



CAUSAL DISCOVERY USING ADAPTIVE LOGICS. TOWARDS A  
MORE REALISTIC HEURISTICS FOR HUMAN CAUSAL  
LEARNING\*

MAARTEN VAN DYCK<sup>†</sup>

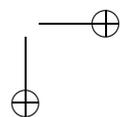
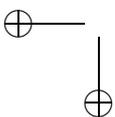
*We shall afterwards take notice of some general rules, by which we ought to regulate our judgements concerning causes and effects; and these rules are form'd on the nature of our understanding, and on our experience of its operations in the judgments we form concerning objects.* [10, p. 149]

1. *Introduction*

In this article I propose a logic that allows one to derive causal statements from probabilistic information. Of course, since long philosophers have been aware that causal statements contain relevant information that is missing from mere statements about the association between events/facts/(whatever one likes). To name but the most obvious thing: the relation between cause and effect is asymmetric. However, this should not be taken as implying the impossibility of a causal discovery logic, but rather as the warning that such a logic will not be possible without making further assumptions about the part of the world one studies — and I will follow large part of the more recent literature in AI and the philosophy of science in opting for the *causal Markov condition* as the most important assumption. I will not focus on the justification for this assumption, neither inquire in the prospects this leaves for the hope of ever attaining an adequate characterization, let alone reduction, of the notion “cause”. But I do believe that this assumption provides

\*I would like to thank: Clark Glymour for his valuable comments on the penultimate draft of this paper, Diderik Batens and a referee to this journal for their comments on some technical issues, and João Marcos for preliminary discussions on the contents of the paper.

<sup>†</sup>The author is Research Assistant of the Fund for Scientific Research – Flanders (Belgium). The research for this paper benefited from a bilateral scientific exchange project funded by the Ministry of the Flemish Community (project BIL/01/80) and the Polish State Committee for Scientific Research (KBN)..



a very good starting point from which to clarify some of the reasoning processes by which we come to judgements about what causes what. And that we do come to such judgements in a reasoned way is something that even David Hume in no way contested (one of the sections in his *Treatise* is even entitled “*Rules by which to judge of causes and effects*” — remark the imperative).

The logic to be proposed is an adaptive logic, and I will briefly explain what this means. Let me for the moment suffice with the remark that as an adaptive logic it is a member of a larger family of logics, which all serve a common goal: to get a better grip on real life reasoning processes. This is particularly important in view of the fact that there exist already automated data-mining programs for causal discovery (see especially [16, 12]), which embody the same assumptions as I use (and which served as the main source of inspiration). Notwithstanding the impressive results these seem to come up with — although this is not altogether uncontroversial<sup>1</sup> — they surely do not offer much insight in how *humans* reason from association to causation. On the other hand, the logic to be presented here will have the advantage of being constructive (and working in a much more piecemeal fashion).

In his most recent book, *The Mind's Arrow*, Clark Glymour claims that the way humans learn the causal structure of the world is illuminated by the Bayes net approach, i.e. by exploring the consequences of the causal Markov condition; at the same time Glymour has to concede that “the algorithms are unlikely psychological models” [6, p. 34], and it is in this respect that I seek to improve on the existing accounts.<sup>2</sup> In this way I hope to contribute to the program described by Alison Gopnik and Clark Glymour in a recent article:

The program we propose is therefore not to theorize that children or scientists are optimal data-miners, but rather to investigate in general how human minds learn causal maps, and how much (and, possibly, how little) their learning processes accord with Bayes net assumptions and heuristics. [7, p. 131]

The main purpose of this article is to show how the Bayes net heuristics can be reformulated, possibly providing a more realistic model for human causal learning.

Readers familiar with the Bayes net approach can skip section 2, and might want to start by having a look at section 6 for the justification of adding the present article to the already existing literature on causal discovery. Section 3

<sup>1</sup> For a very critical appreciation, see [5].

<sup>2</sup> See the concluding section, however, for an important caveat with respect to this claim. There are two possible reasons for claiming the algorithms to be unlikely psychological models. The reformulation that I suggest in the present paper only remedies one of these.

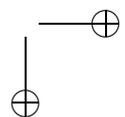
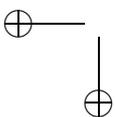
is not supposed to contain anything new for readers already familiar with adaptive logics. It is to be hoped that the other sections contain something new for all.

## 2. *Some background on causal discovery*

The main problem to be tackled concerns the discovery of the causal *structure* of systems. To correctly assess the results it is important to keep in mind that, no matter how important it may be, this is only a preliminary step in many investigations. All that one can conclude with the logic to be presented, is that certain aspects of a system (possibly) have a causal influence on other aspects. Nothing will be said about the (functional) form of these influences, not even about the question whether they are positive (contributory) or negative (inhibitory). This obviously means that the presented results in no way will be sufficient to answer traditional questions about which causes are necessary or sufficient etc.<sup>3</sup> But still, before any of these questions can be answered, one first has to discover the causal structure of the system(s) one studies. This is all the logic will enable us to do: infer (part of) the topology of the causal mechanisms that are responsible for the observed behavior of systems. Moreover, this has to be taken quite literally: it is not assumed that we obtain this information from performing manipulations, but rather from passive observation. Of course, manipulations — if possible — can teach us much more — but they aren't always possible.

Causal structure can be discovered either in just one system, or in a set of systems that are assumed to share the same structure. An example of the first case could be that one tries to discover the causal structure of the workings of one particular piece of household equipment (which buttons influence which functions). A typical example of the second case shows up in medical investigations where one tries to discover causal mechanisms shared by all people sharing certain characteristics. In any case, it has to be possible to gather enough observations about the behavior of the system(s), so that one can start from premisses stating that certain aspects of the system(s) are (un)correlated with other aspects (e.g. the temperature of the refrigerator is statistically independent from the position of the red button, there is a statistical dependency between smoking and having lung cancer). To this end the characteristic aspects of the modelled system will be designated with variables (e.g. position of the red button, temperature of the refrigerator, smoking, having lung cancer), so that the state of the system can be represented by

<sup>3</sup>For an interesting approach to these questions, from the Bayes net perspective, see chapters 9 and 10 in [12].



assigning a value to these variables (position of the red button=position II, temperature of the refrigerator= $6^{\circ}C$ , smoking=no, having lung cancer=yes). It is assumed that every system only has a finite number of variables, and that every variable only has a finite number of possible values.<sup>4</sup> As an example, imagine that we have 180 observations on the behavior of the refrigerator, in which we check the temperature (which is assumed to be either 3, 6, or  $9^{\circ}C$ ) and the position of the red button (either position I, II, or III), and that we have the findings reported in table 1.

position button	temperature	number of cases
I	3	20
	6	19
	9	21
II	3	20
	6	20
	9	20
III	3	19
	6	21
	9	20

Table 1. Observations on the refrigerator

Even if these observations do not properly permit such a conclusion, it could be that one decides on the basis of them that it is true that the position of the red button and the temperature of the refrigerator are independent variables. I will immediately give a definition that makes clear what it means for variables to be independent, but first we need a further extension of this notion.

In addition to information about the independency — and its negation, dependency — of variables, we also need premisses stating that two variables are *conditionally* independent given another (set of) variable(s). For instance, it could be that one finds that the position of the red button and the temperature are dependent, but that if one looks only at those observations in which a second, green, button is always in the same position, the position of the red button and the temperature are independent: in this case it is said that those two variables are conditionally independent given the position of the green button. A standard definition for conditional independency between (sets of) variables is the following:

<sup>4</sup> Even if one measures the values of a continuous variable like temperature, the obtained results will always be intervals, due to the limitations of every measurement. In this way the relevant scale of temperature will be divided in a finite number of possible measurement outcomes.

*Definition 1: (Conditional Independency)* For  $\Lambda = \{\lambda_1, \lambda_2, \dots\}$  a finite set of variables,  $P(\cdot)$  a probability function over these variables,<sup>5</sup> and  $X \subset \Lambda, Y \subset \Lambda, Z \subset \Lambda$ , we then say that  $X$  and  $Y$  are conditionally independent given  $Z$  if

$$P(X = x \mid Y = y \ \& \ Z = z) = P(X = x \mid Z = z) \\ \text{whenever } P(Y = y \ \& \ Z = z) > 0, \quad (1)$$

in which  $x, y, z$  stand for all possible values associated with possible configurations of the variables in  $X, Y$ , and  $Z$  respectively.

