G' [19, p. 61].

Faithful indistinguishability classes may be represented by a *pattern* $\Pi = \langle \mathbb{V}, \mathbb{E} \rangle$. A *pattern* is a partially directed graph: $\mathbb{E}$ may contain both directed $(\ldots \rightarrow \ldots)$ and undirected edges $(\ldots - \ldots)$ edges. Each pattern Π represents a set of graphs $Repr(\Pi)$. Whether $G = \langle \mathbb{V}, \mathbb{E} \rangle \in Repr(\Pi)$ may be determined by the following graphic criterion. $G = \langle \mathbb{V}, \mathbb{E} \rangle \in Repr(\Pi)$ iff

1. G and Π have the same adjacency relations;
2. if $A \rightarrow B$ in Π, then $A \rightarrow B$ in G;
3. if $A \rightarrow B \leftarrow C$ and not $A - C$ in G, then $A \rightarrow B \leftarrow C$ and not $A - C$ in Π.

2.3. The graphoid axioms, incompleteness and partial completeness

In this section, I will shortly dwell on some meta-theoretical results regarding the (semi-)graphoid axioms, since these are relevant for the rest of this paper.[12]

A ternary relation $\amalg$ between *disjoint* subsets of $\mathbb{V}$ is a *semi-graphoid* over $\mathbb{V}$ iff it satisfies the following axiom schemata[13] (cf. [20, p. 176], see also [7]):

(G1) $(\mathbf{A} \amalg \mathbf{B} \mid \mathbf{Q}) \supset (\mathbf{B} \amalg \mathbf{A} \mid \mathbf{Q})$ (Symmetry).
(G2) $(\mathbf{A} \amalg \emptyset \mid \mathbf{Q})$ (Trivial Independence).
(G3) $(\mathbf{A} \amalg \mathbf{B} \cup \mathbf{C} \mid \mathbf{Q}) \supset (\mathbf{A} \amalg \mathbf{B} \mid \mathbf{Q})$ (Decomposition).
(G4) $(\mathbf{A} \amalg \mathbf{B} \cup \mathbf{C} \mid \mathbf{Q}) \supset (\mathbf{A} \amalg \mathbf{B} \mid \mathbf{Q} \cup \mathbf{C})$ (Weak Union).

---

[11] The requirement that P is Markov and faithful to some *DAG* G_0 will prove to be very important in Section 3.2.

[12] In fact, the (semi-)graphoid axioms are not axioms, but axiom schemata. But for reasons of readability, I will use 'axiom' and 'axiom schema' interchangeably.

[13] '$\supset$' denotes material implication; '$\wedge$' denotes classical conjunction.

(G5) $((\mathbf{A} \amalg \mathbf{B} \mid \mathbf{Q} \cup \mathbf{C}) \wedge (\mathbf{A} \amalg \mathbf{C} \mid \mathbf{Q})) \supset (\mathbf{A} \amalg \mathbf{B} \cup \mathbf{C} \mid \mathbf{Q})$ (Contraction).

It is a *graphoid* over $\mathbb{V}$ iff it also satisfies the following extra schema:

(G6) $((\mathbf{A} \amalg \mathbf{B} \mid \mathbf{Q} \cup \mathbf{C}) \wedge (\mathbf{A} \amalg \mathbf{C} \mid \mathbf{Q} \cup \mathbf{B})) \supset (\mathbf{A} \amalg \mathbf{B} \cup \mathbf{C} \mid \mathbf{Q})$ (Intersection).

The graphoid axioms are highly relevant for the following three reasons. Firstly, probabilistic conditional independence is a (semi-)graphoid (Theorem 3). This means that the graphoid axioms may be used to derive *IRs* from other *IRs*. Secondly, however, probabilistic conditional independence is not in general completely axiomatizable (Theorem 4). But this does not alter the fact that, thirdly, some interesting subclasses of (semi-)graphoids *are* completely axiomatizable (Theorem 5).

Theorem 3. *For any probability measure P, $\amalg_P$ is a semi-graphoid. Moreover, if P is strictly positive* (*i.e., if $P(A) = 0$ only for $A = \emptyset$*)*, then $\amalg_P$ is a graphoid* [20, p. 176].

So the axioms (G1)–(G5) are *sound* for probabilistic conditional independence and (G6) is sound in case P is strictly positive. By contrast, (G1)–(G6) are *not in general complete* for probabilistic conditional independence. This follows from the following theorem of Milan Studený:

Theorem 4. *There is no finite set of independent axioms which is complete for probabilistic conditional independence. More specifically, there is no finite set of rules of the form* ($r \geqslant 0$):

$$((\mathbf{A}_1 \amalg \mathbf{B}_1 \mid \mathbf{C}_1) \wedge \cdots \wedge (\mathbf{A}_r \amalg \mathbf{B}_r \mid \mathbf{C}_r)) \supset (\mathbf{A}_{r+1} \amalg \mathbf{B}_{r+1} \mid \mathbf{C}_{r+1})$$

such that for any set T of IRs on any set of variables $\mathbb{V}$ there is a probability measure $P(\mathbb{V})$ such that $\amalg_P = \amalg$, i.e. such that

$$(\mathbf{A} \amalg_P \mathbf{B} \mid \mathbf{Q}) \quad \textit{iff} \quad (\mathbf{A} \amalg \mathbf{B} \mid \mathbf{Q}) \in Cl(T)$$

(*where $Cl(T)$ is the closure of T under the given set of rules*) [21].

As I stated above, there are some interesting subclasses of (semi-)graphoids that may be completely characterized. For example:

Theorem 5. *For each semi-graphoid $\amalg$ generated by a list of total causes there is a probability measure P such that $\amalg_P = \amalg$* [20, p. 180].

A *list of total causes* is a set of *IRs* for which there is a linear ordering $X_1, X_2, \ldots$ of all variables in $\mathbb{V}$ such that the list contains for each X_k exactly one statement of the form $(X_k \amalg \{X_1, \ldots, X_{k-1}\} \setminus \mathbf{J} \mid \mathbf{J})$ for some $\mathbf{J} \subseteq \{X_1, \ldots, X_{k-1}\}$. A (semi-)graphoid is *generated by* a list of total causes iff it is the closure of that list under the (semi-)graphoid axioms [20, p. 179].

Let $P^*(\mathbb{V})$ be Markov to the *DAG* $G_0 = \langle \mathbb{V}, \mathbb{E} \rangle$, i.e. $P^*(\mathbb{V})$ is a distribution that verifies, for all $A \in \mathbb{V}$, the *IR* $(A \amalg \mathit{Nondescendants}(A) \setminus \mathit{Parents}(A) \mid \mathit{Parents}(A))$ that results from applying the Markov condition to G_0. The set of all these *IRs* is a list of total causes. Hence, the (semi-)graphoid $\amalg$ generated by this list is completely axiomatized by the (semi-)graphoid axioms and for all P that are Markov and faithful to G_0: $\amalg_P = \amalg$.

3. The IC algorithm for causal discovery

In Section 3.1, I will present the **IC** algorithm. Then I will show to what extent it is impotent regarding the problem of ignorance (Section 3.2) and give a hint at the solution which I will present in the rest of this paper (Section 3.3).

3.1. The algorithm

The **IC** algorithm is based on the relations between *DAGs*, probability distributions and f.i. classes described in Sections 2.1 and 2.2. It runs as follows [18, pp. 50–51]:

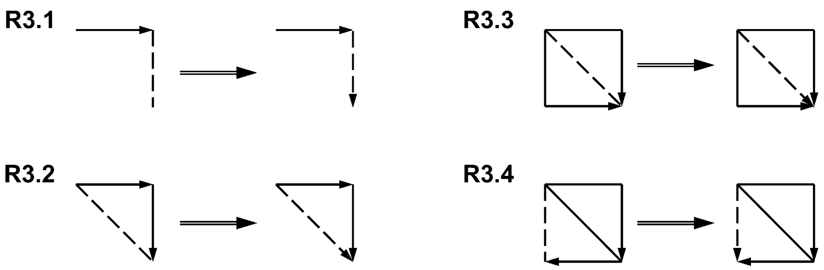

Fig. 1. Rules R3.1–R3.4 According to each of the rules, if a graph contains the left-hand side as a subgraph, an arrow may be added to the dotted line so as to obtain the right-hand side.

Input A probability distribution $P^*(\mathbb{V})$ which is faithful to some underlying *DAG* D_0 (or the list of *IRs* that it verifies).

Output A pattern Π^* representing all *DAGs* that form a complete causal explanation for the *IRs* verified by $P^*(\mathbb{V})$; i.e. all *DAGs* G such that $(\mathbf{A} \amalg_{P^*} \mathbf{B} \mid \mathbf{Q})$ iff $(\mathbf{A} \amalg_G \mathbf{B} \mid \mathbf{Q})$. These *DAGs* form the f.i. class of D_0.

Algorithm The algorithm consists of three consecutive steps:

1. For all $A, B \in \mathbb{V}$ search for a (possibly empty) set $\mathbf{Q} \subseteq \mathbb{V} \setminus \{A, B\}$ such that $(A \amalg_{P^*} B \mid \mathbf{Q})$. Construct an undirected graph G^1 such that $A - B$ iff no such $\mathbf{Q}$ can be found.
2. For all $A, B, C \in \mathbb{V}$ such that $A - C - B$ and not $A - B$ in G^1, check if $C \in \mathbf{Q}$. If it is, then continue. If it is not, then $A \to C \leftarrow B$. The resulting graph is G^2.
3. Starting from G^2, orient as many of the undirected edges as possible subject to two conditions: (i) the orientation should not create a new v-structure; and (ii) the orientation should not create a directed cycle. This is done by closing G^2 under the rules R3.1–R3.4 depicted in Fig. 1.[14]

3.2. *Taking stock of* **IC**

The **IC** algorithm certainly is meritorious as it provides an interesting means to infer causal relations from non-experimental data. However, it is impotent regarding the problem of ignorance. Firstly, its possible inputs are restricted to *full lists* of *IRs*. Secondly, when applied to partial lists of *IRs*, **IC** would lead to unsatisfactory results.