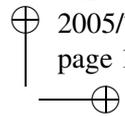
I will use the standard notation  $(X \perp\!\!\!\perp Y \mid Z)$  to express that two sets of variables  $X$  and  $Y$  are independent conditional on the set of variables  $Z$ .  $X$  and  $Y$  are unconditionally independent if  $(X \perp\!\!\!\perp Y \mid \emptyset)$  holds, which for simplicity I will write as  $(X \perp\!\!\!\perp Y)$ . From now on I will also write expressions like the one in (1) as:  $P(X \mid Y \ \& \ Z) = P(X \mid Z)$ .

It is clear that this definition only makes sense if the relevant probabilities are somehow available, whereas strictly speaking all that we can observe are observational data (as exemplified in table 1). There are well known statistical techniques which can be used in deciding which probabilistic conclusions are warranted, but I will not comment on this. As already mentioned, I am primarily interested in the logic behind the reasoning process from (in)dependency to causality. Moreover, as will be noted in section 6, humans often guess the presence or absence of correlations, without bothering about the precise statistical information (which may be hard to assess).

The premisses of the logic for causal discovery will consist of statements about conditional and unconditional (in)dependencies between the selected variables. It has to be assumed that there are no conceptual relations between the different variables (e.g. being inside a box and being outside the same box are not considered to be different variables). The main reason for this assumption is that we will try to find a causal explanation for *every* dependency that holds between the variables; if the possibility of conceptual relations were not excluded, this would imply that there can be cases where we would search for causal relations whereas clearly there are none.

As already mentioned in the introduction, information about the correlations that hold in the system(s) is not enough to justifiably come to causal statements. To this end we need some further constraints on the possible causal structures of systems (constraints, which, as we will see, in some cases impose a direction on the dependency between two variables). These

<sup>5</sup>I assume that conditional probability  $P(a \mid b)$  is defined as  $P(a \ \& \ b)/P(b)$ .



constraints will then be reflected in the inference rules of the causal discovery logic. In this way it will become possible to derive from the premisses all causal structures that are jointly compatible with these premisses and those constraints.

I will follow the important work of Spirtes, Glymour and Scheines [16] and Pearl [12], in opting for the following three assumptions: the *causal Markov condition*, the *faithfulness condition*, and the *acyclicity* of all structures. (The first one is the conceptually most important one, whereas the other two are to be considered methodological presuppositions which render the task of causal discovery more feasible.)

### 2.1. The causal Markov condition

The basic assumption underlying the Bayes net approach is that all causal structures responsible for the observed dependencies and independencies are causal Markov chains. This means that the following condition is always met:

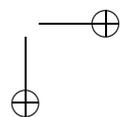
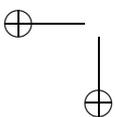
*Definition 2: (Causal Markov Condition) For all distinct variables  $X$  and  $Y$  in a causal sufficient variable set  $\Lambda$ , and  $P(\cdot)$  a probability function over these variables, if  $X$  does not cause  $Y$ , then:*

$$P(X | Y \ \& \ \text{parents}(X)) = P(X | \text{parents}(X)), \tag{2}$$

where  $\text{parents}(X)$  is the subset of all variables in  $V$  that have a direct causal influence on  $X$ .

To fully understand this definition, it is necessary to introduce some further specifications on the notion *causal structure*. To this end it is useful to represent the causal structure of a system by a set of nodes and arrows between them. The nodes represent the variables, and the arrows causal influences (which are always asymmetric). A sequence of nodes connected by arrows not pointing towards each other is called a *path*. Only *acyclic* causal structures will be considered, i.e. all possible structures satisfy the condition that there is no path from one variable to itself. It is said that one variable *causes* another if there is a path from the former to the latter; that one variable has a *direct causal influence* on another if they are connected by an arrow pointing from the former to the latter. (Of course, effect and direct effect are the natural complementary notions.) A set of variables is called *causally sufficient* when for any two variables in the set which have a common cause, this common cause is also in the set.

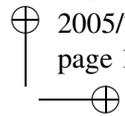
I will not attempt a further characterization of what it is to have causal influence, but I assume that everybody would agree that pushing a ball has a



causal influence on the movement of the ball, that not brushing one's teeth has a causal influence on having caries, that the explosion of a star has a causal influence on its surrounding planets, that the causal influence of my pushing the light switch on the light in the room is mediated by the influence of the electric current on the light bulb, etc. Moreover, it is clear that if one adds the causal Markov condition as a constraint on all causal structures, this will provide a partial characterization of the notion causal influence. To further spell out this implication, let us briefly try to understand what this condition comes down to. It is widely recognized to be a generalization of Reichenbach's *common cause condition* (on this condition, see [13, 15]); and this condition was meant to capture two ideas: that all correlations have a causal explanation, and that common causes *screen off* their correlated effects.

If the causal Markov condition holds of a system, and if one knows that  $P(X | Y \& \text{parents}(X)) \neq P(X | \text{parents}(X))$  (as always  $X, Y$  are variables characterizing the system), then one immediately knows that it must be the case that  $X$  causes  $Y$ . This is already one case in which we see that dependency implies causation. As a special case we can immediately see that if one knows that  $\text{parents}(X)$  is empty, then it must be the case that unconditional dependency between  $X$  and  $Y$  implies that  $X$  causes  $Y$ . And if  $X$  and  $Y$  are the only variables in  $\Lambda$ , then a correlation always implies that one causes the other.<sup>6</sup> But of course, before one can make these particular inferences, one first has to know the set of parents of a variable, that is, one needs enough causal knowledge to start with. And this looks worrisome for the task of causal discovery, since it might seem impossible to get the discovery process started (if one knows the set of parents, large part of the job is already done). But let us first go on with our inquiry in the meanings of definition 2 (and when it comes to meaning, it was already indicated that no reductive analysis will be attempted — clearly, definition 2 would on first analysis turn out to be circular in this respect).

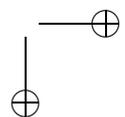
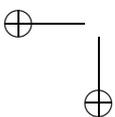
<sup>6</sup>Remark also that if in this case  $X$  and  $Y$  are found to be independent — this will be the case whenever they are not related as cause and effect — their values of course still can vary (according to the probability distribution they satisfy). This can reflect two possible situations: either these variables are governed by an intrinsic stochastic process, or there are *external* causes responsible for the variation. In the second case we speak of external causes, reflecting that these causes are no part of the considered system, which implies that they are supposed to influence at most one variable (the set is causally sufficient). This is another instance of the fact, mentioned further on in the text, that one needs enough causal knowledge to start with. (Consult chapter 1 in [12] for the relevant theorems stating that a causal model is Markovian if all so-called background factors are jointly independent; i.e. each cause, not included in the system influences at most one variable.) Nevertheless, there also exist some algorithms for causal inference in the presence of unobserved common causes.



Another way of stating the causal Markov condition is that conditional on its parents, a variable is independent of every other variable except its effects. In Reichenbach's language, we can say that the set of parents *screens off* a variable and all its non-effects. If for instance a variable  $Y$  is a cause of  $X$ , but does not have a direct causal influence on  $X$ , then this means that its influence can be screened off, i.e. rendered superfluous, since it is entirely transmitted by (part of) the set of parents of  $X$ ; no further dependency between  $Y$  and  $X$  besides the one already holding between  $\text{parents}(X)$  and  $X$  can ever show up. Hence, we can understand the Markov property as a consequence of the idea that causal influences are local in space and time. If one holds fixed (in mind) the direct causes of  $X$ , then the values of its indirect causes will no longer be correlated with the values of  $X$ . Moreover, screening off does not only apply to indirect causes, but also to correlated effects of a common cause. If two variables share at least one variable in their set of causes (whether direct or indirect), this will (very often)<sup>7</sup> show up in a dependency between these two variables, even if neither of the two is an effect of the other. But then, if one is dealing with a causal Markov system, this dependency will disappear if one conditions on the parents of one of the two correlated variables. In the simplest case, consisting of only three variables  $X$ ,  $Y$  and  $Z$ , in which  $Z$  is a direct cause of both  $X$  and  $Y$ , this means that  $Z$  will screen off  $X$  and  $Y$ . In more complex cases in which there can be more than one common cause, which moreover can be indirect causes, sometimes only the complete set of parents will do.

There exists an easy-to-use graphical criterion, called *d-separation* (consult [12, pp. 16–17], or the appendix of the present article), which allows one to judge if a set of variables screens off two (sets of) variables (assuming that one is dealing with a Markov system). The causal intuitions behind this criterion are easily explained. In the causal structure depicted in figure 1, it is clear that it is not enough to hold fixed  $B$  to render the influence of  $A$  on  $D$  superfluous, since  $A$  also influences  $D$  via the path through  $C$ . One also immediately sees that  $B$  and  $C$  are *d-separated* (screened off) by  $A$ . Often there will be more than one set of variables that screen off two variables. Again in figure 1,  $E$  and  $A$  are screened off as well by the variable  $D$ , as by the set  $\{B, C\}$ . More complex cases are also common. Consider the structure in figure 2:  $A$  and  $D$  will only be screened off by the set consisting of  $B$  and  $C$ . In this case the reason is more subtle. It is immediately clear that holding only  $C$  fixed will not suffice, but it is also the case that holding only  $B$  fixed will neither. The reason for the latter is the effect known

<sup>7</sup>I will give some comments on this caveat in section 2.2



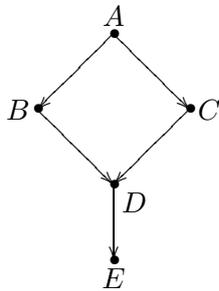


Figure 1.  $A$  screens off  $B$  and  $C$ ;  $\{B,C\}$  screens off  $D$  and  $A$ ; etc.

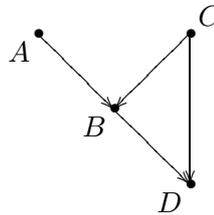
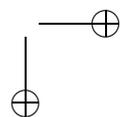
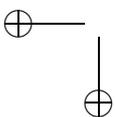
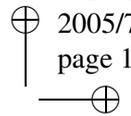


Figure 2. Only  $\{B, C\}$  screens off  $D$  and  $A$

as *explaining away*: if one conditions on a common effect of two independent variables, these two variables will be rendered dependent.<sup>8</sup> An example will make this clear. Imagine a teacher at a girl's college who has the tendency to favour blond girls and smart girls. If one is told that one of his favoured pupils is not smart, this will increase the probability of her being blond. Still, in the total population of his pupils there is no correlation to be found between colour of the hair and IQ. It is only if one conditions on the subset of all favoured pupils that such a dependency shows up. The reason is clear: the state of being a favoured pupil is positively influenced by blondness and smartness, so the knowledge that one of these two factors is missing increases the chance that the other is present. If we now turn back to the situation depicted in figure 2, it is clear that by holding fixed  $B$ , the independent variables  $A$  and  $C$  — that they are independent follows directly from definition 2.1 and the assumption that we are dealing with a Markov system — will become dependent, but then, because of the direct influence of  $C$  on  $D$ ,  $A$  and  $D$  will also be dependent, so not screened off! Only holding fixed both  $B$  and  $C$  will do the job (because when also holding fixed  $C$ , the dependency between  $A$  and  $C$  will disappear).

In line with reasonings like the one in the last paragraph, we can see how particular structures of causal influences give rise to distinct patterns of conditional and unconditional dependencies and independencies. The basic idea behind the causal discovery approach is to exploit those patterns that point

<sup>8</sup>Thus, in figure 1,  $B$  and  $C$  that were independent when conditioned on  $A$  will become dependent again if also conditioned on  $D$ . This is entirely compatible with the causal Markov condition, since this only says that the set of parents screen off from non-effects; it is silent about any other set of variables.



unambiguously to a particular structuring of the influences. Of course, for this to be a justifiable methodology a justification for the Markov condition has to be provided. I will not attempt such here, but refer the reader to the excellent article by Hausman and Woodward [9]. Let me suffice with a few remarks on this topic. First consider the case of mediated causal influences. It is clear that many of the ideas associated with causal mechanisms would fail if indirect causes were not screened off from their effects by intermediate causes, for it is essentially this fact that guarantees the distinctness of the involved causal influences. If intermediate causes would not screen off, then keeping the value of an intermediate cause fixed would still leave the indirect cause efficacious; but then what would it mean to call this an *indirect* cause? To put it another way, the validity of the causal Markov condition guarantees that if one holds fixed the direct causes of  $X$ , then tinkering around with its indirect causes (or anything else that could happen with them that does not have any effect on the direct causes of  $X$ , except, of course, through the path from the indirect to the direct causes) will not have any influence on  $X$ . The case of correlated effects of common causes is similar, but more controversial (the main critiques coming from Nancy Cartwright, e.g. [3]).<sup>9</sup> There seems to be something very counterintuitive involved in assuming that common causes need not screen off. This would imply that causal and informational relevance somehow would come apart in a hard-to-understand way, for then the value that one of both effects takes can always provide further information on the value of the other effect, even if the value of the common cause was already known, i.e. everything of causal relevance was specified already. But how would this informational relevance come about if it is not through causal influences? So it seems that somehow we must be mistaken in our specification of the situation: either there exists a causal influence of one of both effects on the other, or there is some missing common cause that was left unspecified (but that would do the screening off). But this is exactly what the causal Markov condition tells us. These remarks point to one other important fact. If one chooses the wrong (or not enough) variables, the Markov condition will not always hold of causal systems in the world. Even if it is assumed that all causal systems satisfy the Markov condition, this will hold only at the right level of description. This is the most important instance of the fact that one needs a good deal of causal knowledge to start with. But this does not *per se* diminish the value of the Markov condition, which anyway "can play an important heuristic role in discovering causal

<sup>9</sup>The quantum-mechanical EPR correlations involved in Bell's inequalities seem to provide another important source of suspicion; but then, the perplexities that stem from these peculiar systems, underline the fact that something like the Markov condition is deeply entrenched in our conceptions of causality.

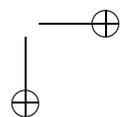
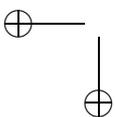
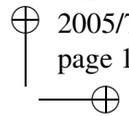
structure, in the sense that its apparent failure suggests that one has left out causally relevant information" [9, p. 580].

To sum up this rather long discussion, let me try to formulate the (for our purposes) most important property of systems satisfying the causal Markov condition as succinct as possible. If  $X$  and  $Y$  are not related as cause and effect, then they will be independent conditional on their respective sets of parents. If moreover they have no common causes, then they will also be independent conditional on the empty set, since in this case their parents will do no relevant screening off (they are no causal intermediaries between  $X$  and  $Y$ , nor can they be causal intermediaries between common causes and  $X$  or  $Y$ , or common causes themselves).<sup>10</sup> Thus, if the causal Markov condition holds and if  $X$  and  $Y$  are distinct variables, then  $\neg(X \amalg Y)$  implies either that  $X$  causes  $Y$ , or that  $Y$  causes  $X$ , or that they are correlated effects of a common set of variables in  $\Lambda$ ; equivalently,  $\neg(X \amalg Y)$  implies either that  $X$  has a direct causal influence on  $Y$ , or that  $Y$  has a direct causal influence on  $X$ , or that there is a set of variables in  $\Lambda$  that screens off  $X$  and  $Y$ . Both statements will be used in building the adaptive logic for causal discovery.

## 2.2. Faithfulness

Before inferring causal relations from probabilistic information becomes feasible, one further complication has to be dealt with. Consider the situation depicted in figure 3 ('+' and '-' stand respectively for contributory and inhibitory influences). It is clear that, depending on the exact form of the influences, it is possible that the influence through  $C$  exactly cancels the direct influence that  $A$  has on  $B$ . But if this possibility is allowed for, one can only make the inference from dependency to causality, not from independency to no causal influence. In that case, there would be much more causal structures compatible with particular patterns of dependencies and independencies, and this would severely diminish the prospects for causal discovery. However, although such a situation is possible, in a large class of cases it is

<sup>10</sup> A more formal way to see this. First remark, as already indicated in footnote 6, that every causal structure for which the external causes — i.e. causes that are not included in the structure — cause at most one variable, is a Markovian causal structure (consult theorem 1.4.1 in [12]). Then recall that it was remarked already that in every causal Markov structure consisting of only two variables that are not related as cause and effect, these two variables are unconditionally independent. Thus: if  $X$  and  $Y$  (supposedly not related as cause and effect) are unconditionally dependent, the structure consisting of only these two variables would be no Markov structure; this would imply that a larger structure containing these variables would be no Markov structure neither, *unless* this structure would contain the common cause that was left out of the structure consisting of just the two variables. So, if  $X$  and  $Y$  are not related as cause and effect and have no common causes among  $\Lambda$ , then if the causal Markov condition holds for the system characterized by  $\Lambda$ , they are unconditionally independent.