The possible inputs of **IC** are restricted to *full lists* of *IRs*. This follows from the requirement that $P^*(\mathbb{V})$ is Markov and faithful to D_0. The graphoid $\amalg$ generated by applying the Markov condition to D_0 is completely axiomatized by (G1)–(G6). Hence for all $A, B \in \mathbb{V}$ and all $\mathbf{Q} \subseteq \mathbb{V} \setminus \{A, B\}$ it is known whether or not $(A \amalg_{P^*} B \mid \mathbf{Q})$; all *IRs* are decided.

Now suppose that **IC** were applied to a *partial list* of *IRs*. A straightforward solution would be to take a *negation as failure* account: all *CIRs* $(A \amalg B \mid \mathbf{Q})$ and *IRs* $(A \amalg B)$ not occurring in the input (failure) should be taken to be false (negation). This approach, however, has a serious drawback: it treats conditional and unconditional sentences on a par. From a pragmatic point of view, I will argue, this is unsatisfactory. Whereas the negation as failure account is sensible with respect to the former, it certainly isn't with respect to the latter.

The negation as failure account is sensible regarding *conditional independence* (*CIRs*). Suppose that A and B are known to be unconditionally dependent, $\sim (A \amalg B)$, that all tested sentences $(A \amalg B \mid \mathbf{Q})$ were falsified, but that $(A \amalg B \mid \mathbf{Q}^*)$ is still undecided for some $\mathbf{Q}^*$. Even if correlation is no proof for direct causation, it is often regarded as a useful indicator—and so it should be, otherwise non-experimental causal research would be a rather idle enterprise (cf. the importance attached to epidemiological evidence by the IARC [10]). This requirement may satisfactorily be met by considering all undecided $(A \amalg B \mid \mathbf{Q}^*)$ as false (negation as failure), *provided faulty applications of this heuristic can be detected and remedied quickly*.

By contrast, the negation as failure account is unsatisfactory for *unconditional independencies* (*UIRs*). Large part of scientific practice consists in finding models that are simple enough to be manageable and useful for prediction, explanation and/or intervention. Models are simulacra that share some, but not all the characteristics of the phenomena

[14] The rules R3.1–R3.4 are necessary [23] and sufficient [12] for obtaining the pattern Π^* representing the intended equivalence class.

under study [6]. They are the mediators between a theory and reality [14]. As such, they do not perfectly mirror this world. Abstraction and idealization are non-negligible aspects of scientific modelling. Moreover, they are already implicitly present in Pearl's framework. The Markov and the Faithfulness condition together imply the Minimality condition [19, p. 31].[15] If $G = \langle \mathbb{V}, \mathbb{E} \rangle$ and $P(\mathbb{V})$ satisfy the Minimality condition, then each edge in G prevents some conditional independence that would otherwise obtain; apart from that, G does not contain any superfluous edges [19, p. 12]. However, if all undecided $(A \amalg B)$-sentences are taken to be false (negation as failure), the resulting model would be gratuitously complex. In scientific practice, if $(A \amalg B)$ is undecided, then it is frequently or even mostly the case that for all $\mathbf{Q}^*$, $(A \amalg B \mid \mathbf{Q}^*)$ is undecided too. Hence, in view of the previous paragraph, negation as failure would allow to infer that $A - B$. Therefore negation as failure should not be applied to undecided *UIRs*. These should be considered as true, *provided faulty applications of this heuristic can be detected and remedied quickly*.

Before I give a short introduction to adaptive logics in Section 3.3, I will first pursue some possible epistemological worries. My suggestions may come across as too rash to some readers. Why should we rely on default assumptions which will most probably be violated in many contexts? And what would be the consequences of any such violation? The following three remarks should help to remove such doubts. Firstly, in scientific practice defeasible assumptions are abundantly used for causal inference. In his short but highly influential paper, Sir Austin Bradford Hill [9, p. 295] addresses the question how to pass from an observed association to a verdict of causation in occupational medicine, when no general body of medical knowledge provides a decisive answer. After listing nine viewpoints from which to investigate the association in question, he argues that no decisive criterion (or set of criteria) exists. Hence, the inference from association to causation is defeasible. Nevertheless, we are pragmatically forced to judge (whether the association is causal or not, whether the assumed cause is suitable for intervention or not), given that we have to take action:[16]

> All scientific work is incomplete—whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time [9, p. 300].

Secondly, the use of default assumptions is not new in the literature on causal inference. For example, Williamson [24, Chapters 5, 6] proposes to use the Maximum Entropy principle (*MaxEnt*) to handle the problem of ignorance. Suppose we are given a partial list of *IRs*. Then many different probability functions will satisfy this list. If the set of these functions is closed and convex, *MaxEnt* selects the single member which is maximally non-committal with regard to missing information. (The *MaxEnt* principle generalizes the Principle of Indifference; see also [16,17].) As the selected probability function will mostly be at odds with any new information, the suggested mechanism is defeasible and non-monotonic (*cf. infra*, Section 5.6). Thirdly, the adaptive logic framework provides us with a dynamic proof theory which allows us (i) to trace the particular assumptions on which each inference is based, and (ii) to trace the consequences of the violation of each particular assumption (i.e. faulty applications of each heuristic can be detected and remedied quickly). As such, the dynamic proof theory allows us to cautiously apply these default assumptions. This will become clear in Section 5.4, where I will present the proof theory of **ALIC** and discuss its epistemological implications.

3.3. Towards an adaptive logic solving the problem of ignorance

In the following sections, I will show how the problem of ignorance may be solved. I will develop an adaptive logic for causal discovery that properly gives shape to the findings of the previous section.

Using an adaptive logic (instead of, for example, some other default logic) has several advantages. Firstly, the standard format of adaptive logics provides a unified framework for handling various non-monotonic consequence relations (defaults, inconsistency-handling mechanisms, ...).[17] Many defeasible consequence relations have success-

[15] $G = \langle \mathbb{V}, \mathbb{E} \rangle$ and $P(\mathbb{V})$ satisfy the Minimality condition if and only if for every $G' = \langle \mathbb{V}, \mathbb{E}' \rangle$ such that $\mathbb{E}' \subset \mathbb{E}$ (i.e. for every subgraph G' of G), $G' = \langle \mathbb{V}, \mathbb{E}' \rangle$ and $P(\mathbb{V})$ do not satisfy the Markov Condition [19, p. 12].

[16] I do not claim that Hill argued in favour of the defaults incorporated in the adaptive logic **ALIC**, only that he argued in favor of defeasible reasoning.

[17] Given this flexibility, it would be fairly easy to devise alternatives to **ALIC**, based on different rationales, to address the problem of ignorance.

fully been translated into the adaptive logic framework (e.g. [1,13]),[18] whereas no other approach is known to have such unifying power. This unified framework makes it possible to easily compare such consequence relations. Moreover, contrary to most other non-monotonic logics (cf. [11]), adaptive logics provide a good proof theory that captures the dynamics involved in non-monotonic reasoning. Finally, as I will argue in Section 5.4, **ALIC**'s dynamic proof theory gives rise to a pragmatic picture of causal inference in which proofs may act as a guide for both scientific research and policy.

Adaptive logics are non-monotonic logics [2–4]. In general, they are characterized by a triple $\mathbf{AL} = \langle \mathbf{LLL}, \Omega,$ adaptive strategy$\rangle$. **LLL** is the *lower limit logic* of **AL**. It is the stable part of **AL**. Semantically, the adaptive models of Γ are a subset of its **LLL**-models (Γ being a premise set). Proof theoretically, all the axioms and inference rules of **LLL** may be applied unconditionally. Hence, if $\Gamma \vdash_{\mathbf{LLL}} \alpha$, then $\Gamma \vdash_{\mathbf{AL}} \alpha$. By contrast, the other axioms or inference rules of **AL** are conditional: they may be applied on the condition that certain other formulas (certain abnormalities) are not derivable. Ω is the set of *abnormalities*. These are formulas that are characterized by a (possibly restricted) logical form and that are presupposed to be false, unless and until proven otherwise. Intuitively speaking, an abnormality $\omega \in \Omega$ is presupposed to be false relative to Γ, unless and until it turns out that Γ forces you to give up this presupposition.[19]

Together, **LLL** and Ω define an upper limit logic **ULL**. Proof theoretically, the **ULL** is obtained by adding to **LLL** an axiom or inference rule that connects abnormality to triviality. Semantically, the **ULL**-models are obtained by selecting those **LLL**-models that verify no abnormality. The following theorem reveals a crucial relation between **LLL**, Ω and **ULL**. ($Dab(\Delta)$ denotes a disjunction of the members of Δ.)