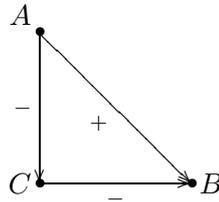
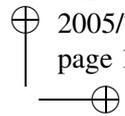


Figure 3.  $A$  and  $B$  can be causally related but independent

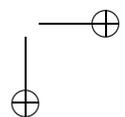
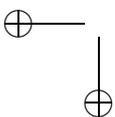
also highly unlikely to occur. The reason for this is clear: there has to be a considerable amount of fine-tuning between the different influences before they will exactly cancel each other. The *faithfulness* assumption states that this is so unlikely that we can assume that it does not occur.<sup>11</sup> This does not only hold for two variables where one causes the other, but also for effects of a common cause; in general one can state the assumption that a probability distribution is faithful to a causal structure as follows (using the aforementioned  $d$ -separation criterion to judge causal relevance in a structure):<sup>12</sup>

*Definition 3: (Faithfulness Condition)* For all disjoint sets of variables  $X$ ,  $Y$  and  $Z$  in  $\Lambda$ , if in the underlying causal structure it is not the case that  $X$  is  $d$ -separated from  $Y$  given  $Z$ , then in the probability distribution  $P$  we have that  $X$  and  $Y$  are conditionally dependent given  $Z$ , where  $X$  and  $Y$  are not empty but  $Z$  may be.

Judea Pearl opts for the suggestive name of *stability*: “This restriction conveys the assumption that all the independencies embedded in [the probability distribution]  $P$  are stable; that is, they are entailed by the structure of the model . . . and hence remain invariant to any change in parameters [specifying the functional form of the influences]” [12, p. 48]. Pearl is referring to the unstableness of any fine-tuning between different causal influences — remember that they have to cancel each other *exactly* for an independency to show up. And indeed, if one is not dealing with goal-directed systems that were explicitly designed to cancel certain unwanted effects, it seems that this unstableness of parameter fine-tuning will be good enough a reason

<sup>11</sup> There is some controversy about how this unlikeliness has to be interpreted, and hence about how justified this assumption is; I will not comment on this issue, but refer the interested reader to [14].

<sup>12</sup> I follow the presentation as given in [4].



to assume faithfulness (although, sometimes we might be mistaken in doing so).

The issues surrounding this assumption point to an important problem with the general strategy behind the Bayes net approach: it is fundamentally sensitive to zero-correlations between variables (whether unconditional or conditional — see the importance of screening off), but of course this is extremely hard to establish in practice. Notice that I sidestepped this issue by assuming that somehow we do have the right probability-distribution at our disposal (and in doing so skip the step from sample statistics to population probability).

In conclusion we can say that whereas the causal Markov condition states that unconditional dependency implies causal relations (whether direct, indirect, or through common causes), the faithfulness assumption states that unconditional independency implies the absence of causal relations. Thus, Markov plus faithfulness imply an *equivalence* between variables being unconditionally dependent and causally related (possibly through common causes, as indicated at page 15 at the end of section 2.1). Any structure for which this is valid is called a causal Bayes net. In section 4 I will put these assumptions to work, and in doing so hopefully make more clear their usefulness and strength, but first I will introduce the basic ideas behind adaptive logics.

### 3. Adaptive logics

Adaptive logics were originally developed to deal with a particular problem arising when a set of premisses give rise to an inconsistency (for an overview, see [2]). It is of course well known that classical logic turns out to be completely impotent in this situation: the consequence set is trivial, following the classical property of *ex falso quodlibet*. In view of this, so-called *paraconsistent* logics were developed; the most obvious way is to drop certain rules of classical logic (e.g. *modus tollens* and *disjunctive syllogism*), so that *ex falso quodlibet* no longer follows. These logics, however, have the important drawback of being rather weak: *no* application of the dropped rules remains valid. Consider what this can come down to: if one uses a paraconsistent logic that invalidates *modus tollens*, then ' $\neg q$ ' will not follow from the following set of premisses  $\{q \supset p, \neg p, r, \neg r\}$ . So, avoiding the triviality that arises from the inconsistency comes at a price; and a price that is hard to pay, since  $p$  and  $q$  have nothing to do with the inconsistency! The solution proposed by Diderik Batens in developing *inconsistency-adaptive logics* was the following: not the inference rules of classical logic should be dropped, but certain *applications* thereof (i.e. the ones leading to triviality). This is implemented by introducing three basic components for such an adaptive logic: a *lower limit logic* (henceforth LLL), an *upper limit logic* (ULL), and

a *marking strategy*. The LLL contains all the unproblematic rules, and thus is a paraconsistent logic of some sort, the ULL is made up of the rules of the LLL *plus* the problematic rules, and thus is classical logic. The important feature of the ULL rules is that they are introduced *conditionally* in an adaptive proof: any application of such a rule can be retracted *whenever* it turns out that the condition is violated (the sentence that was derived is marked as invalid, the line of the proof containing this sentence is out). The condition is the set of formulas that have to behave "*normal*" for the specific ULL rules to be applicable; normality obviously being linked with consistency for inconsistency–adaptive logics, the marking strategy deciding when marking has to occur. (In the simple example given above, the application of modus tollens to derive  $\neg q$  would have to be retracted if it turned out that  $\neg p$  behaved inconsistently, i.e. if  $p$  would be derived somewhere in the proof — obviously this can only be the case in the presence of further premises.)<sup>13</sup> It is clear that all the rules from the LLL can be treated as unconditional rules (these being the rules that cannot cause trouble, i.e. that never lead to triviality in the presence of inconsistencies). But of course, if the sentences to which one applies such unconditional rules were derived on a condition, then the sentence that is being introduced at a new line in the proof will also contain a condition: the union of all the conditions of the earlier sentences of which it is a consequence (if an earlier line has to be considered invalid at a certain point of the proof, all its consequences obviously also have to be).

The foregoing already indicates how the idea of adaptive logics can be extended to deal with other problems besides the presence of inconsistencies. Particular interesting are so-called *ampliative* adaptive logics. For these logics, the LLL will often be classical logic, and the ULL will allow one to go further than permitted by classical logic, where the particular criteria for marking (and definitions of abnormality) will determine when "further" means "too far". Examples are an adaptive logic for induction [1], and one for abduction [11]. As was already stressed extensively by David Hume, causal reasoning is a species of inductive reasoning, so it will not come as a surprise that the adaptive logic for causal discovery will also be of this kind. All these logics share the interesting characteristic that they allow one to formalize what makes a particular ampliative argument a *correct* one (of course relative to the chosen adaptive logic), which is notoriously impossible from the viewpoint of classical logic (as Hume taught us).

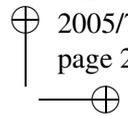
<sup>13</sup> Further complications can arise if one considers the possibility of expressions stating a disjunction of abnormalities: in these cases a line will sometimes have to be marked when its condition is present in such a disjunction. So the link between marking and abnormal behaviour of members of a condition is not always as straightforward as might be thought on first impression, this being determined by the marking strategy; the adaptive logic for causal discovery to be presented here will, however, display such a straightforward connection.

The fact that lines that were already derived can be marked as invalid at a later stage in a proof gives adaptive logics a dynamic character. The most interesting dynamics occurs if the normality of formulas is not decidable from the premisses (remember that the non-propositional part of classical logic is not decidable!). In this case, all one can do is assume the normality of a formula until it can explicitly be shown to behave abnormal, that is, by writing down a line of the proof that expresses the abnormality of that formula.<sup>14</sup> (And even then, more can happen: it can also be the case that a line that is marked will be unmarked later on in the proof. Those who are becoming curious or suspicious by remarks like these, should look for an article explaining the properties of adaptive logics in more detail — here, I will only expand on those properties that my (simple) adaptive logic will expose.) But also if one is dealing with a decidable marking criterion, this dynamic character can arise; if new premisses are added to a proof after part of it was already completed, it can very well be that some consequence of these premisses (maybe together with the old premisses) turns out to be an abnormality for some line that was already derived, causing this line to be marked. This fact makes clear that adaptive logics are non-monotonic (i.e. adding further formulas to a set of premisses can invalidate some consequences of the original set). The first kind of cases can be called instances of an *internal* dynamics (this is linked with the proof theory proper, reflecting some properties of natural reasoning), whereas the second kind exemplify an *external* dynamics (this is linked with the inference relation being non-monotonic). A combination of both kinds of dynamics of course can arise. To get a better grip on all these properties, the study of adaptive logics also incorporates semantic conceptions (see e.g. [2]), but in the present article I will stay on the proof-theoretic level.

Before turning to the adaptive logic for causal discovery, let me briefly recapitulate. Adaptive logics always have something like the following proof format. Every line of the proof consists of five elements (the presence of a fifth element being characteristic for adaptive logics):

- (1) a line number,
- (2) the sentence derived,
- (3) the line numbers of the sentences from which (2) is derived,
- (4) the rule of inference that justifies the derivation,
- (5) the set of sentences on the normal behaviour of which we rely in order for (2) to be derivable by (4) from the sentences of the lines enumerated in (3).

<sup>14</sup> As already remarked in footnote 13, depending on the marking strategy one opts for, marking can already occur when a line is written down that states a disjunction of abnormalities. As the adaptive logic that I will introduce does not have that characteristic, I will give no further comments on this.