Theorem 6 *(Derivability Adjustment Theorem). $\Gamma \vdash_{\mathbf{ULL}} \alpha$ iff there is a finite $\Delta \subseteq \Omega$ such that $\Gamma \vdash_{\mathbf{LLL}} \alpha \vee Dab(\Delta)$.*

Proof. See Batens [4, pp. 230–231]. □

In general, the following proof theoretic relations hold between **AL**, **LLL**, and **ULL**: If Γ is normal (i.e. if no *Dab*-formulas are **LLL**-derivable from Γ), then

$$Cn_{\mathbf{LLL}}(\Gamma) \subset Cn_{\mathbf{AL}}(\Gamma) = Cn_{\mathbf{ULL}}(\Gamma)$$

If Γ is abnormal, then

$$Cn_{\mathbf{LLL}}(\Gamma) \subset Cn_{\mathbf{AL}}(\Gamma) \subset Cn_{\mathbf{ULL}}(\Gamma) = \text{the set of all formulas}$$

One of the best known adaptive strategies is *reliability*. It determines how to treat minimal *Dab*-consequences. $Dab(\Delta)$ is a minimal *Dab*-consequence of Γ iff $\Gamma \vdash_{\mathbf{LLL}} Dab(\Delta)$ and there is no $\Delta^* \subset \Delta$ such that $\Gamma \vdash_{\mathbf{LLL}} Dab(\Delta^*)$. According to *reliability*, a formula is unreliable relative to Γ if it is a disjunct of a *minimal Dab-consequence* of Γ.[20]

4. LIC: the lower limit logic of ALIC

In this section I will present **LIC**, a non-adaptive logic for causal discovery. By itself, **LIC** does not add much to **IC**, but its significance derives from the role it plays in the formulation of **ALIC**. I will describe both its language (Section 4.1), its semantics (Section 4.2) and its proof theory (Section 4.3). I will conclude this section by briefly discussing soundness and (in)completeness for **LIC** (Section 4.4).

[18] Other results can be found at http://logica.UGent.be/centrum/writings/.

[19] As I will show in Section 5.4, the proof theory of adaptive logics is such that it makes sense to write "unless *and until* it turns out that Γ forces you to give up this presupposition".

[20] Another well-known strategy is *minimal abnormality*. The choice of strategy affects both the proof theory and the semantics of the adaptive logic. In this paper, I will stick to reliability.

4.1. *The language of* **LIC**

Let $\mathbb{V}$ be a *finite set of finite random variables*. I will assume that all variables in $\mathbb{V}$ are different (different name = different variable) and that they are logically independent.[21] Let $\mathcal{W}^f$ be the set of *factual propositions*, i.e. the smallest set satisfying the following conditions:

(L1) If $X \in \mathbb{V}$ and x_i is a value of X, then $X = x_i \in \mathcal{S}$.
(L2) If $\alpha \in \mathcal{S}$, then $\alpha \in \mathcal{W}^f$.
(L3) If $\alpha, \beta \in \mathcal{W}^f$, then $\sim\alpha$, $(\alpha \wedge \beta)$, $(\alpha \vee \beta) \in \mathcal{W}^f$.

The elements of $\mathbb{V}$ and of $\mathcal{W}^f$ by themselves do not belong to the language of **LIC**, but they serve as the basis for the formulation of the latter. The atomic sentences of **LIC** are either *probabilistic* ($\mathcal{W}^p$) or *causal* ($\mathcal{W}^c$). Complex sentences ($\mathcal{W}$) are built from atomic ones by means of classical connectives. So $\mathcal{W}^p$, $\mathcal{W}^c$ and $\mathcal{W}$ are the smallest sets satisfying the following conditions:

(L4) If $\mathbf{A}, \mathbf{B}, \mathbf{Q} \subset \mathbb{V}$ are disjoint sets of variables, then $(\mathbf{A} \amalg \mathbf{B})$, $(\mathbf{A} \amalg \mathbf{B} \mid \mathbf{Q}) \in \mathcal{W}^p$.

Convention 2. If $A, B \in \mathbb{V}$ and $\mathbf{Q} \subseteq \mathbb{V} \setminus \{A, B\}$, then I will write $(A \amalg B)$ and $(A \amalg B \mid \mathbf{Q})$ instead of $(\{A\} \amalg \{B\})$ and $(\{A\} \amalg \{B\} \mid \mathbf{Q})$. Likewise, if $\mathbf{Q}$ is a singleton $\{Q\}$, I will write Q instead of $\{Q\}$.

(L5) If A and $B \in \mathbb{V}$, then $A \to B$, $A - B$, and $A \Rightarrow B \in \mathcal{W}^c$.
(L6) If $\alpha \in \mathcal{W}^p \cup \mathcal{W}^c$, then $\alpha \in \mathcal{W}$.
(L7) If $\alpha \in \mathcal{W}$, then $\sim\alpha \in \mathcal{W}$.
(L8) If $\alpha, \beta \in \mathcal{W}$, then $(\alpha \wedge \beta)$, $(\alpha \vee \beta)$, $(\alpha \supset \beta)$, $(\alpha \equiv \beta)$, $(\alpha \veebar \beta) \in \mathcal{W}$.[22]

4.2. *The semantics of* **LIC**

The semantics of **LIC** should meet certain obvious requisites. For one thing, each model should assign appropriate truth values to the atomic probabilistic sentences, to the atomic causal sentences and to the classical complex sentences (see (Sv1)–(Sv12)). Moreover, no model should allow for cyclicity (see (Sc.3)). Finally, the probabilistic sentences verified by each model should be Markov and Faithful to the causal sentences it verifies (see (Sp9)).

To make sure that the semantics hereunder is recursive, I need to introduce an ordering relation $\prec$ over $\mathbb{V}$. This ordering may be done by lexicographic order, or by the Gödel numbers of the variables. Hence it is not the case that $A \prec B$ iff A is an ancestor of B in some *DAG*.

- There is no $A \in \mathbb{V}$ such that $A \prec A$.
- For all $A, B \in \mathbb{V}$: either $A \prec B$, or $B \prec A$, but not both.
- For all $A, B, C \in \mathbb{V}$: $A \prec B$ and $B \prec C$ implies $A \prec C$.

An **LIC**-model is a triple $M = \langle \mathbb{R}^+, \mathsf{c}, \mathsf{p} \rangle$. $\mathbb{R}^+$ is the set of nonnegative real numbers (including 0). c is a function which determines the causal relations holding between members of $\mathbb{V}$. p is a probability distribution over $\mathbb{V}$.

c is a partial function that maps couples of variables $\langle A, B \rangle$ to the set $\{l, r, n\}$, subject to the following conditions:

(Sc1) $\mathsf{c} : \mathbb{V} \times \mathbb{V} \to \{l, r, n\}$.
(Sc2) $\mathsf{c}(\langle A, B \rangle)$ is only defined for $A \prec B$.
(Sc3) There is no $\{X_1, \ldots, X_n\} \subseteq \mathbb{V}$ such that either $\mathsf{c}(\langle X_i, X_{i+1} \rangle) = l$ for all $1 \leqslant i \leqslant n-1$ and $\mathsf{c}(\langle X_1, X_n \rangle) = r$, or $\mathsf{c}(\langle X_i, X_{i+1} \rangle) = r$ for all $1 \leqslant i \leqslant n-1$ and $\mathsf{c}(\langle X_1, X_n \rangle) = l$[23] (acyclicity).

[21] So $\mathbb{V}$ cannot contain both G, ‘gender’, and K, ‘being a king’, since by definition, all kings are male.
[22] The last logical connective, $\veebar$, is the exclusive disjunction, cf. (Sv12). Although the exclusive disjunction can be easily omitted, I include it so as to present the **LIC**-axioms in an intuitive way.
[23] Note that I use the shorthand notation for the values of c throughout this section (cf. Convention 3).

In the interest of the readability of the metatheorems, however, I introduce the following convention:

Convention 3. "$\mathsf{c}(\langle A, B\rangle) = l$" should be read as: "If $A \prec B$, then $\mathsf{c}(\langle A, B\rangle) = l$. If $B \prec A$, then $\mathsf{c}(\langle B, A\rangle) = r$." In the same manner, I will use "$\mathsf{c}(\langle A, B\rangle) \neq n$" as an abbreviation for "If $A \prec B$, then $\mathsf{c}(\langle A, B\rangle) \neq n$. If $B \prec A$, then $\mathsf{c}(\langle B, A\rangle) \neq n$." etc.

p is a function assigning a 'weight' to sentences in $\mathcal{W}^f$. It is defined as the composition of two other functions, m and o. m assigns a weight to the elements of a set $\mathfrak{S}$. o maps factual propositions to subsets of $\mathfrak{S}$.

(Sp1) Let $\mathfrak{S}$ be a set with at least the cardinality of the sample space defined by $\mathbb{V}$.
(Sp2) $\mathsf{m} : \mathfrak{S} \to \mathbb{R}^+$.
(Sp3) $\sum_{\alpha\in\mathfrak{S}} \mathsf{m}(\alpha) = 1$.
(Sp4) For all $\beta \subseteq \mathfrak{S} : \mathsf{m}(\beta) = \sum_{\alpha\in\beta} \mathsf{m}(\alpha)$.
(Sp5) For all $\beta \subseteq \mathfrak{S} : \mathsf{m}(\beta^c) = 1 - \mathsf{m}(\beta)$.
(Sp6) For all $\beta, \gamma \subseteq \mathfrak{S} : \mathsf{m}(\beta \cap \gamma) = \mathsf{m}(\beta) + \mathsf{m}(\gamma) - \mathsf{m}(\beta \cup \gamma)$.
(Sp7) $\mathsf{o} : \mathcal{S} \to \wp(\mathfrak{S})$, where $\mathcal{S}$ is the set of atomic factual propositions $X = x_i$.

o can be extended to a function mapping all factual propositions to subsets of $\mathfrak{S}$: $\mathsf{o} : \mathcal{W}^f \to \wp(\mathfrak{S})$, with

- $\mathsf{o}(\sim\alpha) = (\mathsf{o}(\alpha))^c$ for all $\alpha \in \mathcal{W}^f$,
- $\mathsf{o}(\alpha \wedge \beta) = \mathsf{o}(\alpha) \cap \mathsf{o}(\beta)$ for all $\alpha, \beta \in \mathcal{W}^f$, and
- $\mathsf{o}(\alpha \vee \beta) = \mathsf{o}(\alpha) \cup \mathsf{o}(\beta)$, for all $\alpha, \beta \in \mathcal{W}^f$.

p is composed of m and o. Moreover, by condition (Sp9), p is Markov and faithful to c.

(Sp8) $\mathsf{p}(x) = \mathsf{m}(\mathsf{o}(x))$.
(Sp9) If $A, B \notin \mathbf{Q}$, then $(A \amalg B \mid \mathbf{Q})$ iff every undirected path between A and B is d-separated by $\mathbf{Q}$. More precisely, $\frac{\mathsf{p}(A=a\wedge(B=b\wedge\mathbf{Q}=\mathbf{q}))}{\mathsf{p}(B=b\wedge\mathbf{Q}=\mathbf{q})} = \frac{\mathsf{p}(A=a\wedge\mathbf{Q}=\mathbf{q})}{\mathsf{p}(\mathbf{Q}=\mathbf{q})}$ for all values a of A, b of B and $\mathbf{q}$ of $\mathbf{Q}$ (whenever $\mathsf{p}(B = b \wedge \mathbf{Q} = \mathbf{q}) > 0$) iff
  0.1. $\mathsf{c}(\langle A, B\rangle) = n$, and
  0.2. for any n-tuple $\langle X_1, \ldots, X_n\rangle$ such that (i) $n \geqslant 3$, (ii) $X_1 = A$, (iii) $X_n = B$, and (iv) $\mathsf{c}(\langle X_i, X_{i+1}\rangle) \neq n$ for all $1 \leqslant i \leqslant n-1$, at least one of the following conditions is satisfied:
    (a) There is some $1 \leqslant i \leqslant n-2$ such that $\mathsf{c}(\langle X_i, X_{i+1}\rangle) = \mathsf{c}(\langle X_{i+1}, X_{i+2}\rangle)$ ($= r$ or $= l$) and $X_{i+1} \in \mathbf{Q}$.
    (b) There is some $1 \leqslant i \leqslant n-2$ such that $\mathsf{c}(\langle X_i, X_{i+1}\rangle) = l$, and $\mathsf{c}(\langle X_{i+1}, X_{i+2}\rangle) = r$, and $X_{i+1} \in \mathbf{Q}$.
    (c) There is some $1 \leqslant i \leqslant n-2$ such that $\mathsf{c}(\langle X_i, X_{i+1}\rangle) = r$, and $\mathsf{c}(\langle X_{i+1}, X_{i+2}\rangle) = l$, and $X_{i+1} \notin \mathbf{Q}$ and for all X such that there is an m-tuple $\langle Y_1, \ldots, Y_m\rangle$ such that $Y_1 = X_{i+1}$, and $Y_m = X$, and for all $1 \leqslant i \leqslant m-1$, $\mathsf{c}(\langle Y_i, Y_{i+1}\rangle) = r$, $X \notin \mathbf{Q}$.

A valuation function v_M determined by a model $M = \langle\mathbb{R}^+, \mathsf{c}, \mathsf{p}\rangle$ is a function that satisfies the following conditions:

(Sv1) $\mathsf{v}_M : \mathcal{W} \to \{0, 1\}$.
(Sv2) $\mathsf{v}_M((\mathbf{A} \amalg \mathbf{B})) = 1$ iff $\frac{\mathsf{p}(\mathbf{A}=\mathbf{a}\wedge\mathbf{B}=\mathbf{b})}{\mathsf{p}(\mathbf{B}=\mathbf{b})} = \mathsf{p}(\mathbf{A} = \mathbf{a})$ for all values $\mathbf{a}$ of $\mathbf{A}$ and $\mathbf{b}$ of $\mathbf{B}$ (whenever $\mathsf{p}(\mathbf{B} = \mathbf{b}) > 0$).
(Sv3) $\mathsf{v}_M((\mathbf{A} \amalg \mathbf{B} \mid \mathbf{Q})) = 1$ iff $\frac{\mathsf{p}(\mathbf{A}=\mathbf{a}\wedge(\mathbf{B}=\mathbf{b}\wedge\mathbf{Q}=\mathbf{q}))}{\mathsf{p}(\mathbf{B}=\mathbf{b}\wedge\mathbf{Q}=\mathbf{q})} = \frac{\mathsf{p}(\mathbf{A}=\mathbf{a}\wedge\mathbf{Q}=\mathbf{q})}{\mathsf{p}(\mathbf{Q}=\mathbf{q})}$ for all values $\mathbf{a}$ of $\mathbf{A}$, $\mathbf{b}$ of $\mathbf{B}$ and $\mathbf{q}$ of $\mathbf{Q}$ (whenever $\mathsf{p}(\mathbf{B} = \mathbf{b} \wedge \mathbf{Q} = \mathbf{q}) > 0$).
(Sv4) $\mathsf{v}_M(A \to B) = 1$ iff $\mathsf{c}(\langle A, B\rangle) = r$.
(Sv5) $\mathsf{v}_M(A - B) = 1$ iff $\mathsf{c}(\langle A, B\rangle) \neq n$.
(Sv6) $\mathsf{v}_M(A \Rightarrow B) = 1$ iff there is a series of variables $C_1, \ldots, C_m \in \mathbb{V}$ ($m \geqslant 2$) such that $A = C_1$, $B = C_m$, and for all $1 \leqslant i \leqslant m-1$: $\mathsf{c}(\langle C_i, C_{i+1}\rangle) = r$.
(Sv7) $\mathsf{v}_M(\sim\alpha) = 1$ iff $\mathsf{v}_M(\alpha) = 0$.
(Sv8) $\mathsf{v}_M(\alpha \wedge \beta) = 1$ iff $\mathsf{v}_M(\alpha) = \mathsf{v}_M(\beta) = 1$.
(Sv9) $\mathsf{v}_M(\alpha \vee \beta) = 1$ iff $\mathsf{v}_M(\alpha) = 1$ or $\mathsf{v}_M(\beta) = 1$.

(Sv10) $\mathsf{v}_M(\alpha \supset \beta) = 1$ iff $\mathsf{v}_M(\alpha) = 0$ or $\mathsf{v}_M(\beta) = 1$.
(Sv11) $\mathsf{v}_M(\alpha \equiv \beta) = 1$ iff $\mathsf{v}_M(\alpha) = \mathsf{v}_M(\beta)$.
(Sv12) $\mathsf{v}_M(\alpha \veebar \beta) = 1$ iff $\mathsf{v}_M(\alpha) \neq \mathsf{v}_M(\beta)$.

Definition 6 *(Truth in a model).* α is true in a model $M = \langle \mathbb{R}^+, \mathsf{c}, \mathsf{p} \rangle$ (abbreviated as $M \models \alpha$) $=_{df}$ $\mathsf{v}_M(\alpha) = 1$.

4.3. The proof theory of **LIC**

The proof theory of **LIC** should meet the following obvious requirements. Firstly, it should determine the behavior of the classical connectives $\sim, \wedge, \ldots$, of the $\amalg$-relation and of the causal relations $\to$, $-$ and $\Rightarrow$. Secondly, it should mimic the **IC** algorithm. These requirements are met by the following axiom schemata and inference rules:

(R1) $\alpha, \alpha \supset \beta / \beta$.
(A1) The axiom schemata of propositional classical logic.
(A2) The (semi-)graphoid axiom schemata.
(A3) $\sim (A \Rightarrow A)$.
(A4) $A - B \equiv (A \to B \vee B \to A)$.
(A5) $A \Rightarrow B \equiv (A \to B \vee \bigvee\{(A \Rightarrow C \wedge C \to B) \mid C \in \mathbb{V} \setminus \{A, B\}\})$.
(A6) $A - B \veebar \bigvee\{(A \amalg B \mid \mathbf{Q}) \mid \mathbf{Q} \subseteq \mathbb{V} \setminus \{A, B\}\}$.
(R2) $A - C, C - B, \sim (A - B) \;/\; (A \to C \wedge B \to C) \equiv \bigvee\{(A \amalg B \mid \mathbf{Q}) \wedge \sim (A \amalg B \mid \mathbf{Q} \cup \{C\}) \mid \mathbf{Q} \subset \mathbb{V} \setminus \{A, B\}\}$.

That the **IC** algorithm is adequately mimicked, can be easily ascertained. The steps in the algorithm are based on a few basic premises. Firstly, no directed cyclic paths are allowed. This is mimicked by (A3) and (A5). Secondly, two variables are adjacent if *and only if* no disjoint set screens them off. This is mimicked by (A6). Thirdly, three variables A, B, C form a v-structure (i.e. $A \to C \leftarrow B$ and not $A - B$) if *and only if* there is a $\mathbf{Q} \subset \mathbb{V} \setminus \{A, B\}$ such that $(A \amalg B \mid \mathbf{Q})$ and not $(A \amalg B \mid \mathbf{Q} \cup \{C\})$.

4.4. Soundness, but no completeness for **LIC**

The inference rules and axiom schemata listed in Section 4.3 are not complete with respect to the semantics of Section 4.2. This follows from Theorem 4, given that I will not restrict the possible sets of premises to those sets of *IRs* for which the graphoid axioms are complete. Consider a set T of *IRs* for which there is no P such that $(A \amalg_P B \mid \mathbf{Q})$ iff $(A \amalg B \mid \mathbf{Q}) \in Cl(T)$ (where $Cl(T)$ is the closure of T under the graphoid axioms). From Theorem 4 it follows that such a T exists. Hence each model $M = \langle \mathbb{R}^+, \mathsf{c}, \mathsf{p} \rangle$ of T verifies at least one $IR \notin Cl(T)$. So all models verify the disjunction of these additional *IRs*, but there is no guarantee that their disjunction is **LIC**-derivable from T.

Conversely, the rules and axioms in Section 4.3 are sound with respect to the semantics in Section 4.2.