In addition to a structural rule by which one introduces premises in a proof (always with an empty fifth element), there are two kinds of inference rules: the unconditional one (being all valid applications of the LLL rules), and the conditional one (all the applications of the extra ULL rules). The fifth element of a line, together with the marking strategy, determines when that line will have to be marked in view of the other lines written down in the proof.

#### 4. An adaptive logic for causal discovery

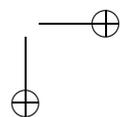
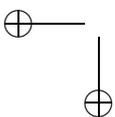
##### 4.1. The basics

The basic idea is very simple: the logic for causal discovery will allow one to commit the classic fallacy of *cum hoc, ergo propter hoc* — that is, of inferring a direct causal relation from a correlation between variables; well, not quite, it allows one to do this *unless* it can be shown that this dependency is screened off by another set of variables ...

It is clear how this will be implemented: there will be a LLL validating a certain set of consequences from a set of premisses, and a ULL that is more daring and posits that any unconditional dependency is due to a direct causal relation. If this turns out to be too daring, that is, if the condition of a line in the proof is violated — obviously, this will be the case whenever an unconditional dependency is shown to be screened off — then this line will have to be retracted. It is the ULL that will allow one to go ahead in inferring causal relations — and as such will be essential to get the discovery process started — but, as will become clear, it is the LLL that will help to infer the direction of the direct influences posited by the ULL.

Before we go on, let me first clarify the notational conventions that I will use. As indicated in section 2, the objects we are talking about are the nodes of a causal structure (corresponding to the variables of a system), and the causal arrows between them. So, any structure can be described with a finite set of names for the nodes, say  $\Lambda = \{A, B, \dots, A_n\}$ , and predicative expressions like ' $A \rightarrow B$ ', which obviously states that the characteristic of the system denoted by  $A$  has a direct causal influence on the one denoted by  $B$ . As meta-variables for  $A, B$ , etc. I use lower case Greek letters  $\alpha, \beta$ , etc. Upper case Greek letters are used for sets of nodes. Besides the arrows, I also have to introduce the symbol  $(\cdot \amalg \cdot | \cdot)$  for stating probabilistic (in)dependency between (sets of) nodes. That there is a causal path between  $\alpha$  and  $\beta$  will be expressed by the predicate  $\mathcal{P}(\alpha, \beta)$ , which can be given a recursive definition ( $\gamma$  is an arbitrary member of  $\Lambda$ ):

*Definition 4:* (Causal Path)  $\mathcal{P}(\alpha, \beta) =_{\text{def}} \alpha \rightarrow \beta \vee (\mathcal{P}(\alpha, \gamma) \wedge \gamma \rightarrow \beta)$ .



A very central place is of course occupied by the notion of screening off;  $SO_{\alpha\beta}$  will denote the set of all formulas that state that two probabilistically dependent nodes  $\alpha$  and  $\beta$  are screened off by a non-empty set of nodes in  $\Lambda$  by:

*Definition 5: (Screening Off)* For  $\alpha$  and  $\beta$  which satisfy  $\neg(\alpha \amalg \beta)$ :  $SO_{\alpha\beta} = \{(\alpha \amalg \beta | \Delta) : \alpha, \beta \notin \Delta; \emptyset \subset \Delta \subset \Lambda\}$ .

The union of all subsets of nodes that screen off  $\alpha$  and  $\beta$  is denoted by  $\Psi(SO_{\alpha\beta})$ . The disjunction of all formulas that state that  $\alpha$  and  $\beta$  are screened off is denoted by  $DSO_{\alpha\beta} = \bigvee \{(\alpha \amalg \beta | \Delta) : \alpha, \beta \notin \Delta; \emptyset \subset \Delta \subset \Lambda\}$  (again this is only defined for probabilistically dependent  $\alpha$  and  $\beta$ ).

#### 4.2. The Lower Limit Logic

The rules and axioms characterizing the LLL can be divided in two categories (that together will constitute the unconditional rule of the adaptive logic)<sup>15</sup>. I will only focus on the second kind, since the first kind is of course already well-documented: these are rules and axioms characterizing classical predicate logic with identity. The second kind of rules and axioms express properties satisfied by all faithful, causal Markov structures, that is, by causal Bayes nets.

I will first state a list of axiom-schemes and inference rules, then I will briefly comment on them, and finally introduce some derived inference rules, which will be at the core of causal discovery. The list of axioms and rules introduced is not intended to be exhaustive, but suffices for my purposes.

$$\text{A1 } \neg\mathcal{P}(\alpha, \alpha)$$

$$\text{A2 } \neg(\alpha \amalg \beta) \equiv (\alpha \rightarrow \beta \vee \beta \rightarrow \alpha \vee DSO_{\alpha\beta} \vee \alpha = \beta)$$

$$\text{A2}' \neg(\alpha \amalg \beta) \equiv (\mathcal{P}(\alpha, \beta) \vee \mathcal{P}(\beta, \alpha) \vee (\exists \gamma)(\mathcal{P}(\gamma, \alpha) \wedge \mathcal{P}(\gamma, \beta)) \vee \alpha = \beta)$$

$$\text{R1 } \frac{DSO_{\alpha\beta}, [\mathcal{P}(\alpha, \gamma) \wedge \mathcal{P}(\gamma, \beta)] \vee [\mathcal{P}(\beta, \gamma) \wedge \mathcal{P}(\gamma, \alpha)] \vee [\mathcal{P}(\gamma, \alpha) \wedge \mathcal{P}(\gamma, \beta)] \vee (\exists \delta)[\mathcal{P}(\delta, \gamma) \wedge \mathcal{P}(\gamma, \alpha) \wedge \mathcal{P}(\delta, \beta)] \vee (\exists \delta)[\mathcal{P}(\delta, \alpha) \wedge \mathcal{P}(\delta, \gamma) \wedge \mathcal{P}(\gamma, \beta)]}{}$$

<sup>15</sup> However, when giving some examples of proofs with the adaptive logic, I will, for clarity's sake, still indicate applications of these different kinds separately. So the fourth element of a line that is the result of the application of the unconditional rule, will state the name of specific rules (e.g. CL for all valid applications of a rule of classical logic, DR1 for an application of that LLL-rule), instead of a generic name.

$$\gamma \in \Psi(\mathcal{SO}_{\alpha\beta})$$

$$\text{R2 } \frac{\neg(\alpha \amalg \beta), \quad \gamma \in \Psi(\mathcal{SO}_{\alpha\beta})}{[\mathcal{P}(\alpha, \gamma) \wedge \mathcal{P}(\gamma, \beta)] \vee [\mathcal{P}(\beta, \gamma) \wedge \mathcal{P}(\gamma, \alpha)] \vee [\mathcal{P}(\gamma, \alpha) \wedge \mathcal{P}(\gamma, \beta)] \vee (\exists \delta)[\mathcal{P}(\delta, \gamma) \wedge \mathcal{P}(\gamma, \alpha) \wedge \mathcal{P}(\delta, \beta)] \vee (\exists \delta)[\mathcal{P}(\delta, \alpha) \wedge \mathcal{P}(\delta, \gamma) \wedge \mathcal{P}(\gamma, \beta)]}$$

$$\text{R3 } \frac{\gamma \in \Psi(\mathcal{SO}_{\alpha\beta})}{\neg(\alpha \rightarrow \gamma \wedge \beta \rightarrow \gamma)}$$

Axiom **A1** states the assumption that all structures are acyclic (so that there can be no feed-back loops). Axioms **A2** and **A2'** are clearly not independent, but are the two equivalent ways of expressing what I singled out in section 2 (on pages 15 and 17) as the most important property of faithful causal Markov systems. Rules **R1**, **R2**, and **R3** mirror some properties of the  $d$ -separation criterion, and thus express assumptions about causal relevance. **R1** looks more gruesome than it really is; it states that if two nodes are screened off, and if a third node either lies on a path between these two nodes, or is a common cause of them, or lies on a path from a common cause to one of both nodes, that then this third node must be a member of one of the sets that screen off the two nodes. This is clearly in line with the meaning of screening off that I discussed in section 2. (In appendix A I give a proof of the fact that **R1** follows from the  $d$ -separation criterion.) **R2** inversely states that if two nodes are screened off by a third node, then this node either has to lie on a path between the two nodes, or has to be the common cause of these nodes, or has to lie on a path from a common cause to one of both nodes. (A proof is to be found in the appendix). **R3** is an immediate consequence of what was called explaining away on page 13. (See again the appendix for the very short proof.)