Theorem 7 *(Soundness for* **LIC***). If $\Gamma \vdash \alpha$, then $\Gamma \models \alpha$.*

Proof. See Appendix A. □

5. ALIC: the adaptive logic for causal discovery

Regarding the problem of ignorance, **LIC**, the non-adaptive logic presented in Section 4, is almost on a par with the **IC** algorithm. In case the input or premise set Γ consists of a full list of *IRs*, neither **LIC** nor **IC** will run into difficulty and they will produce the same causal output or consequence set: $A - B \in Cn_{\mathbf{LIC}}(\Gamma)$ iff $A - B \in \Pi(\Gamma)$ or $A \to B \in \Pi(\Gamma)$ or $B \to A \in \Pi(\Gamma)$ and $A \to B \in Cn_{\mathbf{LIC}}(\Gamma)$ iff $A \to B \in \Pi(\Gamma)$. Likewise, neither **LIC** nor **IC** will lead to satisfactory results in case the input or premise set Γ consists of a partial list of *IRs*.

As I said, **LIC** is *almost* on a par with **IC**. Since any set of **LIC**-wffs (i.e. any $\Gamma \subset \mathcal{W}$) may serve as a premise set, combining observational knowledge with background knowledge, micro-level knowledge, common sense knowledge, etc. poses no technical problems. For example, if background knowledge shows that A causes B, or if common sense dictates that C cannot cause D, since it succeeds D in time, it is straightforward to include $A \to B$ or $\sim (C \to D)$ in the premise set.[24]

In this section, I will present **ALIC**, the adaptive logic for causal discovery which properly solves the problem of ignorance. In Section 3.2, I have outlined the basic heuristics governing **ALIC**. In the face of an undecided *CIR*, $(A \amalg B \mid \mathbf{Q})$, we should conceive of it as false, *provided faulty applications of this heuristic can be detected and remedied quickly*. If not, non-experimental causal research would be a rather idle enterprise. In other words, *CIRs* should be presupposed to be false, unless and until proven otherwise. This heuristic does not apply to *UIRs*. In the face of an undecided *UIR*, $(A \amalg B)$, we should conceive of it as true, *provided faulty applications of this heuristic can be detected and remedied quickly*. If not, our scientific models would be gratuitously complex. So *UIRs* should be presupposed to be true, unless and until proven otherwise. These heuristics will be formalized singly in Sections 5.1 and 5.2, where I will present two auxiliary logics $\mathbf{ALIC^{RS}}$ and $\mathbf{ALIC^{At}}$. Then I will combine both auxiliary logics (and their respective heuristics) in one single logic: **ALIC**. First I will describe its proof theory (Section 5.4). Then I will describe its semantics and some straightforward meta-theoretic results (Section 5.5). Finally, I will briefly discuss the relation between **ALIC** and *MaxEnt*-based causal discovery (Section 5.6).

5.1. $\mathbf{ALIC^{RS}}$: *the Reckless Statistician's account of CIRs*

The heuristic regarding conditional independence statements may metaphorically be called the heuristic of the Reckless Statistician. If all *CIRs* $(A \amalg B \mid \mathbf{Q})$ are undecided, and if this heuristic is added to **LIC**, then one may infer causation, $A - B$, from correlation, $\sim (A \amalg B)$, contrary to one of the best-known warnings in introductory statistics courses. Any statistician that would apply this heuristic blindly, would deservedly be called reckless. But as I will show in the following sections, the adaptive logical framework affords a way to apply it properly. The Reckless Statistician is tempered.[25]

The heuristic of the Reckless Statistician naturally leads to the following adaptive logic: $\mathbf{ALIC^{RS}} = \langle \mathbf{LIC}, \Omega^{\mathbf{RS}},$ *reliability*$\rangle$. The set of abnormalities $\Omega^{\mathbf{RS}}$ contains all *CIRs* regarding pairs of variables (not pairs of sets of variables).

$$\Omega^{\mathbf{RS}} = \{(A \amalg B \mid \mathbf{Q}) \mid A, B \in \mathbb{V}, \mathbf{Q} \subseteq \mathbb{V} \setminus \{A, B\}\}$$

By adding axiom schema (A7) to **LIC**, a suited upper limit logic $\mathbf{ULL^{RS}}$ is obtained.[26]

(A7) $\sim (A \amalg B \mid \mathbf{Q})$.

The $\mathbf{ULL^{RS}}$-semantics consists of the **LIC**-models for which c is such that

if $\mathsf{c}(\langle A, B\rangle) = \mathsf{c}(\langle B, C\rangle) = r$, then $\mathsf{c}(\langle A, C\rangle) = r$,
if $\mathsf{c}(\langle A, B\rangle) = \mathsf{c}(\langle B, C\rangle) = l$, then $\mathsf{c}(\langle A, C\rangle) = l$,
if $\mathsf{c}(\langle A, B\rangle) = l$ and $\mathsf{c}(\langle B, C\rangle) = r$, then $\mathsf{c}(\langle A, C\rangle) \neq n$.

I will not discuss the proof theory, nor the semantics of $\mathbf{ALIC^{RS}}$ as these are not significant for the present paper and moreover are really straightforward (see [4, Sections 3 and 4]).

5.2. $\mathbf{ALIC^{At}}$: *the Atomist's account of UIRs*

Where the first heuristic could be called the heuristic of the Reckless Statistician, the second one, regarding *UIRs*, may metaphorically be called the heuristic of the Atomist. If all *UIRs* $(A \amalg B)$ are undecided, and if this heuristic is

[24] In [12] a framework is presented which also allows to combine the **IC**-algorithm with background knowledge. However, this framework cannot deal with the problem of ignorance. Moreover, it cannot deal with background knowledge consisting of complex formulas (in [12], background knowledge consists of a pair $\mathcal{K} = \langle \mathbf{F}, \mathbf{R}\rangle$ in which $\mathbf{F}$ is the set of directed edges which are forbidden and $\mathbf{R}$ is the set of directed edges which are required).

[25] The same intuition, viz. that $\sim (A \amalg B)$ implies $A - B$, unless and until some set $\mathbf{Q}$ is found such that $(A \amalg B \mid \mathbf{Q})$, lays at the basis of the logic presented in [22].

[26] The reader can easily check that Theorem 6 holds for **LIC**, $\Omega^{\mathbf{RS}}$ and $\mathbf{ULL^{RS}}$.

added to **LIC**, then one may infer that no causal relations whatsoever exist; i.e., $\sim(A - B)$ for all $A, B \in \mathbb{V}$. So the heuristic of the Atomist results in models that verify as little causal relations as possible.[27]

All this naturally leads to the following adaptive logic: $\mathbf{ALIC^{At}} = \langle \mathbf{LIC}, \Omega^{\mathbf{At}}, \textit{reliability} \rangle$, where $\Omega^{\mathbf{At}}$ contains all negations of *UIRs* regarding pairs of variables.

$$\Omega^{\mathbf{At}} = \{\sim(A \amalg B) \mid A, B \in \mathbb{V}\}$$

By adding axiom schema (A8) to **LIC**, a suited upper limit logic $\mathbf{ULL^{At}}$ is obtained.[28] The $\mathbf{ULL^{At}}$-semantics consists of the **LIC**-models for which $\mathsf{c}(\langle A, B \rangle) = n$ for all $A, B \in \mathbb{V}$.

(A8) $(A \amalg B)$.

Again, I will not further discuss the proof theory or the semantics of $\mathbf{ALIC^{At}}$.

5.3. **ALIC**: *outline*

ALIC is the result of combining $\mathbf{ALIC^{RS}}$ and $\mathbf{ALIC^{At}}$. Its lower limit logic is **LIC** and its strategy is *reliability*. Its set of abnormalities is the union of $\Omega^{\mathbf{RS}}$ and $\Omega^{\mathbf{At}}$:

$$\mathbf{ALIC} = \langle \mathbf{LIC}, \Omega^{\mathbf{RS}} \cup \Omega^{\mathbf{At}}, \textit{reliability} \rangle$$

As I will show in Section 5.4, the proof theory of **ALIC** is dynamic. Lines of a proof may be marked (but also unmarked) as the proof continues. Formulas occurring on marked lines are not considered as derived. Notwithstanding this dynamics, the set of **ALIC**-consequences of some premise set Γ is fixed: $Cn_{\mathbf{ALIC}}(\Gamma) = Cn_{\mathbf{ALIC^{RS}}}(Cn_{\mathbf{ALIC^{At}}}(\Gamma))$.[29] This stability is reflected in the **ALIC**-semantics (see Section 5.5).

5.4. *The proof theory of* **ALIC**

ALIC-proofs are dynamic. Lines in a dynamic proof consist of five elements: (i) a line number k, (ii) a formula α, (iii) the line numbers of the formulas from which α is derived, (iv) the rule by which α is derived, and (v) a *condition* Υ.[30] The condition is a (possibly empty) set of abnormalities. It determines whether α is derived or not. Intuitively, if all members of Υ may be considered as false, then α is derived on line k. Otherwise, line k is marked and α is no longer considered as derived.

The proof theory of **ALIC** consists of three generic deduction rules, and a marking definition. The deduction rules allow one to add a line to the proof. By adding a line, the proof is brought to a next *stage*. The marking definitions determine, at each stage s of the proof, which lines are marked and which are unmarked.

The generic deduction rules are as follows (Γ is a premise set):

PREM If $\alpha \in \Gamma$, one may add a line comprising of the following elements: (i) an appropriate line number, (ii) α, (iii) –, (iv) PREM, and (v) $\emptyset$.

RU If $\beta_1, \ldots, \beta_n \vdash_{\mathbf{LIC}} \alpha$ and each of the β_i occurs in the proof on lines $i_1, \ldots, i_n$ that have conditions $\Upsilon_1, \ldots, \Upsilon_n$ respectively, one may add a line comprising of the following elements: (i) an appropriate line number, (ii) α, (iii) $i_1, \ldots, i_n$, (iv) RU, and (v) $\Upsilon_1 \cup \cdots \cup \Upsilon_n$.

RC If $\beta_1, \ldots, \beta_n \vdash_{\mathbf{LIC}} \alpha \vee Dab(\Theta)$ (for some $\Theta \subseteq \Omega^{\mathbf{RS}} \cup \Omega^{\mathbf{At}}$) and each of the β_i occurs in the proof on lines $i_1, \ldots, i_n$ that have conditions $\Upsilon_1, \ldots, \Upsilon_n$ respectively, one may add a line comprising of the following elements: (i) an appropriate line number, (ii) α, (iii) $i_1, \ldots, i_n$, (iv) RC, and (v) $\Upsilon_1 \cup \cdots \cup \Upsilon_n \cup \Theta$.

---

27 An even better label would have been ‘the Greedy Statistician’, but ‘the Atomist’ was chosen to avoid entanglement with existing greedy search strategies (cf. [24, p. 38]).

28 The reader can again easily check that Theorem 6 holds for **LIC**, $\Omega^{\mathbf{At}}$ and $\mathbf{ULL^{At}}$.

29 Although the proof theories of $\mathbf{ALIC^{RS}}$ and of $\mathbf{ALIC^{At}}$ are also dynamic, $Cn_{\mathbf{ALIC^{RS}}}(\Gamma)$ and $Cn_{\mathbf{ALIC^{At}}}(\Gamma)$ are fixed as well.

30 For a general characterization of the dynamic proof theory of adaptive logics, see [4, pp. 227–229]. For the proof theory of combined adaptive logics, see [2, pp. 61–62].

As I stated above, α is derived from Γ at stage s of a proof from Γ iff α occurs at some line k of the proof, and line k is not marked at stage s. Marking is governed by the marking definition, which is applied at each stage of the proof. Let $Dab^{\mathbf{At}}(\Delta)$ denote a *Dab*-formula with $\Delta \subseteq \Omega^{\mathbf{At}}$; and let $Dab^{\mathbf{RS}}(\Delta)$ denote a *Dab*-formula with $\Delta \subseteq \Omega^{\mathbf{RS}}$.

Definition 7 *(minimal $Dab^{\mathbf{At}}$-formula).* $Dab(\Delta)$ is a minimal $Dab^{\mathbf{At}}$-formula at stage s of a proof iff

(1) $\Delta \subseteq \Omega^{\mathbf{At}}$.
(2) At stage s, $Dab(\Delta)$ is the second element of a line i on the condition $\emptyset$.
(3) There is no $\Delta' \subset \Delta$ such that $Dab(\Delta')$ satisfies condition (2).

Definition 8 *(minimal $Dab^{\mathbf{RS}}$-formula).* $Dab(\Delta)$ is a minimal $Dab^{\mathbf{RS}}$-formula at stage s of a proof iff

(1′) $\Delta \subseteq \Omega^{\mathbf{RS}}$.
(2′) At stage s, $Dab(\Delta)$ is the second element of a line i' on the condition $\Theta \subseteq \Omega^{\mathbf{At}}$.
(3′) Line i' is unmarked at stage s.
(4′) There is no $\Delta' \subset \Delta$ such that $Dab(\Delta')$ satisfies conditions (2′) and (3′).

Where $Dab(\Delta_{11}), \ldots, Dab(\Delta_{1n})$ are the minimal $Dab^{\mathbf{At}}$-formulas at stage s, $U_s^{\mathbf{At}}(\Gamma) = \Delta_{11} \cup \cdots \cup \Delta_{1n}$ is the set of unreliable **At**-formulas at stage s. Likewise, where $Dab(\Delta_{21}), \ldots, Dab(\Delta_{2m})$ are the minimal $Dab^{\mathbf{RS}}$-formulas at stage s, $U_s^{\mathbf{RS}}(\Gamma) = \Delta_{21} \cup \cdots \cup \Delta_{2m}$ is the set of unreliable **RS**-formulas at stage s.

Now everything is in place to present the marking definition, the application of which proceeds stepwise. At each stage s of the proof, lines are *first* marked/unmarked in view of $U_s^{\mathbf{At}}(\Gamma)$. *Then* lines are marked/unmarked in view of $U_s^{\mathbf{RS}}(\Gamma)$.

Definition 9 *(Marking for* **ALIC***).* Step 1: line i is marked at stage s iff, where Υ is its condition, $\Upsilon \cap U_s^{\mathbf{At}}(\Gamma) \neq \emptyset$. Step 2: after step 1, line i is marked at stage s iff, where Υ is its condition, $\Upsilon \cap U_s^{\mathbf{RS}}(\Gamma) \neq \emptyset$.[31]

Notwithstanding the dynamics of **ALIC**-proofs, the set of **ALIC**-consequences of some premise set Γ is fixed, well-defined and proof-independent (see Definition 11).

Definition 10. α is *finally derived* from Γ on line i of a proof at stage s iff (i) α is the second element of line i, (ii) line i is not marked at stage s, and (iii) any extension of the proof in which line i is marked may be further extended in such a way that line i is unmarked.

Definition 11. $\Gamma \vdash_{\mathbf{ALIC}} \alpha$ (α is *finally derivable* from Γ) iff α is finally derived on a line of a proof from Γ.

Let us briefly return to the worries raised in Section 3.2, viz. that the suggestions which eventually were incorporated in **ALIC** come across as too rash. I argued that default assumptions for causal inference are used both in scientific practice (cf. [9]) and in the literature on causal modeling (cf. [24]). I also argued that the proof theory of adaptive logics allows us (i) to trace the particular assumptions on which each inference or (intermediate) conclusion is based (cf. the condition of each line in a proof), and (ii) to trace the consequences of the violation of each particular assumption (cf. the marking definition). This gives rise to the following pragmatic picture of causal inference: given a partial list of *IRs*, **ALIC** allows us to derive a set of consequences (giving rise to a *DAG* or a pattern representing our causal beliefs). Some formulas are derived on the empty condition and hence are indubitable relative to the premises. All other consequences may be accepted provisionally. Whether these may serve as a ground for action will depend upon circumstances. Some interventions may be based on relatively slight evidence, while others need fair or even very strong evidence (cf. [9, p. 300]). Hence, if α is derived on line i on the non-empty condition Υ_i, we may either

[31] The marking definition depends on the adaptive strategy, *in casu quo* reliability. Minimal abnormality would result in a different definition (see [2, Section 6]).

decide to take action on the basis of α, or we may find that more information is needed regarding the members of Υ_i. As such, **ALIC**s dynamic proofs may act as a guide for both scientific research and policy.[32]

5.5. The semantics of ALIC

The semantics of **ALIC** bears out that final derivability is a stable notion.[33] For each premise set Γ, the set of its **ALIC**-models is a subset of its **LIC**-models: $\mathcal{M}^{\mathbf{ALIC}}(\Gamma) \subseteq \mathcal{M}^{\mathbf{LIC}}(\Gamma)$. This subset is obtained by a two-step selection. For any $M \in \mathcal{M}^{\mathbf{LIC}}$, two abnormal parts of M are defined as follows:

Definition 12. $Ab^{\mathbf{At}}(M) = \{\omega \in \Omega^{\mathbf{At}} \mid M \models \omega\}$ and $Ab^{\mathbf{RS}}(M) = \{\omega \in \Omega^{\mathbf{RS}} \mid M \models \omega\}$.

Now the selection runs as follows:

$$\mathcal{M}^0(\Gamma) =_{df} \mathcal{M}^{\mathbf{LIC}}(\Gamma), \quad \text{where } \mathcal{M}^{\mathbf{LIC}}(\Gamma) = \{M \in \mathcal{M}^{\mathbf{LIC}} \mid M \models \Gamma\}$$

$Dab^{\mathbf{At}}(\Delta)$ is a $Dab^{\mathbf{At}}$-consequence of Γ iff $\Delta \subseteq \Omega^{\mathbf{At}}$ and $Dab^{\mathbf{At}}(\Delta)$ is verified by all $M \in \mathcal{M}^0(\Gamma)$. Where $Dab^{\mathbf{At}}(\Delta_{11}), Dab^{\mathbf{At}}(\Delta_{12}), \ldots$ are the minimal $Dab^{\mathbf{At}}$-consequences of Γ, $U^1(\Gamma) =_{df} \Delta_{11} \cup \Delta_{12} \cup \cdots$

$$\mathcal{M}^1(\Gamma) =_{df} \{M \in \mathcal{M}^0(\Gamma) \mid Ab^{\mathbf{At}}(M) \subseteq U^1(\Gamma)\}$$

$Dab^{\mathbf{RS}}(\Delta)$ is a $Dab^{\mathbf{RS}}$-consequence of Γ iff $\Delta \subseteq \Omega^{\mathbf{RS}}$ and $Dab^{\mathbf{RS}}(\Delta)$ is verified by all $M \in \mathcal{M}^1(\Gamma)$. Where $Dab^{\mathbf{RS}}(\Delta_{21}), Dab^{\mathbf{RS}}(\Delta_{22}), \ldots$ are the minimal $Dab^{\mathbf{RS}}$-consequences of Γ, $U^2(\Gamma) =_{df} \Delta_{21} \cup \Delta_{22} \cup \cdots$

$$\mathcal{M}^2(\Gamma) =_{df} \{M \in \mathcal{M}^1 \mid Ab^{\mathbf{RS}}(M) \subseteq U^2(\Gamma)\}$$

This concludes the selection of the **ALIC**-models of Γ:

$$\mathcal{M}^{\mathbf{ALIC}}(\Gamma) = \mathcal{M}^2(\Gamma)$$

The meta-theoretical properties of adaptive logics are straightforward and have been studied extensively [2–4]. If the proof theory of the **LLL** is sound and complete with respect to its semantics, then so is the resulting adaptive logic's proof theory regarding to the adaptive semantics. However, the proof theory of **LIC** is sound, but not complete, with respect to the **LIC**-semantics. Hence the **ALIC**-proof theory is sound, but not complete, with regard to the **ALIC**-semantics.