The following easy-to-prove derived rules are of main import for causal discovery.

$$\text{DR1 } \frac{\alpha \neq \beta, \quad (\alpha \amalg \beta), \quad \alpha \rightarrow \gamma \vee \gamma \rightarrow \alpha \vee DSO_{\alpha\gamma}, \quad \beta \rightarrow \gamma \vee \gamma \rightarrow \beta \vee DSO_{\beta\gamma}}{\alpha \rightarrow \gamma \vee DSO_{\alpha\gamma}, \quad \beta \rightarrow \gamma \vee DSO_{\beta\gamma}}$$

$$\text{DR2 } \alpha \neq \beta, \quad DSO_{\alpha\beta}, \quad \gamma \notin \Psi(\mathcal{SO}_{\alpha\beta}),$$

$$\frac{\alpha \rightarrow \gamma \vee \gamma \rightarrow \alpha \vee DSO_{\alpha\gamma}, \quad \beta \rightarrow \gamma \vee \gamma \rightarrow \beta \vee DSO_{\beta\gamma}}{\alpha \rightarrow \gamma \vee DSO_{\alpha\gamma}, \quad \beta \rightarrow \gamma \vee DSO_{\beta\gamma}}$$

As one can notice, these rules allow one to infer the direction of causal influences given the right kind of causal (!) premises; i.e. given the absence of the third disjunct in the (already partly causal) premisses on lines 3 and 4 (or lines 4 and 5 for the second derived rule) and in the conclusion. Together with R3, which also excludes certain directions, this will make possible the derivation of (parts of the) causal structures responsible for observed (in)dependencies. But before we can do that, we have to get rid of the extra disjuncts; and this is where the adaptive character of the logic comes into play.

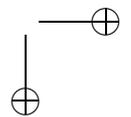
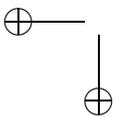
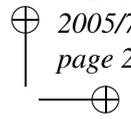
### 4.3. AL<sub>cd</sub>

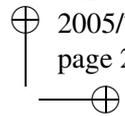
As explained in section 3, an adaptive logic is not only characterized by a LLL but also by an ULL and a marking strategy, where the ULL incorporates some presuppositions ("normalities") not made by the LLL. In section 4.1, I already introduced the most important idea of the logic for causal discovery, henceforth called AL<sub>cd</sub>. The ULL validates *cum hoc, ergo propter hoc* for any two nodes, so the presence of a formula stating that these two nodes are screened off is considered an *abnormality*. (As indicated this name goes back to the origin of adaptive logic in inconsistency adaptive logics, so call it otherwise if you might feel uncomfortable with this name; but it is not an altogether unnatural perspective to consider "correlation = causation" as a normality.)

The ULL of AL<sub>cd</sub> is obtained by adding the following rule to the LLL that I introduced in section 4.2:

$$\text{Rc} \frac{\neg(\alpha \amalg \beta)}{\alpha \rightarrow \beta \vee \beta \rightarrow \alpha \vee \alpha = \beta}$$

This rule will always be applied conditionally in an AL<sub>cd</sub> proof: if a line is added to a proof as a result of an application of Rc, then this line will have as its fifth element  $\mathcal{SO}_{\alpha\beta}$ . It is immediately clear from A2 that the LLL doesn't make the presupposition that the ULL makes, and that the validity of the presupposition is enough to guarantee the correctness of the ULL rule from the perspective of the LLL (a fact known as "derivability adjustment theorem" in the study of adaptive logics). As a definition for a marked line we obviously have:





*Definition 6: (Marked Line) Where  $\Theta$  is the fifth element of line  $i$ , line  $i$  is marked iff a formula  $\delta$  is unconditionally derived for some  $\delta \in \Theta$ .<sup>16</sup>*

So, from the moment that a line in a proof states that  $\alpha$  and  $\beta$  are screened off, all lines that were derived on the condition that  $\alpha$  and  $\beta$  are not independent conditional on any variable in  $\Lambda$  have to be marked.

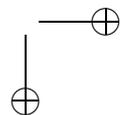
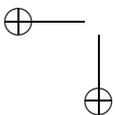
Now, there is one more important thing to remark. If one looks at the two derived rules of the LLL, it is clear that they give rise to two useful variants that are valid in  $AL_{cd}$ : all lines containing the causal statements and the disjuncts of screening off formulas (*DSO*) can be replaced by lines containing only the causal statements and having as condition the absence of any formula stating that the nodes mentioned in the causal statements are screened off. In these variants the interplay between ULL and LLL becomes most clear: the ULL introduces the conditional lines with causal statements, and the LLL helps to select causal statements specifying one unequivocal direction for the causal influences. Finally we have to remark that in view of the presence of the disjunct  $\alpha = \beta$  in  $Rc$ , we will also have to introduce in any proof premisses stating that two nodes with different names are really distinct.

### 5. Some examples

In this section I will present three examples of causal discovery using  $AL_{cd}$ . While giving these examples, I will also comment on an apparent weakness of this approach, that might already have been bothering some attentive readers; i.e. is the distinction between ULL and LLL really needed? (After all, if one already knows that there is no screening off, **A2** suffices to introduce causal statements in a proof.)

Let me start with a very simple example, just to get the taste of the thing. Suppose we have a system with the causal structure depicted in figure 4. This will give rise to the stated premisses,<sup>17</sup> which immediately allow us to

<sup>16</sup>Those familiar with adaptive logics will notice that I opt for the "simple strategy", thus assuming that disjunctions of abnormalities cannot be derived (if they could, one should opt for another strategy, "reliability" or "minimal abnormality"). The absence of such disjunctions is no matter of principle, but rather of convention. In almost any natural application of a logic for causal discovery the premisses will consist of statements stating (un)conditional (in)dependency between variables, as inferred from observations; thus the abnormalities will almost always be given in a straightforward form. However, it is important to know that the mere possibility of premisses that would introduce disjunctions of abnormalities in a proof pose no principled problems, and only require a minimal technical revision of  $AL_{cd}$ . (I don't incorporate this in the logic presented here, to keep the basic format as simple as possible for readers not familiar with adaptive logics but rather interested in causal discovery.)



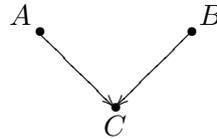


Figure 4. Example 1

derive the underlying structure.

1	$A \neq B$	–	PREM	$\emptyset$
2	$A \neq C$	–	PREM	$\emptyset$
3	$B \neq C$	–	PREM	$\emptyset$
4	$(A \amalg B)$	–	PREM	$\emptyset$
5	$\neg(A \amalg C)$	–	PREM	$\emptyset$
6	$\neg(B \amalg C)$	–	PREM	$\emptyset$
7	$A \rightarrow C \vee C \rightarrow A$	2,5	Rc	$\mathcal{SO}_{AC}$
8	$B \rightarrow C \vee C \rightarrow B$	3,6	Rc	$\mathcal{SO}_{BC}$
9	$A \rightarrow C$	1,4,7,8	DR1	$\mathcal{SO}_{AC}$
10	$B \rightarrow C$	1,4,7,8	DR1	$\mathcal{SO}_{BC}$

Now for something more serious. Consider again the structure depicted in figure 2. The premisses then should be the following:

1	$A \neq B$	–	PREM	$\emptyset$
2	$A \neq C$	–	PREM	$\emptyset$
3	$A \neq D$	–	PREM	$\emptyset$
4	$B \neq C$	–	PREM	$\emptyset$
5	$B \neq D$	–	PREM	$\emptyset$
6	$C \neq D$	–	PREM	$\emptyset$
7	$\neg(A \amalg B)$	–	PREM	$\emptyset$
8	$(A \amalg C)$	–	PREM	$\emptyset$
9	$\neg(A \amalg D)$	–	PREM	$\emptyset$
10	$\neg(B \amalg C)$	–	PREM	$\emptyset$
11	$\neg(B \amalg D)$	–	PREM	$\emptyset$
12	$\neg(C \amalg D)$	–	PREM	$\emptyset$
13	$(A \amalg D B \& C)$	–	PREM	$\emptyset$

With RC we can infer the following causal relations, but it is also clear that in view of the premisses, line 15 immediately has to be marked:

<sup>17</sup> That it must give rise to these premisses follows from the assumption that it is a faithful Markov structure — if this assumption holds, it will immediately be shown that  $AL_{cd}$  will allow us to derive the right structure; if this assumption doesn't hold,  $AL_{cd}$  will clearly fail, but we knew this already, didn't we?