5.6. ALIC and MaxEnt

Inspection of the **ALIC** semantics gives us a clear view on what it does at the level of probability functions. Recall that an **(A)LIC**-model is a triple $M = \langle \mathbb{R}^+, \mathsf{c}, \mathsf{p}\rangle$, where p is a probability distribution over $\mathbb{V}$. Intuitively, the set of **ALIC**-models of Γ consists of those **LIC**-models of Γ that verify no more abnormalities ($\omega \in \Omega^{RS} \cup \Omega^{At}$) than required by Γ. Whether $M \models \omega$, wholly depends on p. So **ALIC** indirectly selects those probability functions p that satisfy the premises, but verify no more abnormalities than required.

The **ALIC** semantics and the *MaxEnt* principle discussed in Section 3.2, are somehow similar in that they both provide a mechanism to select one or more members from a set of probability distributions (a credal set) satisfying Γ. However, if this credal set is closed and convex, *MaxEnt* selects one single probability function, whereas the set of **ALIC**-models of Γ will usually not be a singleton (and its members may also differ qua p). Moreover, there is no guarantee that for any $M \in \mathcal{M}^{\mathbf{ALIC}}(\Gamma)$, p has maximal entropy.[34]

Given this semantic difference, **ALIC** and *MaxEnt*-based causal discovery, such as the framework of Williamson [24, §§5.6–5.7], will typically lead to different results if the premises consist of a partial list of *IRs*. **ALIC** tends to

[32] This gives rise to some kind of 'reverse falsificationism'. As the acceptability of α depends on the falsehood of Υ_i's members, try to prove their truth (e.g. by gathering new data and performing new conditional independence tests). Accept α in case such proofs fail.

[33] For a general characterization of the semantics of adaptive logics, see [4, pp. 229–230]. For the semantics of combined adaptive logics, see [2, pp. 56–57].

[34] Entropy is defined as follows [24, p. 80]: $H = -\sum_{v @ \mathbb{V}} p(v) \log p(v)$ where $v @ \mathbb{V}$ is any conjunction of assignments of values to the members of $\mathbb{V}$ (cf. state descriptions for a system with families of related properties in [5, pp. 58, 70, 76]).

output unshielded colliders. Suppose that $\mathbb{V} = \{A, B, C\}$ and $\Gamma = \{\sim (A \amalg B), \sim (A \amalg C)\}$. Then **ALIC** allows to derive $A - B$, $A - C$ and $\sim (B - C)$ on the conditions $\{(A \amalg B \mid C)\}$, $\{(A \amalg C \mid B)\}$ and $\{\sim (B \amalg C)\}$ respectively. Hence $(B \to A) \wedge (C \to A)$ is derivable on the union of these conditions. By contrast, in Williamson's framework unshielded colliders are less frequent. Regarding our example, it would lead to the conclusion that $(B \amalg C \mid A)$, which rules out $(B \to A) \wedge (C \to A)$. (I leave it to the reader to check both facts.)

Which framework is most suitable is hard to determine a priori. Note that Γ may either be the result of observations from an underlying *DAG* in which $B \to A \leftarrow C$, in which case **ALIC** produces the best output, or from an underlying *DAG* in which e.g. $B \to A \to C$, in which case *MaxEnt*'s output is the best. Which framework is most suitable is under-determined by Γ.

6. Concluding remarks

In this paper, I claimed that, in scientific practice, the problem of ignorance is ubiquitous, persistent and far-reaching. I also claimed that Pearl's **IC** algorithm cannot be applied in cases of ignorance. Finally, I put forward an adaptive logic, **ALIC**, which properly solves the problem of ignorance without thereby losing the strong points of **IC**.

ALIC allows one to derive both classical, probabilistic and causal conclusions from any set of probabilistic and/or causal premises. Hence it is greatly apt for combining observational knowledge with background knowledge, common sense knowledge, etc. What is more important: **ALIC** assigns an adequate truth value to all undecided *UIR* and *CIR* (i.e. to all *IR* that are undecided even in the light of the available background knowledge, common sense knowledge, etc). This assignment is based on two rationales: firstly, that scientific models should not be overly complex, and secondly, that correlation is a useful (but not infallible) indicator of causation. But what is most important, if the interpretation of an undecided *IR* turns out to be fallacious (e.g. in the light of new premises), **ALIC** adapts itself to the premises and faulty applications of the above rationales are remedied adequately.

Acknowledgements

I would like to thank Joke Meheus, Dirk Batens, Erik Weber and Maarten Van Dyck for their helpful comments. I would also like to thank Jon Williamson, Jan-Willem Romeijn, and the other participants of the Progic 2007 (Canterbury) workshop on combining probability and logic, as well as both anonymous referees, for their criticisms, their questions and their suggestions.

Appendix A. Soundness for LIC

Theorem 7 *(Soundness for* **LIC***). If $\Gamma \vdash \alpha$, then $\Gamma \models \alpha$.*

Proof. Consider an **LIC**-proof of α from Γ. Each line of this proof either contains a premise, or an instance of (A1)–(A6), or a formula which is derived from previous formulas by either (R1) or (R2). I will show that all **LIC**-models of Γ verify α, so that $\Gamma \models \alpha$.

Consider an **LIC**-model $M = \langle \mathbb{R}^+, \mathsf{c}, \mathsf{p} \rangle$ such that v_M verifies all members of Γ. I will show that for each line i in the proof, if β is the formula derived on line i, then $v_M(\beta) = 1$. It follows that $v_M(\alpha) = 1$.

(PREM) If β is a premise, then $\beta \in \Gamma$ and, by hypothesis, $v_M(\beta) = 1$.

(R1) If $v_M(\beta') = v_M(\beta' \supset \beta) = 1$, then $v_M(\beta) = 1$ (by (Sv10)).

(A1) By (Sv7)–(Sv12), $v_M(\beta) = 1$ if β is an instance of (A1)—i.e. an axiom of propositional classical logic.

(A2) Likewise, $v_M(\beta) = 1$ if β is an instance of (A2)—i.e. a (semi-)graphoid axiom. I will prove this for the case of (G5). The proofs for the other (semi-)graphoid axioms are left to the reader.
Suppose that $v_M((\mathbf{A} \amalg \mathbf{B} \mid \mathbf{Q} \cup \mathbf{C})) = 1$ (*) and that $v_M((\mathbf{A} \amalg \mathbf{C} \mid \mathbf{Q})) = 1$ (**). From (*) it follows by

(Sv3) that for all relevant values of **A**, **B**, **C** and **Q**,[35]

$$\frac{\mathsf{p}(\mathbf{A}=\mathbf{a}\wedge\mathbf{B}=\mathbf{b}\wedge\mathbf{Q}=\mathbf{q}\wedge\mathbf{C}=\mathbf{c})}{\mathsf{p}(\mathbf{B}=\mathbf{b}\wedge\mathbf{Q}=\mathbf{q}\wedge\mathbf{C}=\mathbf{c})}=\frac{\mathsf{p}(\mathbf{A}=\mathbf{a}\wedge\mathbf{Q}=\mathbf{q}\wedge\mathbf{C}=\mathbf{c})}{\mathsf{p}(\mathbf{Q}=\mathbf{q}\wedge\mathbf{C}=\mathbf{c})}$$

and from (**) it follows by (Sv3) that for all relevant values of **A**, **B**, **C** and **Q**,

$$\frac{\mathsf{p}(\mathbf{A}=\mathbf{a}\wedge\mathbf{Q}=\mathbf{q}\wedge\mathbf{C}=\mathbf{c})}{\mathsf{p}(\mathbf{Q}=\mathbf{q}\wedge\mathbf{C}=\mathbf{c})}=\frac{\mathsf{p}(\mathbf{A}=\mathbf{a}\wedge\mathbf{Q}=\mathbf{q})}{\mathsf{p}(\mathbf{Q}=\mathbf{q})}$$

From these equations it follows that

$$\frac{\mathsf{p}(\mathbf{A}=\mathbf{a}\wedge\mathbf{B}=\mathbf{b}\wedge\mathbf{Q}=\mathbf{q}\wedge\mathbf{C}=\mathbf{c})}{\mathsf{p}(\mathbf{B}=\mathbf{b}\wedge\mathbf{Q}=\mathbf{q}\wedge\mathbf{C}=\mathbf{c})}=\frac{\mathsf{p}(\mathbf{A}=\mathbf{a}\wedge\mathbf{Q}=\mathbf{q})}{\mathsf{p}(\mathbf{Q}=\mathbf{q})}$$

But then, by (Sv3), $v_M((\mathbf{A}\amalg\mathbf{B}\cup\mathbf{C}\mid\mathbf{Q}))=1$.

(A3, A4, A5) If β is an instance of (A3), (A4) or (A5), then $v_M(\beta)=1$. The proofs for these cases are straightforward and left to the reader.

(A6) If β is an instance of (A6), then $v_M(\beta)=1$. For the first part, suppose that $v_M(A-B)=1$. It has to be shown that for all $\mathbf{Q}\subseteq\mathbb{V}\setminus\{A,B\}$, $v_M((A\amalg B\mid\mathbf{Q}))=0$. So suppose that for some such $\mathbf{Q}^*$, $v_M((A\amalg B\mid\mathbf{Q}^*))=1$. By (Sv3) it follows that for all relevant values of A, B and $\mathbf{Q}^*$,

$$\frac{\mathsf{p}(A=a\wedge B=b\wedge\mathbf{Q}^*=\mathbf{q}^*)}{\mathsf{p}(B=b\wedge\mathbf{Q}^*=\mathbf{q}^*)}=\frac{\mathsf{p}(A=a\wedge\mathbf{Q}^*=\mathbf{q}^*)}{\mathsf{p}(\mathbf{Q}^*=\mathbf{q}^*)}$$

But then, by (Sp9), $\mathsf{c}(\langle A,B\rangle)=n$. Hence, by (Sv5), $v_M(A-B)=0$, which contradicts our supposition.

For the second part, suppose that $v_M(A-B)=0$. It has to be shown that for some $\mathbf{Q}^*\subseteq\mathbb{V}\setminus\{A,B\}$, $v_M((A\amalg B\mid\mathbf{Q}^*))=1$ (i.e. that $\mathbf{Q}^*$ blocks all paths between A and B—cf. (Sp9) and (Sv3)). Define $\mathbf{Q}^*$ as follows (cf. [23, Lemma 3.1]):

$$\mathbf{Q}^*=\{X\mid A\neq X\neq B \text{ and either } X\Rightarrow A \text{ or } X\Rightarrow B\}$$

Suppose that some path $\mathcal{P}=\langle X_1,\ldots,X_n\rangle$ (with $n\geqslant 3$, $X_1=A$, and $X_n=B$) is not blocked by $\mathbf{Q}^*$. This means that for this path, $\mathbf{Q}^*$ satisfies none of the conditions (a), (b) and (c) of (Sp9). Hence, by the definition of $\mathbf{Q}^*$, for all X_i $(2\leqslant i\leqslant n-1)$:[36]

($a^\sim$) if $X_{i-1}\to X_i\to X_{i+1}$ or $X_{i-1}\leftarrow X_i\leftarrow X_{i+1}$, then $\sim(X_i\Rightarrow A)$ and $\sim(X_i\Rightarrow B)$
($b^\sim$) if $X_{i-1}\leftarrow X_i\to X_{i+1}$, then $\sim(X_i\Rightarrow A)$ and $\sim(X_i\Rightarrow B)$
($c^\sim$) if $X_{i-1}\to X_i\leftarrow X_{i+1}$, then $(X_i\Rightarrow A)$ or $(X_i\Rightarrow B)$.

Let us now consider all adjacency relations in $\mathcal{P}$, starting with X_1. If $X_1\leftarrow X_2$, then $X_2\in\mathbf{Q}^*$. But this contradicts either ($a^\sim$) (if $X_2\leftarrow X_3$) or ($b^\sim$) (if $X_2\to X_3$). So $X_1\to X_2$ (§).

What about X_2 and X_3? If $X_2\leftarrow X_3$, then $X_2\Rightarrow B$ (by (§), ($c^\sim$) and acyclicity). But then $X_3\in\mathbf{Q}^*$, which contradicts either ($a^\sim$) (if $X_3\leftarrow X_4$) or ($b^\sim$) (if $X_3\to X_4$). It follows that $X_2\to X_3$ (§§). If $n=3$, $X_2\in\mathbf{Q}^*$, which contradicts ($a^\sim$), so $n\geqslant 4$.

So what about X_3 and X_4? By the same reasoning, (§) and (§§) together imply that $X_3\to X_4$ (§§§), and hence that $n\geqslant 5$. But this implies that $X_4\to X_5$ (§§§§) and $n\geqslant 6$, etc. So $\mathcal{P}$ consists of an infinite number of nodes, which is impossible.

To conclude, $\mathbf{Q}^*$ blocks all paths between A and B and $v_M((A\amalg B\mid\mathbf{Q}^*))=1$.

(R2) If β is derived by means of (R2), and if $v_M(\beta')=1$ for all the β' used for this derivation, then $v_M(\beta)=1$. Suppose that $v_M(A-C)=v_M(C-B)=v_M(\sim(A-B))=1$. It has to be shown that $v_M(A\to C\wedge B\to C)=1$ implies that for some $\mathbf{Q}$, $v_M((A\amalg B\mid\mathbf{Q})\wedge\sim(A\amalg B\mid\mathbf{Q}\cup\{C\}))=1$, *and vice versa*.

For the first direction, suppose that $v_M(A\to C\wedge B\to C)=1$. Given that $v_M(\sim(A-B))=1$, there is some $\mathbf{Q}^*$ such that $v_M((A\amalg B\mid\mathbf{Q}^*))=1$ (see the soundness proof for (A6)). Now suppose

[35] By ‘relevant values’ I mean those values for which the conditional probabilities in question are defined.

[36] Note that for all X_i in question, $A\neq X_i\neq B$.

that $v_M((A \amalg B \mid \mathbf{Q}^* \cup \{C\})) = 1$ (†). Since $v_M(A \to C \wedge B \to C) = 1$, the triple $\langle A, C, B\rangle$ is a path between A and B. By (†), $\langle A, C, B\rangle$ must satisfy one of the conditions of (Sp9). Trivially, it cannot satisfy (a) or (b). By condition (c), $C \notin \mathbf{Q}^* \cup \{C\}$, which is impossible. Hence, contra (†), $v_M((A \amalg B \mid \mathbf{Q}^* \cup \{C\})) = 0$ and, by (Sv7), $v_M(\sim (A \amalg B \mid \mathbf{Q}^* \cup \{C\})) = 1$. So, by (Sv8), $v_M((A \amalg B \mid \mathbf{Q}^*) \wedge \sim (A \amalg B \mid \mathbf{Q}^* \cup \{C\})) = 1$.

The proof for the reverse direction is left to the reader. □

References

[1] D. Batens, Towards the unification of inconsistency handling mechanisms, Logic and Logical Philosophy 8 (2000) 5–31.
[2] D. Batens, A general characterization of adaptive logics, Logique et Analyse 173–175 (2001) 45–68.
[3] D. Batens, The need for adaptive logics in epistemology, in: D. Gabbay, S. Rahman, J. Symons, J.-P. Van Bendegem (Eds.), Logic, Epistemology and the Unity of Science, Kluwer, Dordrecht, 2004, pp. 459–485.
[4] D. Batens, A universal logic approach to adaptive logics, Logica Universalis 1 (2007) 221–242.
[5] R. Carnap, Logical Foundations of Probability, Routledge and Kegan Paul, London, 1950.
[6] N. Cartwright, How the Laws of Physics Lie, Clarendon Press, Oxford, 1983.
[7] A.P. Dawid, Conditional independence in statistical theory, Journal of the Royal Statistical Society. Series B (Methodological) 41 (1) (1979) 1–31.
[8] R. Haenni, S. Hartmann, Causality, uncertainty and ignorance, Minds and Machines 16 (3) (2006) (Special issue).
[9] A.B. Hill, The environment and disease: Association or causation? Proceedings of the Royal Society of Medicine 58 (1965) 295–300.
[10] IARC, Preamble to the IARC monographs on the evaluation of carcinogenic risks to humans, http://monographs.iarc.fr/ENG/Preamble/CurrentPreamble.pdf, 2006.
[11] D. Makinson, Bridges from Classical to Nonmonotonic Logic, Texts in Computing Series, King's College Publications, London, 2005.
[12] C. Meek, Causal inference and causal explanation with background knowledge, in: S. Hanks, P. Besnard (Eds.), Proceedings of Eleventh Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, San Francisco, 1995, pp. 403–418.
[13] J. Meheus, Do we need paraconsistency in commonsense reasoning, in: C. Delrieux, J. Legris (Eds.), Computer Modeling of Scientific Reasoning, Universidad Nacional Del Sur, Bahia Blanca, Argentina, 2003, pp. 135–145.
[14] M.S. Morgan, M. Morrison (Eds.), Models as Mediatiors. Perspectives on Natural and Social Science, Cambridge University Press, Cambridge, 1999.
[15] R.E. Neapolitan, Learning Bayesian Networks, Prentice Hall Series in Artificial Intelligence, Pearson Education, 2004.
[16] J.B. Paris, A. Vencovská, In defense of the maximum entropy inference process, International Journal of Approximate Reasoning 17 (1997) 77–103.
[17] J.B. Paris, A. Vencovská, Common sense and stochastic independence, in: D. Corfield, J. Williamson (Eds.), Foundations of Bayesianism, Kluwer, Dordrecht, 2001, pp. 204–240.
[18] J. Pearl, Causality. Models, Reasoning, and Inference, Cambridge University Press, Cambridge, 2000.
[19] P. Spirtes, C. Glymour, R. Scheines, Causation, Prediction, and Search, MIT Press, Cambridge, MA, 2000.
[20] W. Spohn, On the properties of conditional independence, in: P. Humphreys (Ed.), Patrick Suppes: Scientific Philosopher, vol. I, Kluwer Academic Publishers, Dordrecht, 1994, pp. 173–196.
[21] M. Studený, Conditional independence relations have no finite complete characterization, in: S. Kubík, J. Vísek (Eds.), Transactions of the Eleventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, vol. B, Prague, 1990, Academia, Prague, 1992, pp. 377–396.
[22] M. Van Dyck, Causal discovery using adaptive logics. Towards a more realistic heuristics for human causal learning, Logique et Analyse 185–188 (2004) 5–32.
[23] T. Verma, J. Pearl, An algorithm for deciding if a set of observed independencies has a causal explanation, in: D. Dubois, M.P. Wellman, B. D'Ambrosio, P. Smets (Eds.), Proceedings of the Eighth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, San Francisco, 1992, pp. 323–330.
[24] J. Williamson, Bayesian Nets and Causality. Philosophical and Computational Foundations, Oxford University Press, New York, 2005.