14	$A \rightarrow B \vee B \rightarrow A$	1,7	Rc	$SO_{AB}$
15	$A \rightarrow D \vee D \rightarrow A$	3,9	Rc	$SO_{AD} \sqrt{13}$
16	$B \rightarrow C \vee C \rightarrow B$	4,10	Rc	$SO_{BC}$
17	$B \rightarrow D \vee D \rightarrow B$	5,11	Rc	$SO_{BD}$
18	$C \rightarrow D \vee D \rightarrow C$	6,12	Rc	$SO_{CD}$

The fact that we knew in advance that line 15 would have to be marked is the weakness I mentioned in the beginning of this section. This clearly has to do with the fact that we are dealing with a decidable marking criterion (given the premisses, it can always be decided in advance if a line will have to be marked). But this need not mean that the adaptive approach is empty posturing for our purposes. It could very well be that the premisses we are considering are incomplete, and that there is a relevant missing variable that screens off  $B$  and  $C$ . This would mean that the Markov condition will not hold for the considered variables, and that  $AL_{cd}$  will give us wrong answers. But, and this is the important point, this logic has got the resources to adapt itself to new relevant information, and so to correct for the initial "mistakes". It is not the marking in light of already given premisses that is really important, but the marking after the addition of new premisses. (In the words of section 3,  $AL_{cd}$  displays an external dynamics.) One can never be *guaranteed* that there is not somewhere a variable that screens off two variables: an adaptive logic still allows one to go on in a sensible way in view of this uncertainty. This should be enough to clear up the suspicion that I stated in the beginning of this section: in reality we never *know* for sure that there is no screening off, so ULL and LLL really do behave differently. One could answer to this that from a logical point of view reality is only illusory, and that if the premisses state that two objects are conditionally dependent given any other object in our language (all variables included in the structure), then one knows there is no screening off. But even adopting this point of view, there still remains an important difference between ULL and LLL: whenever the premisses *don't* give *all* information on the conditional (in)dependencies holding between two variables — this is not only a logical unavoidable situation (remember we are talking premisses) but also a very realistic situation — ULL, but not LLL, assumes that there is no screening off and introduces causal statements.

Let us now continue the proof ('CL' as a fourth element of a line indicates that this line is the result of an application of rules of classical logic).

19	$A \rightarrow B$	1,8,14,16	DR1	$SO_{AB}$
20	$C \rightarrow B$	1,8,14,16	DR1	$SO_{BC}$
21	$\neg(A \rightarrow B \wedge D \rightarrow B)$	13	R3	$\emptyset$
22	$\neg D \rightarrow B$	19,21	CL	$SO_{AB}$
23	$B \rightarrow D$	17,22	CL	$SO_{BD}$

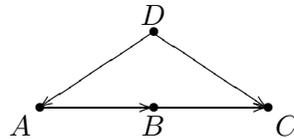


Figure 5. Example 3

24	$\neg(C \rightarrow B \wedge B \rightarrow D \wedge D \rightarrow C)$	-	A1	$\emptyset$
25	$\neg D \rightarrow C$	20,23,24	CL	$\mathcal{SO}_{BC} \cup \mathcal{SO}_{BD}$
26	$C \rightarrow D$	18,25	CL	$\mathcal{SO}_{BC} \cup \mathcal{SO}_{BD} \cup \mathcal{SO}_{CD}$

As one can see,  $AL_{cd}$  indeed enables us to infer the structure responsible for the observed (in)dependencies.<sup>18</sup> In this example one can clearly see the importance of the fact that applications of unconditional rules carry over the conditions attached to lines of a proof. For example, line 26 will have to be marked from the moment it is found out that either  $C$  and  $B$ , or  $B$  and  $D$ , or  $C$  and  $D$  are screened off. This is indeed what one should expect, since this line was derived by using the axiom scheme stating that there can be no cyclic paths; if e.g.  $B$  and  $C$  are screened off, it could be the case that they are effects of a common cause, but then there could be no cyclic path anyway; so this could no longer provide any information on the directionality of the causal influence between  $C$  and  $D$ . It is this behavior that justifies the claim that  $AL_{cd}$  has got the resources to adapt itself to new relevant information.

The last example is meant to show the general limitations of the causal discovery approach: not all structures can be recovered using only probabilistic information (of course, it is always possible that temporal information, or performed manipulations, provide further clues that can help to solve the resulting indeterminacy). An example is the structure depicted in figure 5. Again the premisses follow from the fact that this is supposed to be a faithful Markov system, and the rest of the proof proceeds normally (I will not introduce the lines that will immediately have to be marked).

1	$A \neq B$	-	PREM	$\emptyset$
---	------------	---	------	-------------

<sup>18</sup>Of course, I am cheating in my examples, since I start by inferring the (in)dependencies from the structure. As already remarked, given the correctness of the assumptions of faithfulness and Markov property, this is an innocuous strategy — this in no way means that these assumptions are innocuous, and our examples can surely provide no support for them: this can only be provided by tackling real-life problems, and showing that one can also recover the right causal structures there.

2	$A \neq C$	–	PREM	$\emptyset$
3	$A \neq D$	–	PREM	$\emptyset$
4	$B \neq C$	–	PREM	$\emptyset$
5	$B \neq D$	–	PREM	$\emptyset$
6	$C \neq D$	–	PREM	$\emptyset$
7	$\neg(A \amalg B)$	–	PREM	$\emptyset$
8	$\neg(A \amalg C)$	–	PREM	$\emptyset$
9	$\neg(A \amalg D)$	–	PREM	$\emptyset$
10	$\neg(B \amalg C)$	–	PREM	$\emptyset$
11	$\neg(B \amalg D)$	–	PREM	$\emptyset$
12	$\neg(C \amalg D)$	–	PREM	$\emptyset$
13	$(A \amalg C   B \& D)$	–	PREM	$\emptyset$
14	$(B \amalg D   A)$	–	PREM	$\emptyset$
15	$A \rightarrow B \vee B \rightarrow A$	1,7	Rc	$SO_{AB}$
16	$A \rightarrow D \vee D \rightarrow A$	3,9	Rc	$SO_{AD}$
17	$B \rightarrow C \vee C \rightarrow B$	4,10	Rc	$SO_{BC}$
18	$C \rightarrow D \vee D \rightarrow C$	6,12	Rc	$SO_{CD}$
19	$B \rightarrow C$	5,14,17,18	DR2	$SO_{BC}$
20	$D \rightarrow C$	5,14,17,18	DR2	$SO_{CD}$
21	$\neg(D \rightarrow A \wedge B \rightarrow A)$	14	R3	$\emptyset$
22	$(A \rightarrow B \wedge A \rightarrow D) \vee (B \rightarrow A \wedge A \rightarrow D) \vee (A \rightarrow B \wedge D \rightarrow A)$	15,16,21	CL	$SO_{AB} \cup SO_{AD}$

It is easy to check that all structures compatible with the conclusions of this proof will give rise to the same premisses; it is clear that on basis of these premisses alone it is impossible to decide between these structures. So, although there is considerable number of cases in which our assumptions impose a directionality on the causal influences between different nodes, it is not the case that this will always be possible. Nevertheless, when applying  $AL_{cd}$  to model human reasoning in causal discovery, it will immediately become clear that in practice this reasoning is rich in content-based assumptions that have to be added to the premisses. These will often claim the impossibility of particular causal influences, most importantly because of some temporal information or knowledge about the possible causal mechanisms responsible for the observed (in)dependencies, and will enable one to further delimit the possible causal structures.

## 6. On human causal learning

Recent literature in psychology reports that there is empirical evidence supporting the hypothesis that human adults and even children represent causal

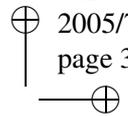
relationships in ways that can be described as causal Bayes nets, and moreover that even children might make causal inferences without discriminating between potential causes and potential effects beforehand [8].

The heuristics proposed in the present article is clearly based on the assumptions underlying the Bayes nets approach, and its final output will always be a Bayes net (or a disjunction of Bayes nets); at the same time it works in a completely piecemeal and constructive fashion. The natural way to derive causal structure using  $AL_{cd}$  is as follows: start with the data on a few nodes and look for the possible causal structure, then consider more nodes — possibly, it can turn out that some of these provide screening off conditions of earlier nodes, leading to a revision of earlier formed hypotheses — and so on. The major way in which the heuristics that is being proposed here might be an advancement over the usual approaches is that when new information is introduced, one need not start computing the possible causal structures again *de novo* (due to the non-monotonic character of the logic behind the heuristics). On this ground the present heuristics might be claimed to be a possible starting point for providing a more realistic model for human causal inference.

It must be stressed that  $AL_{cd}$  is a *logic* for causal inference. It is not meant to provide an entirely accurate description of human reasoning, but a normative model against which it can be judged — however, not any model will do, only a relevant one. A good normative model need not coincide with the actual reasoning processes used, but must resemble them closely enough, so that the latter can meaningfully be judged against the former. As already indicated by the reference to the limited memory and processing capacity of human reasoners, it is utterly unrealistic to expect from people trying to uncover the causal structure of a system that they should assess all data at once; e.g. in common situations, not all independencies are remarked at once.<sup>19</sup> As is clear, this is the kind of situation for which  $AL_{cd}$  is designed. It allows reasoners to be daring, without being foolish; i.e. *cum hoc, ergo propter hoc* may be applied, but always with a condition attached to the conclusion. It is important to stress that newly added premisses causing the revision of a conditional conclusion do not contradict the former premisses, but nevertheless can invalidate some of the conclusions drawn from them — an important characteristic of all non-monotonic reasoning; it is not that one was mistaken, but rather that one was missing some crucial information.

One last caveat: in assuming that we start from the population correlations, I sidestepped one major problem for the discovery of causal structure in the world. This is another important aspect in which human reasoners

<sup>19</sup>The set of conditions attached to a derived conclusion can also play a useful heuristic role in this, by pointing towards interesting hypotheses to be investigated.



differ from automated search programs: they often guess population correlations — e.g. by pretending that two observed cases were a two-hundred — without starting from all numerical statistical information.<sup>20</sup> Obviously this is problem is in no way remedied by the heuristics proposed here. Moreover, in [8] it is exactly this problem that is singled out as one of the major obstacles for the causal Bayes net approach to provide realistic models of human reasoning.

*Appendix A. R1, R2, R3 and d-separation*

The *d*-separation criterion, as introduced by Judea Pearl [12, pp. 16–17], reads as follows (adapted to our notation and terminology):

*Definition 7: (d-Separation)* Two nodes  $\alpha$  and  $\beta$  are said to be *d*-separated by a set of nodes  $\Delta$  iff for every sequence of arrows and nodes connecting  $\alpha$  and  $\beta$  (not necessarily through a directed path):

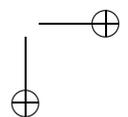
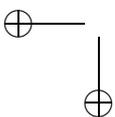
- (1) this sequence contains a path  $\eta \rightarrow \delta \rightarrow \epsilon$  or a fork  $\eta \leftarrow \delta \rightarrow \epsilon$  such that the middle node  $\delta$  is in  $\Delta$ , OR
- (2) this sequence contains an inverted fork  $\eta \rightarrow \sigma \leftarrow \epsilon$  such that the middle node  $\sigma$  is not in  $\Delta$  and such that no effect of  $\sigma$  is in  $\Delta$ .

If two nodes in a Markov structure are *d*-separated, then the two corresponding variables are independent conditional on the set of variables that are responsible for the *d*-separation, and the faithfulness condition states the inverse.

Let us now look at rule R1. The first line states that two nodes  $\alpha$  and  $\beta$  are *d*-separated, so that there is at least one non-empty<sup>21</sup> set  $\Delta$  such that members of this set satisfy the conditions stated in definition 7. If then it is the case, as stated in the second line of R1, for a node  $\gamma$  that either it lies on a path between  $\alpha$  and  $\beta$ , or it is a common cause of  $\alpha$  and  $\beta$ , or it lies on a path from a common cause to  $\alpha$  or  $\beta$ , then it follows that it cannot be the case that this node is not a member a set of nodes that screens off  $\alpha$  and  $\beta$ . (The second condition of definition 7 cannot be satisfied for the sequences of arrows and nodes between  $\alpha$  and  $\beta$  constituted by these paths; and since the definition states a necessary and sufficient condition, any node  $\gamma$  that is such a middle node as mentioned in the first condition will be a member of

<sup>20</sup>This guessing could be — partly — corrected by adopting Bayesian learning algorithms (cf. [17])

<sup>21</sup>This does not follow from the definition of *d*-separation proper, but from the definition of  $DSO_{\alpha\beta}$ .



a set of nodes that screens off  $\alpha$  and  $\beta$ , if they were only connected through these paths. Moreover, since we exclude acyclic structures, these nodes will be members of a set of nodes that screen off  $\alpha$  and  $\beta$  anyhow. *Proof:* The only reason that such a node  $\gamma$  would be no member of a set of nodes that screen off could be that it is positioned at, or is an effect of, an inverted fork of another sequence connecting  $\alpha$  and  $\beta$ ; but even this would not be enough, for if this sequence would contain another inverted fork  $\eta \rightarrow \sigma \leftarrow \epsilon$  or a path  $\eta \rightarrow \delta \rightarrow \epsilon$  or a fork  $\eta \leftarrow \delta \rightarrow \epsilon$ , then  $\alpha$  and  $\beta$  still can be  $d$ -separated by a set containing  $\gamma$  but not containing  $\sigma$  and any of its effects, or also containing  $\delta$  — so only if none of this is the case could  $\gamma$  not be a member of a set of nodes that screen off; this means that for this to hold we should have either  $\alpha \rightarrow \sigma \leftarrow \beta \wedge \mathcal{P}(\sigma, \gamma)$  or  $\alpha \rightarrow \gamma \leftarrow \beta$ , but then we would have at least one directed cycle — this can easily be seen by drawing all the possibilities.) So, it follows that  $\gamma \in \Psi(\mathcal{SO}_{\alpha\beta})$ . QED.

Now for rule R3. If it is the case that  $\alpha \rightarrow \gamma \leftarrow \beta$ , then by definition 7 it immediately follows that the set of nodes  $d$ -separating  $\alpha$  and  $\beta$  cannot contain  $\gamma$ , so that  $\gamma \notin \psi(\mathcal{SO}_{\alpha\beta})$ . QED.

Centre for Logic and Philosophy of Science  
University Ghent, Belgium  
E-mail: maarten.vandyck@UGent.be

#### REFERENCES

- [1] Diderik Batens. On a logic of induction. To appear.
- [2] Diderik Batens. Inconsistency-adaptive logics. In Ewa Orłowska, editor, *Logic at Work. Essays Dedicated to the Memory of Helena Rasiowa*, pages 445–472. Physica Verlag (Springer), Heidelberg, New York, 1999.
- [3] Nancy Cartwright. Causal diversity and the markov condition. *Synthese*, 121:3–27, 1999.
- [4] Gregory F. Cooper. An overview of the representation and discovery of causal relationships using bayesian networks. In Clark Glymour and Gregory F. Cooper, editors, *Computation, Causation, and Discovery*, pages 3–62. AAAI Press/MIT Press, Menlo Park (Cal.)/Cambridge (Mass.), 1999.
- [5] David Freedman and Paul Humphreys. Are there algorithms that discover causal structure? *Synthese*, 121:29–54, 1999.
- [6] Clark C. Glymour. *The Mind's Arrow. Bayes Nets and Graphical Causal Models in Psychology*. MIT Press, Cambridge (Mass.), 2001.



- [7] Alison Gopnik and Clark Glymour. Causal maps and bayes nets: a cognitive and computational account of theory-formation. In Peter Carruthers, Stephen Stich, and Michael Siegal, editors, *The Cognitive Basis of Science*, pages 117–132. Cambridge University Press, Cambridge, 2002.
- [8] Alison Gopnik, Clark Glymour, David M. Sobel, Laura E. Schulz, Tamar Kushnir, and David Danks. A theory of causal learning in children: Causal maps and bayes nets. *Psychological Review*, 111:3–32, 2004.
- [9] Daniel M. Hausman and James Woodward. Independence, invariance and the causal markov condition. *British Journal for the Philosophy of Science*, 20:521–583, 1999.
- [10] David Hume. *A Treatise of Human Nature*. Clarendon Press, Oxford, second edition, 1978. Ed. by L.A. Selby–Bigge.
- [11] Joke Meheus, Liza Verhoeven, Maarten Van Dyck, and Dagmar Provijn. Ampliative adaptive logics and the foundation of logic-based approaches to abduction. In Lorenzo Magnani, Nancy J. Nersessian, and Claudio Pizzi, editors, *Logical and Computational Aspects of Model-Based Reasoning*, pages 39–71. Kluwer Academic, Dordrecht, 2002.
- [12] Judea Pearl. *Causality. Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, 2000.
- [13] Hans Reichenbach. *The Direction of Time*. University of California Press, Berkeley, 1956.
- [14] James M. Robins, Richard Scheines, Peter Spirtes, and Larry Wasserman. Uniform consistency in causal inference. Technical Report 725, Carnegie Mellon University Department of Statistics, 2000.
- [15] Wesley C. Salmon. *Scientific Explanation and the Causal Structure of the World*. Princeton University Press, Princeton, 1984.
- [16] Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. Springer Verlag, New York, 1993.
- [17] Mark Steyvers, Joshua B. Tenenbaum, Eric-Jan Wagenmakers, and Ben Blum. Inferring causal networks from observations and interventions. *Cognitive Science*, 27:453–489, 2003.

